Proceedings of

Measuring Behavior 2005

5th International Conference on Methods and Techniques in Behavioral Research

Wageningen, The Netherlands
30 August – 2 September 2005

Editors
Lucas P.J.J. Noldus
Fabrizio Grieco
Leanne W.S. Loijens
Patrick H. Zimmerman

Noldus
Information Technology

2005

Bibliographical reference

Proceedings of Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research (Wageningen, The Netherlands, 30 August – 2 September 2005). Edited by L.P.J.J. Noldus, F. Grieco, L.W.S. Loijens and P.H. Zimmerman. Wageningen: Noldus Information Technology.

ISBN 90-74821-71-5

Copyright

P.O. Box 268
6700 AG Wageningen
The Netherlands
www.noldus.com

Printed in the Netherlands

Table of Contents

Symposia

Symposium 1: Automatic facial expression analysis and synthesis

Symposium 2: Defining and measuring dominant-submissive behavior

Symposium 3: Innovation in pain research: development and validation of novel behavioral assays for the prediction of analgesic efficacy

Symposium 4: The automation of observing and analyzing rodent behavior: possibilities and limitations

Symposium 5: Measuring and analyzing dispersal

Symposium 6: Standardized measurement methods for patients with neurological disorders

Symposium 7: Innovative use of the water maze to assess neurocognitive abilities in laboratory rodents

Symposium 8: Annotating and measuring meeting behavior

Symposium 9: Automatic detection of abnormal human behavior using video processing

Session 3: Measuring physiology and activity

Session 4: Sound, speech and gesture analysis

Session 5: Recording behavior: methods and techniques

Session 6: Measuring body motion

Session 7: Measuring the user experience

Session 8: Tracking and analyzing fish movement

Session 9: Automated behavior recognition, part 1

Session 10: Automated behavior recognition, part 2

Session 11: Analyzing the sequential structure of behavior

Session 12: Posture and gait analysis

Session 13: Observational methods: novel applications

Session 14: New techniques in human factors research

Poster group 2: Behavioral phenotyping of rodents

Poster group 3: New animal models and test paradigms

Poster group 4: Measuring body posture and gait

Poster group 5: Analyzing social behavior

Poster group 6: Innovation in observational data collection

User Meetings

User meeting 1: EthoVision

User meeting 2: The Observer

User meeting 3: Theme

Program Chair's Welcome

On behalf of the Scientific Program Committee I welcome you to Measuring Behavior 2005, the 5th International Conference on Methods and Techniques in Behavioral Research. I trust you arrived safe and sound and are eager to join the conference. After four very successful conferences in Utrecht, Groningen and Amsterdam, it is now Wageningen to set the stage for science and pleasure. You are invited to meet Wageningen, its University and Research Centers and explore and enjoy its lovely surroundings. But what is more important, you will be served a stimulating scientific program.

Throughout the years more and more scientists from all over the world discover this conference as the place to be for exchanging ideas on how to measure behavior. In fact, its growing popularity nearly got us in trouble this 5th conference, since we were getting overbooked and had to turn down many requests! One of the attractive aspects of this conference is that it creates the possibility for communication between scientists of very different backgrounds. From behavioral ecologists to psychiatrists, from physiologists to linguists. They are all intrigued with behavior one way or another, but their research questions, objects and applications may greatly differ. They may literally and figuratively speak very different languages but what binds them is a common goal: to enhance the accuracy, speed, ease and reliability of measuring, quantifying, analyzing or visualizing behavior. New methods, new techniques, new applications. As always the focus of Measuring Behavior is on 'how' and not on 'why' questions. This focus results in a very productive exchange and cross-fertilization of researchers from fields that would otherwise never meet.

As you can learn from the program, the conference includes a selection of events such as oral and poster presentations, special symposia, demonstrations, special interest group meetings, tutorials, and user meetings. Do not miss the special symposia where topics are grouped into a central theme. To name but a few: detecting abnormal human behavior, innovation in pain research, innovative use of the water maze and measuring olfaction in insects and mammals. Be eager to learn and visit the tutorials. They provide a valuable opportunity to instruct you in specific methods, techniques and equipment you can use in your behavioral research. There is a whole selection of topics such as measuring human and animal movement, remote physiological monitoring, data manipulation and management. Go deep and dive into a Special Interest Group (SIG) where participants focus on a specific methodological or technical topic and go for in-depth discussions.

Research centers in and around Wageningen invite you to visit their facilities during the scientific tours. There is a choice of eight different sites. This includes my own institute, the Netherlands Institute of Ecology (NIOO-KNAW), which has one of its centers, the Centre for Terrestrial Ecology, close to Wageningen. I look forward to showing you around.

There is so much to share and learn, it may become hard to drag you away to our informal social events, with good food and a relaxing atmosphere. But do not miss the opportunity to have dinner at the Rosendael Castle and to visit one of the Netherlands most beautiful nature conservation sites "De Blauwe Kamer". On behalf of the Scientific Program Committee and the Local Organizing Committee I wish you an immeasurably fun and interesting stay in Wageningen.

Louise E.M. Vet
Chair, Scientific Program Committee

Editors’ Preface

This volume contains short papers of most presentations at Measuring Behavior 2005. It is the first time printed conference proceedings are published in addition to the program and abstracts books, and we are pleased with the positive response of presenters to our request to submit papers for the proceedings. We understand that for most authors, the ultimate destination of their research findings is an article in a peer-reviewed journal with a high impact factor. These conference proceedings do not attempt to compete with such journals. Instead, this book offers a complementary publication opportunity, in which authors can elaborate on technical details of their methods and instruments – including drawings, photographs and software screenshots – in a way not allowed by most research journals. A selection of the papers presented at the conference will be published as full papers in *Behavior Research Methods*, the journal that has been devoting a special issue to our conference since Measuring Behavior ’98. In order to maintain a clear distinction between the short papers in the proceedings and the full papers in *Behavior Research Methods*, the length of short papers was limited to 4 pages for oral presentations and 2 pages for posters.

The response to this year’s conference has been very positive in terms of submissions and participation. In line with the cross-disciplinary aims of the conference, we had a good balance between human and animal behavior topics, and between methodological and technical sessions. Eleven special symposia dealt with topics as diverse as measuring and analyzing facial expression, dominant-submissive behavior, pain, rodent behavior, dispersal, neurological disorders, learning and memory, meeting behavior, abnormal human behavior, zebra fish behavior, and olfaction. This year’s special interest groups covered interesting new data acquisition and signal analysis techniques in the area of measurement of human motion, telemetric and behavioral monitoring, and Virtual Reality for skills training and performance measurement.

Last but not least we pride ourselves with the contribution of three very interesting keynote speakers. Jeffrey Cohn (Pittsburgh, USA) spoke about automated facial image analysis; Richard Morris (Edinburgh, UK), the inventor of the classical water maze task for rodents, addressed the move from spatial learning to episodic-like and semantic-like memory; and Sergio Velastin (Kingston upon Thames, UK) reviewed the state of the art in intelligent camera surveillance for the detection of abnormal human behavior. Dr Velastin’s lecture received extensive coverage in Dutch radio programs and newspaper articles, illustrating the timely nature of the topic.

The organization of Measuring Behavior 2005 would not have been possible without the diligent efforts of a large number of volunteers and the staff members in the conference office. We thank in particular the members of the Scientific Program Committee, chaired by Louise Vet, for supporting us in reviewing abstracts and giving feedback to authors. All submissions were reviewed by at least three (for oral papers) or two (for posters and demonstrations) reviewers, none of whom are associated with the organizing company. This allowed us to keep the quality of the work presented at Measuring Behavior at a high level, and to avoid a conflict of interest. As editors we have done our best to assist authors with textual refinement of their paper, keeping in mind the multidisciplinary readership of this volume.

Our colleagues of the Local Organizing Committee did an excellent job in making it all happen, before and during the conference. Together with the tour guides and the student volunteers, they deserve to be applauded. We are also grateful to our sponsors and exhibitors, who are listed at the end of this volume.

Invaluable support for the submission and review process and for preparing the proceedings has been provided by the conference secretariat and local organizing committee members Mechteld Ballintijn, Marjoleine Bessels, Cécile Bruisten, Yvonne Leander and Joeke van Santen.

Our final thanks go to all Measuring Behavior 2005 authors who have enabled us to put together a high-quality, diverse and exciting conference program and proceedings.

Lucas Noldus
Fabrizio Grieco
Leanne Loijens
Patrick Zimmerman

Noldus Information Technology bv
Wageningen

The Measuring Behavior Conferences

Measuring Behavior is a unique event. The mission of the conference is to present innovations and share ideas and experiences about methods, techniques and tools for the study of human or animal behavior, independent of the species being studied. While most conferences focus on a specific domain, Measuring Behavior creates bridges between disciplines by bringing together people who may otherwise be unlikely to meet each other. At a Measuring Behavior meeting, you will find yourself among ethologists, behavioral ecologists, neuroscientists, developmental psychologists, human factors researchers, movement scientists, psychiatrists, usability testers and others! While the research questions and applications may be highly diverse, all delegates share an interest in methods, techniques and tools for studying behavior. Experience tells us that the focus on methodological and technical themes can lead to a very productive cross-fertilization between research fields.

The first meeting, Measuring Behavior ’96, chaired by Berry Spruijt, was a spin-off from the European project “Automatic Recording and Analysis of Behavior”. The plan to share the results of our project with colleagues quickly evolved into an international event. Organized by Noldus Information Technology and hosted by Utrecht University, Measuring Behavior ’96 attracted 153 participants from 25 countries. The 2-day program included 70 presentations and 4 scientific tours. Menno Kruk wrote a report of the meeting, which was published in *Trends in Neurosciences* (vol. 20, pp. 187-189, 1997).

The second conference, Measuring Behavior ‘98, under chairmanship of Jaap Koolhaas, brought more than 275 delegates from 32 countries together at the campus of the University of Groningen. During three full conference days, there were 140 presentations grouped into 14 thematic symposia, well balanced between human and animal research. The program included ample time for posters and demonstrations of software or equipment by participants. We had 6 lab tours, 20 set-ups for ongoing technical training and two companies organized user meetings. A ‘video digitization service’ allowed delegates to take a look into the world of digital video, which was new for most at that time. Finally, some 20 companies exhibited scientific books, instruments and software. For the first time, selected presentations were published as full papers in the journal *Behavior Research Methods, Instruments & Computers*.

Measuring Behavior 2000 was held in Nijmegen and hosted by Alexander Cools, attracting over 300 delegates from around the world, who attended more than 160 oral and poster presentations. New to the program at Measuring Behavior 2000 were special interest groups and workshops. Once again, different companies exhibited a wide range of research instruments and software, and delegates could visit different labs on scientific tours, receive ongoing technical training and attend user meetings organized by Noldus Information Technology.

We saw a further increase in attendance when Measuring Behavior 2002 was held at the Vrije Universiteit Amsterdam with Gerrit van der Veer serving as program chair. 325 delegates from 37 countries took part in a busy program, which included - for the first time - tutorials, short courses taught by expert instructors. Twelve of these were organized and received high ratings.

And now you find yourself at the fifth Measuring Behavior conference in Wageningen. Measuring Behavior 2005 is special in several ways. First of all, we have a record number of participants, as many as 450! The interest shown in the conference has been truly worldwide, with participants traveling to the Netherlands from 40 different countries. We also have a record number of presentations, with almost 300 oral papers, posters and demonstrations. This includes 3 renowned keynote speakers, 11 well-prepared symposia and 5 special interest groups. The conference publications have been expanded with a Conference CD and printed Conference Proceedings, containing short papers of most presentations. Finally, for Noldus Information Technology as the organizing company, this conference marks the opening of our new building, where we proudly receive all delegates for the welcome reception.

The local organizing committee has done its best to prepare an optimal mix of scientific, technical, social and culinary ingredients. We hope that you find Measuring Behavior 2005 a rewarding experience and wish you a pleasant stay in Wageningen.

Lucas P.J.J. Noldus
Chair, Local Organizing Committee
Managing Director, Noldus Information Technology bv

Acknowledgements

Scientific Program Committee

- Louise Vet (Program Chair), Netherlands Institute of Ecology (NIOO-KNAW), Nieuwersluis, and Laboratory of Entomology, Wageningen University, Wageningen, The Netherlands
- Ring Carde, Department of Entomology, University of California, Riverside, CA, U.S.A.
- Jerome Casas, Institut de Recherche sur la Biologie de l'Insecte, University of Tours, Tours, France
- Alexander Cools, Department of Psychoneuropharmacology, Radboud University Nijmegen, Nijmegen, The Netherlands
- Wim Crusio, Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA, U.S.A.
- Marcel Dicke, Laboratory of Entomology, Wageningen University, Wageningen, The Netherlands
- Robert Gerlai, Department of Psychology, University of Toronto, Mississauga, Ontario, Canada
- Eco de Geus, Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands
- Stan Gielen, Department of Medical Physics and Biophysics, Radboud University Nijmegen, Nijmegen, The Netherlands
- Ilan Golani, Department of Zoology, Tel Aviv University, Tel Aviv, Israel
- Jaap Harlaar, Department of Rehabilitation, Vrije Universiteit Medical Center, Amsterdam, The Netherlands
- Frans van der Helm, Man-Machine Systems & Control group, Department of Mechanical Engineering, Delft University of Technology, Delft, The Netherlands
- Piet Hoogeboom,Training, Human Factors and Cockpit Operations, National Aerospace Laboratory, Amsterdam, The Netherlands
- John Krantz, Department of Psychology, Hanover College, Hanover, IN, U.S.A.
- Menno Kruk, Department Medical Pharmacology, Leiden/Amsterdam Center for Drug Research (LACDR), Leiden, The Netherlands
- Johan van Leeuwen, Experimental Zoology Group, Wageningen University, Wageningen, The Netherlands
- Gitte Lindgaard, Department of Psychology, Carleton University, Ottawa, Canada
- Josina Mojet, Agrotechnology & Food Innovations, Wageningen University, Wageningen, The Netherlands
- Pertti Panula, Neuroscience Center, Institute of Biomedicine/Anatomy, University of Helsinki, Helsinki, Finland
- Fabio Paterno, Laboratory on Human Interfaces in Information Systems, ISTI - C.N.R., Pisa, Italy
- Vicenç Quera, Department of Behavioral Science Methods, University of Barcelona, Barcelona, Spain
- Matthias Rauterberg, User-Centered Engineering, Faculty of Technology Management, Eindhoven University of Technology, Eindhoven, The Netherlands
- Berry Spruijt, Department Animal, Science and Society, Utrecht University, Utrecht, The Netherlands
- Franz Josef van der Staay, Animal Sciences Group, Wageningen University & Research Center, Lelystad, The Netherlands
- Hans van Trijp, Marketing and Consumer Behavior Group, Social Sciences Group, Wageningen University, Wageningen, The Netherlands and Unilever Research, Vlaardingen, The Netherlands
- Jonathan Vaughan, Department of Psychology, Hamilton College, Clinton, NY, U.S.A.
- Gerrit van der Veer, Department of Information Management and Software Engineering, Vrije Universiteit, Amsterdam, The Netherlands
- Peter Wittenburg, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Local Organizing Committee

- Lucas Noldus, chair (symposia, special interest groups)
- Joeke van Santen (website, venue, secretariat, general communication, registrations)
- Mechteld Ballintijn (website, logistics, general communication, program and abstracts books)
- René Bourgonje (user meetings)
- Cécile Bruisten (program and abstracts books)
- Ruud de Keijzer (conference CD)
- Eveline Mulder (social events, scientific tours, exhibition, merchandising, website)
- Andrew Spink (tutorials)

Scientific Tours

- Bart van de Bunt, Numico Research
- Marcel Dicke, Laboratory of Entomology, Wageningen University
- Jos van Doorn, MARIN
- Tobias Heffelaar, Noldus Information Technology bv
- Jorrit Kuipers, Green Dino Virtual Realities
- Johan van Leeuwen, Experimental Zoology Group, Wageningen University
- Astrid Poelman, Agrotechnology & Food Innovations bv
- Froukje Rienks, NIOO-KNAW

Proceedings Editors

- Fabrizio Grieco
- Leanne Loijens
- Lucas Noldus
- Patrick Zimmerman

Conference Secretariat

- Marjoleine Bessels
- Melvin de Bruijn
- Yvonne Leander

Conference Publications

Measuring Behavior 2005 resulted in the following conference publications:

Program book

The program book contains a complete overview of all conference events, schedules of sessions and presentations, social events, information about exhibitors, conference organization, and practical information.

Abstracts book

The abstracts book contains the summaries of all 296 conference presentations, listed alphabetically by first author.

Conference CD

Bundled with the abstracts book is the conference CD. This disk, published for the first time for Measuring Behavior 2005, contains 220 papers presented at the conference. A search function is provided, allowing you to find papers by author, co-author, title or keyword. Several papers on the conference CD are accompanied by digital video clips.

Printed proceedings

This book, published for the first time in conjunction with Measuring Behavior 2005, contains 230 papers presented at the conference, including those that reached the editors after the conference and did not make it to the conference CD. Because most journals allow little space for detailed descriptions and illustrations of methods, tools, and setups, the Measuring Behavior proceedings offer a unique collection of papers with a methodological and technical emphasis. All papers focus on the details of methods and techniques, rather than a discussion of scientific hypotheses, data and results. This makes it a unique volume for your professional library.

Special issue of *Behavior Research Methods*

The conference organization has an agreement with *Behavior Research Methods* (BRM), a peer-reviewed journal (ISSN 0743-3808) published by the American Psychonomic Society (www.psychonomic.org). This journal (previously called *Behavior Research Methods, Instruments & Computers*) devotes a special issue to the Measuring Behavior conference. Copies of the special issues can be obtained directly from Psychonomic Society Publications at www.psychonomic.org/brmic/special.htm. The links below take you to the table of contents of the special issues that have been published so far:

- Measuring Behavior '98: www.noldus.com/events/mb98/brmic.htm
- Measuring Behavior 2000: www.noldus.com/events/mb2000/brmic.html
- Measuring Behavior 2002: www.noldus.com/events/mb2002/brmic.html

Conference website

After each conference, the Measuring Behavior website is converted into an archival site, with abstracts of all presentations, which remain accessible. The websites of past conferences form a valuable resource on methods and techniques for behavioral research. These are links to the websites of all five conferences:

- Measuring Behavior '96: www.noldus.com/events/mb96/mb96.htm
- Measuring Behavior '98: www.noldus.com/events/mb98/mb98.htm
- Measuring Behavior 2000: www.noldus.com/events/mb2000/
- Measuring Behavior 2002: www.noldus.com/events/mb2002/
- Measuring Behavior 2005: www.noldus.com/events/mb2005/

Keynote speaker

Richard Morris

Division of Neuroscience, Edinburgh University, Edinburgh, United Kingdom

About the speaker

Richard Morris is an elected Fellow of the Royal Society of Edinburgh, of the Royal Society in London, and the Academy of Medical Sciences. He was recently elected to the American Academy of Arts and Science. He has won several awards for his research, notably the Zotterman Medal of the Swedish Physiological Society in 1999 given at the Nobel Forum in Stockholm. He is also active in aspects of science administration, including a period as Chair of the British Neuroscience Association, and current membership of the Advisory Boards of a Research Centre in Tokyo, a Max Planck Institute in Munich and of the Picower Center for Learning and Memory at M.I.T. in Cambridge, USA.

Moving on from spatial learning to episodic-like and semantic-like memory

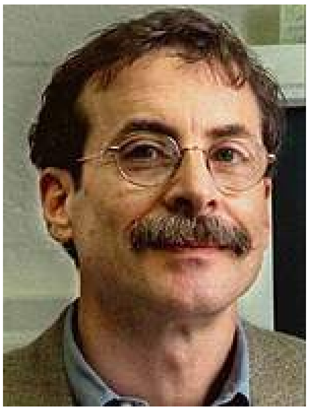
Spatial learning - in T-mazes, radial-mazes and the water maze - has long been a popular choice amongst behavioral neuroscientists interested in the neurobiology of learning and memory. Drawing largely on studies using the water maze, I shall describe classical findings that have emerged over the past 20 years, ranging from standardized spatial reference memory protocols through to delayed-matching-to-place and other procedures expressly designed to address specific theoretical issues relating to hippocampal function, the role of synaptic plasticity in memory, and animal models of neurodegenerative disease. A key theme is that the water maze, no less than types of apparatus, is merely an apparatus; it is important to maintain a sharp distinction between the water maze as an apparatus and the water maze as a set of distinct training protocols. The value and pitfalls of different procedures will be highlighted.

Increasingly, students of the neurobiology of learning and memory are interested in investigating wider issues, notably how spatial and contextual memory can provide a framework for remembering events. My research group has recently developed a new paired-associate paradigm for investigating episodic-like and semantic-like memory in animals. We call the apparatus the ‘Event Arena’ and within it, rats are trained to find a specific flavor of food in a particular location. Learning may take place over 1-trial (episodic-like), or over several trials and days (semantic-like). The characteristics of the two styles of training will be described, together with illustrations of the flexibility of this new appetitive paradigm for investigating a wider range of issues, such as the puzzle of system-level memory consolidation. Certain implications of the apparatus for the development of new automated systems for tracking animals and for automatically recognizing their behavioral actions will also be discussed.

Jeffrey F. Cohn

Division of Psychology, University of Pittsburgh, Pittsburgh, PA, U.S.A.
and the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, U.S.A.

About the speaker



Jeffrey Cohn is Professor of Psychology at the University of Pittsburgh and Adjunct Faculty at the Robotics Institute, Carnegie Mellon University. He earned his PhD in Clinical Psychology from the University of Massachusetts in Amherst and completed Clinical Internship at the University of Maryland Medical Center. For the past 20 years, he has conducted investigations in the theory and science of emotion, depression, and nonverbal communication. He has co-led interdisciplinary and inter-institutional efforts to develop advanced methods of automatic analysis of facial expression and prosody and applied these tools to research in human emotion, communication, biomedicine, biometrics, and human-computer interaction. He has published over 90 papers on these topics. His research has been supported by grants from the National Institute of Mental Health, the National Institute of Child Health and Human Development, the National Science Foundation, the Office of Naval Research, and the Defense Advanced Research Projects Agency.

Automatic facial image analysis

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions. The face can express emotion sooner than people verbalize or even realize their feelings. To make optimal use of the information afforded by facial expression, reliable, valid and efficient methods of measurement are critical.

Two major advances toward this goal were the human-observer-based Facial Action Coding System (FACS) and facial electromyography (EMG). The third major advance - automatic facial image analysis - combines the best features of FACS and EMG: Comprehensive description of facial expression (FACS) and quantitative, automatic measurement, which was previously possible only by using invasive facial sensors (EMG).

This talk reviews previous approaches to facial measurement (FACS and facial EMG) and the advances in automatic facial image analysis they inform. It reviews technical and conceptual challenges that lay ahead and the approaches of leading investigators. It describes capabilities and limitations of current systems, applications in emotion and social interaction, and what we have already learned using these new systems about the configuration and timing of facial expression.

Keynote speaker

Sergio Velastin

Digital Imaging Research Centre, School of Computing & Information Systems, Kingston University, Kingston-upon-Thames, United Kingdom

About the speaker

Dr. Sergio A Velastin obtained his doctoral degree from the University of Manchester (UK) for research on vision systems for pedestrian and road-traffic analysis. He then worked in industrial R&D and project management before joining the Dept. of Electronic Engineering in Kings College London (University of London). In 2001, he and his team joined the Digital Imaging Research Centre in Kingston University, attracted by its size and growing reputation in the field, where he is currently a Reader. He was Technical Coordinator of the EU-funded project PRISMATICA working on the integration of technology (networking, video/audio processing, wireless transmission) and human-based processes for improving personal security in public transport systems. His research interests include computer vision for pedestrian monitoring and personal security as well as distributed visual surveillance systems. Dr. Velastin is a member of the IEE, IEEE and the British Machine Vision Association (BMVA).

Intelligent CCTV surveillance: Advances and limitations

The installation of closed circuit television (CCTV) cameras in urban environments is now commonplace and well-known. The UK leads the world with an estimated 4 million public cameras installed (20% of the world's deployment). Public attitudes to these systems reflect the balance needed between two conflicting requirements: (a) Concerns over invasion of privacy and fears of authoritarian control of the population, and (b) Welcoming the increased safety in public spaces and reductions in crime and antisocial behavior. Recent events and what appears to be the effectiveness of the CCTV infrastructure in assisting law enforcers to understand how the events took place and the people involved, seem to have tipped the balance towards (b), at least momentarily. What we should not forget however, is that such events are thankfully very rare but that there is a cumulative significant effect of mundane daily events that we need to deal with. For example, it has been estimated that what in the UK is called "antisocial behavior" costs the country around € 5000 million a year. A single London Borough (municipality) spends annually around € 1 million to remove graffiti, 44% of women feel unsafe at bus stops at night, a single bus company in a major city is known to have replaced 8,000 windows in one year and a study showed that in a single day in the UK there were around 66,000 reports of nuisance or loutish behavior. At the same time, there are reports that when CCTV has been installed there has been a 35% reduction in crime over 5 years. Indeed, crime in general is decreasing while uncivil behavior is on the increase. It turns out that one of the major limitations of conventional CCTV systems is the impracticality of deploying sufficient number of people to be in front of television screens observing largely uneventful video. As long as this is the case, CCTV will tend to remain a reactive tool. The inability of being truly pro-active, producing timely alarms and eventually being able to prevent incidents, is what ultimately limits these systems. As a preamble to the main associated symposium, this talk will illustrate some of the efforts that the research and industrial communities are making towards realizing automated means of detecting video events involving human activity. It will show the kind of progress made but also the current limitations of this technology.

Automatic facial expression analysis and synthesis

M. Pantic
Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

With an ever-increasing role of computers and other digital devices in our society, one of the main foci of the research in Artificial Intelligence is on Impacts of Emerging Human-Machine Systems. A related, crucial issue is that of Human-Machine Interaction (HMI). A long-term goal in HMI research is to approach the naturalness of human-human interaction. This means integrating "natural" means that humans employ to interact with each other into the HMI designs. With this motivation, automatic speech recognition and synthesis have been the topics of research for decades. Recently, other human interactive modalities such as body and facial gestures have also gained interest as potential modes of HMI. The objective of this symposium is to review recent advances in automatic facial expression analysis and synthesis and their potential applicability to natural HMI.

Facial expression is one of the most cogent, naturally preeminent means for human beings to communicate emotions, to clarify and stress what is said, to signal comprehension, disagreement, and intentions, in brief, to regulate interactions with the environment and other persons in the vicinity. Automated analyzers and synthesizers of facial expressions have, therefore, numerous applications in behavioral science, medicine, security, and HMI. They form the essence of automated tools for lip reading, bimodal speech analysis, videoconferencing, face and visual speech animation, virtual character animation, affective computing, etc. In the robotics area, the facial expression is also very important for natural/comfortable man-machine interaction. KISMET (AI Lab., MIT) and Face Robot (Science University of Tokyo) are good examples.

Although several promising approaches have been reported in the literature in the last ten years, automatic detection, interpretation and synthesis of human facial expressions remain rather difficult to achieve. This Symposium on *Automatic facial expression analysis and synthesis* represents an internationally respected scientific forum for presenting and discussing different approaches to automatic facial expression detection, analysis, synthesis, and their applications to the field of HMI.

Towards a real-time and distributed system for face detection, pose estimation and face-related features

J. Nesvadba, A. Hanjalic, P. Fonseca, B. Kroon, H. Celik and E. Hendriks

This paper discusses the challenges and possibilities of automatic detection and analysis of human faces in the scene. The evolution of storage capacity, computation power and connectivity in Consumer-Electronics(CE)-, in-vehicle-, medical-IT- and on-chip- networks allow the implementation of grid-computing-based real-time and distributed face-related analysis systems. A combination of facial-related analysis components – Service Units (SUs) – such as omni-directional face detection, pose estimation, face tracking and facial feature localization provide a necessary set of basic visual descriptors required for advanced facial- and human-related feature analysis SUs, such as face recognition and facial-based mood interpretation. Smart reuse of the available computational resources across individual CE devices or in-vehicle- and medical-IT- networks in combination with descriptor databases facilitate the establishment of a powerful analytical system applicable for various domains and applications. The SUs for advance analysis of human faces developed at Philips Research Laboratories (NatLab) in cooperation with Delft University of Technology are described.

Learning spatio-temporal models of facial expressions

M. Pantic, I. Patras and M.F. Valstar

This paper explains first the benefits of a robust facial expression analyzer for fields as diverse as psychology, medicine, security, education, and HMI. Then, it reports on two different methods aimed at automatic recognition of facial muscle activations (i.e., facial action units, AUs) from nearly frontal-view face video. The first proposed method constructs temporal templates from the input face video and applies a two-stage learning machine, combining a kNN algorithm and rule-based reasoning, to recognize shown individual AUs and AU combinations. The second method exploits particle filtering to track facial fiducial points such as the mouth and eyebrow corners in the input face video and applies temporal rules to recognize AUs and their temporal segments (i.e., onset, apex, offset) occurring alone or in combination in the input image sequence. Finally, a case-based reasoning system is discussed, which is capable of classifying facial expressions (given in terms of AUs) into the emotion categories learned from the user. The utilized case base is a dynamic, incrementally self-organizing event-content-addressable memory that allows fact retrieval and evaluation of encountered events based upon the user preferences and the generalizations formed from prior input.

The CMU/Pitt automated facial image analysis System

T. Kanade and J.F. Cohn

Both the configuration and the timing of facial actions are important in emotion expression and recognition. To investigate the timing and configuration of facial actions, the CMU/Pitt Automated Facial Image Analysis (AFA) System has been developed. The latest version of the system is based on Active Appearance Models (AAMs). These are generative, parametric models and consist of a shape component and an appearance component. The shape component of is a triangulated mesh that moves like a face undergoing both rigid motion (head pose variation) and non-rigid motion (expression) in response to changes in the parameters. The appearance component of the AAM is an image of the face, which itself can vary under the control of the parameters. As the parameters are varied, the appearance varies so as to model effects such as the emergence of furrows and wrinkles and the visibility of the teeth as the mouth opens. Two disadvantages of traditional AAMs are that they are 2D and rigid head motion and non-rigid facial motion are confounded in the shape model. To address these problems, we recently developed an extension to AAMs that augments the usual 2D shape model with a 3D shape model. This advance separates the 3D rigid motion of the head and 3D non-rigid facial expression into two disjoint sets of parameters and recovers the 3D shape of the face. This approach works well as long as out-of-plane head motion is small to moderate. As out-of-plane head motion becomes large, automatic recovery of 3D shape becomes increasingly difficult because of self-occlusion. To solve this problem, a single AAM is fitted to multiple images captured simultaneously from synchronized cameras.

Bimodal emotion recognition

N. Sebe, E. Bakker, I. Cohen and T. Huang

Recent technological advances have enabled human users to interact with computers in ways previously unimaginable. Beyond the confines of the keyboard and mouse, new modalities for human-computer interaction such as voice, gesture, and force-feedback are emerging. Despite important advances, one necessary ingredient for natural interaction is still missing - emotions. Emotions play an important role in human-to-human communication and interaction, allowing people to express themselves beyond the verbal domain. The ability to understand human emotions is desirable for the computer in several applications. This paper describes the challenging problem of bimodal emotion recognition and advocates the use of probabilistic graphical models when fusing the different modalities. We test our audio-visual emotion recognition approach on 38 subjects with 11 HCI-related affect states. The experimental results show that the average person-dependent emotion recognition accuracy is greatly improved when both visual and audio information is used in classification.

A robust scheme for facial analysis and expression recognition

S. Ioannou, M. Wallace, K. Karpouzis and S. Kollias

Facial analysis includes a number of processing steps. One of those is the extraction and tracking of the movement of facial components and facial fiducial points. Due to noise, illumination variations and low resolution capturing devices, the detection of facial feature points can be inaccurate. Hence, mechanisms are required that can automatically evaluate the quality of each computed mask, assigning a confidence level to it. The emotion recognition system can take advantage of each feature's confidence level when analyzing them. Exploitation of anthropometric knowledge in the form of a set of criteria, evaluating the relation of the extracted features, can form such a mechanism. This paper explains further how the detected facial features can be used to extract the Feature Points considered in the definition of the Facial Animation Parameters (FAPs). Finally, we discuss techniques such as clustering and neurofuzzy methods that can transfer variations of the FAP variables into rules for recognition of the user's emotional state and then adapt these rules to specific user's characteristics.

Emotion and facial expressions in creating embodied agents

T.D. Bui, D. Heylen, A. Nijholt and M. Poel

This paper describes the work done on Obie, an embodied conversational agent framework. An embodied conversational agent, or talking head, consists of various components. The authors have created a face model and a facial muscle model in such a way that realistic facial expressions can be produced in real-time on a standard PC. In particular, they have defined a face model that allows high quality and realistic facial expressions, which is still sufficiently simple in order to keep the animation real-time and is able to assist the muscle model to control the deformations. They have also implemented a muscle model that produces realistic deformation of the facial surface, handles multiple muscle interaction correctly and produces bulges and wrinkles in real-time. Besides this graphical part, the developed system accounts for the actions (dialogue) and emotions of the agent. Namely, the authors have implemented an emotion model and a mapping from emotions to facial expressions. For the animation, it is particularly important to deal with the problem of combining different facial movements temporally. In this respect, the dynamic aspects of facial movements and the combination of facial expressions in different channels that are responsible for different tasks received special attention.

Fast facial animation design for emotional virtual humans

S. Garchery, A. Egges and N. Magnenat-Thalmann

Designing facial animation parameters according to a specific model can be time consuming. This paper presents a fast approach to design facial animations based on minimal information (only feature points). All facial deformations are automatically computed from MPEG-4 feature points. An extension of this approach that allows to personalize or to customize the deformations according to different characteristics and then with minimal manual interaction is also presented. Different prototypes of the facial animation system, available on different platforms, are described. How emotions and expression can be incorporated into the facial animation system is demonstrated as well. Finally, different approaches to emotions and personality are presented.

Towards a real-time and distributed system for face detection, pose estimation and face-related features

J. Nesvadba[1], A. Hanjalic[2], P. M. Fonseca[1], B. Kroon[1/2], H. Celik[1/2], E. Hendriks[2]
[1] *Philips Research, Eindhoven, The Netherlands*
[2] *Delft University of Technology, Delft, The Netherlands*

Abstract

The evolution of storage capacity, computation power and connectivity in *Consumer-Electronics*(CE)-, in-vehicle-, medical-IT- and on-chip-networks allow the easy implementation of grid-computing-based real-time and distributed face-related analysis systems. A combination of facial-related analysis components - *Service Units* (SUs) – such as face detection, pose estimation, face tracking and facial feature localization provide a necessary set of basic visual descriptors required for advanced facial- and human-related feature analysis SUs, such as face recognition and facial-based mood interpretation. Smart reuse of the available computational resources across individual CE devices or across in-vehicle- or medical-IT-networks in combination with descriptor databases facilitate the establishment of a powerful analytical system applicable for various domains and applications.

Keywords

Face detection, pose estimation, face tracking, content management.

1 Introduction

Through the fast evolution of processing power, storage capacity and connectivity 1 in CE-, in-vehicle- and medical-IT-networks, generic *Multimedia-Content-Analysis-* (MCA-) and computer-vision-based analysis solutions start to reach human brain's semantic levels. Powered by smart usage of scattered processing power, storage and bandwidth available across those networks, realization of real-time high-level semantic analysis systems do not belong to the realm of fiction any more. Multiple cross-domain and cross-organizational collaborations 2, combinations of state-of-the-art network and grid-computing solutions, and usage of recently standardized interfaces facilitated the set-up of an advanced analytical system, further referenced to as *CASSANDRA Framework* (CF) 3. This prototyping framework enables distributed computing scenario simulations for e.g. *Distributed Content Analysis* (DCA) across CE In-Home networks, but also the rapid development and assessment of complex multi-MCA-algorithm-based applications and system solutions. Furthermore, the modular nature of the framework - logical MCA and computer vision components are wrapped into so-called *Service Units* (SU) - eases the split between system-architecture- and algorithmic-related work and additionally facilitate reusability, extensibility and upgradeability of those SUs. Additionally, the modularization allows smart network management systems to balance the processing load across the available resources in applicable networks (e.g., CE In-Home networks). Such an elaborated DCA system can be seen as basis for *Ambient Intelligence* (AmI) applicable in various domains, such as CE, medical IT, car infotainment and personal healthcare.

In many of these application domains, one of the most important elements is the human face. Therefore, indication of its location, its identity and even its expression provide useful semantic information. For this reason, one of the most prominent AmI-related problems is the availability of a reliable real-time face-analysis system. Consequently, various face-related SUs have been or are being jointly researched 2, implemented and integrated into the CF, further described in this paper. These comprise SUs such as omni-directional face detection, face tracking, face recognition, face online learning, facial features- and facial points-analysis. In combination, these SUs provide the basic visual descriptors for advanced facial- and human-related feature analysis and applications.

2 Distributed Face Analysis System

The realization of a real-time distributed face analysis system requires modularization of face analysis algorithms and standardization of face-related descriptors, which is the basic concept of the CF. In 1, the first attempt of such a modularization is described for the specific case of a face recognition system; this system includes the required underlying SUs *Face Detection* (SU FD) and *Face Tracking* (SU FT). CF-based evaluations highlighted the limiting capabilities of the implemented face detectors 1 in providing the necessary information for reliable face recognition. Consequently, new face detection algorithms are currently being researched that shall be able not only to localize faces regardless of their spatial orientation but also to achieve higher overall detection performances. Furthermore, these new algorithms will allow the implementation of mid-level SUs such as SU *Pose Estimation* (SU PE) (see ***Figure 1***) - providing indication of the spatial orientation information of localized faces; additionally, SU *Facial Features* (SU FF) will determine position of ears, nose, eyes, etc. All collected facial data is thereafter used as input for SUs *Face Recognition* (SU FR), *Online Face Clustering* (SU OFC), *Facial Feature Points (SU FFP)* and *Facial Expression* (SU FE; emotion/mood interpretation) analysis, which are currently also under investigation. ***Figure 1*** illustrates the relation between such SUs.

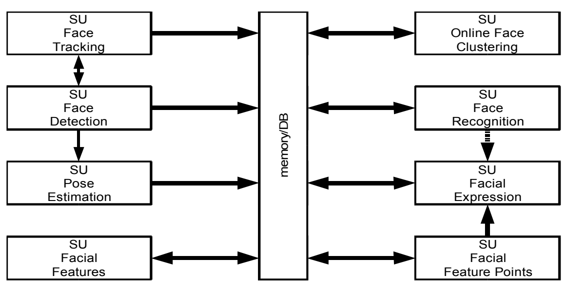


Figure 1. *Face-analysis-related SUs.*

2.1 Existing Face Detection Algorithms

To extract face-related features like pose, gaze direction, identity, facial expression and mood, face detection is an essential step. With this in mind, face detection has been and still is extensively researched. One of the various face detection algorithms, a low complexity color-based method, performs detection in the compressed domain. This method is unequaled in computational efficiency but is not capable of handling monochrome video and due to its extreme low-complexity, it only performs satisfactorily under controlled conditions. To overcome these disadvantages, another algorithm was developed based on the Viola Jones-based learning method. However, this method has the shortcoming of only being able to detect upright frontal faces. Both algorithms are briefly described in the next sections.

2.1.1 Compressed Domain Face Detection

The compressed domain face detection algorithm 4 uses a feature-based approach to determine the presence and location of multiple frontal faces using only DCT coefficients extracted from compressed content (images). Face detection is accomplished by first performing skin color segmentation based on a model built from the statistical color properties of a large set of manually segmented faces. After applying binary morphological operators on the segmented image, specific subsets of the input AC coefficients are used, along with the brightness properties of the input image to determine in SU FF the location of specific facial features (eyes, eyebrows and mouth). Finally, using a model of typical frontal faces, face candidates are generated based on the location of these facial features. Face candidates are then ranked according to their size, their percentage of skin color pixels and the intensity of their facial features. Finally, the most relevant face candidate is chosen for each individual skin color region.

As illustrated in *Figure 2*, even though the face detector is intended for detection of frontal faces, it is also able to correctly determine the location of faces that are rotated and tilted up to a certain limit.

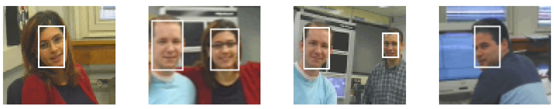


Figure 2. *Examples of correctly detected faces.*

2.1.2 Viola-Jones-based Face Detection

Besides the compressed-domain-based face detector described in the previous section, a Viola-Jones based face detection algorithm 55 was implemented for evaluation purposes. This image-based detection algorithm works on uncompressed images and has proven to be robust under various lighting conditions. The method is based on a cascade of boosted classifiers of simple Haar-wavelet like features on different scales and positions. The features are brightness- and contrast-invariant and consist of two or more rectangular region pixel-sums that can be efficiently calculated by the Canny integral image. The feature set is overcomplete and an adaptation of the AdaBoost learning algorithm is proposed to select and combine features into a linear classifier. To speed up detection a cascade of classifiers is used such that every classifier can reject an image. All classifiers are trained to reject part of the candidates such that on average only a low amount of features are used per position and scale.

After all possible face candidates are obtained, a grouping algorithm reduces groups of face candidates into single positive detections.

This detection method has been mapped to a smart camera 78. The smart camera detects multiple frontal faces of different sizes in images and allows small rotations (± 10°). The face detection application is running at a rate of 4 frames per second.

2.2 Current Face Detection Research

As explained, the methods discussed in paragraph 2.1 are sensitive to color conditions and face pose. Current research addresses these limitations in an attempt to develop algorithms that allow the extraction of face-related features in uncontrolled scenarios regardless of pose and illumination conditions. The main difficulty in developing an omnidirectional face detector is related to the fact that the 2-D visual appearance of an object depends on its pose. To distinguish a face from other objects regardless of its pose, either a set of pose-dependent detectors operating in parallel, a complex "brute force" learning method, or a 3-D model fitting technique is required. For the first kind of detectors (parallel pose-dependent detectors), the in-plane and out-of-plane pose range (i.e.: rotation axis perpendicular or parallel to the image viewing plane respectively) is partitioned into a number of areas for which an independent detector is designed - this kind of omnidirectional face detector is called a multiview detector.

In the following sections, examples of face detectors that use the first two of these techniques are analyzed and their applicability for robust and real-time omnidirectional face detection in video content is discussed.

2.2.1 The Schneiderman-Kanade Method

In 9, Schneiderman and Kanade describe an object detection method applied to face detection. The proposed algorithm was one of the first efficient face detectors in literature that could determine the location of non-upright frontal faces. Besides being able to attain multiview face detection, it copes with variations in pose by using two specific classifiers trained separately: one for detection of frontal faces and another for detection of profile faces. The profile detector is trained for right profile view points and applying it on the vertical mirrored image allows for left profile face detection. As a result, faces with in-plane rotation between -15º and +15º and full-profile faces (-90º to +90º rotation out of plane) can be detected. For each view-point (profile, right-frontal and left-profile), the corresponding detector scans the original image and its downscaled versions at several locations. Images are analyzed with windows of size 48 × 56 for the frontal detector and 64 × 64 for the profile detector. The decision is based on a Bayesian classifier on joint values and positions of visual *attributes*. An *attribute* is here defined as a group of quantized wavelet coefficients in given sub-bands. In total, 17 different attributes are involved, a detailed description of which can be found in 9. Attributes are sampled at regular intervals over the detection window (coarse resolution).

2.2.2 Viola-Jones-based Methods

In Section 2.1.2, a Viola-Jones frontal face detector was presented. In this section extension of that method for omnidirectional detection is discussed.

Omnidirectional face detection could be achieved simply by training a Viola-Jones detector with face images of all poses. However, this would imply that a huge number of selected features would be needed in order to incorporate all different face appearances. Naturally, the complexity of the algorithm would become unbearable, especially for real-time implementations. In order to avoid this problem, a multiview Viola-Jones detector – i.e., in which a single detector is designed for each pose range – may be developed. It may be achieved according to one of the two following strategies: all detectors could perform classification in parallel or a single selector could be used for detection using the information of a pre-processing pose estimator, i.e. SU PE. Both approaches are described in existing literature.

In 10, Viola and Jones propose to train a C4.5 decision tree on 12 poses, 10 levels deep without pruning. The paper covers both in-plane and out-of-plane rotation, but does not present a complete solution. It is argued by the authors that a pose estimator would have approximately the complexity of one detector, which renders the method only twice as intensive as a frontal detector. The pose estimator/single classifier approach should thus be faster than the parallel classifiers approach. For this reason, a potentially robust real-time multiview Viola-Jones-based classifier system employing different kinds of base classifiers is envisioned. The classifiers in this system can be divided into two groups:

1. A pose estimator can quantize poses in order to reduce the classification problem for other classifiers. The pose estimator can be used on all image positions and scales prior to detection such that for non-face areas the pose will be arbitrary.
2. A pose-specific face detector classifies between face and non-face; detectors can be cascaded and of multiple types; simple detectors are used to quickly reduce false alarms without sacrificing recall, while more complex (and slower) detectors may be used to increase precision by validating remaining face candidates.

It may be observed that the original Viola-Jones detector is actually a cascade of classifiers; thus, an omnidirectional face detector may be actually built from a large tree structure of simple classifiers. Current research work may thus be regarded as an attempt to identify and design the optimal structure of such a system.

2.2.3 The Convolutional Face Finder

The third face detector, a *Convolutional Face Finder* (CFF) 11, is based on a multi-layer *Convolutional Neural Network* (CNN). CNNs were originally intended and designed for handwritten digit recognition.

It is designed for faces rotated between -20º and +20º in-plane, and between -60º and +60º out-of-plane and relies, unlike previous methods, only on a single detector.

The CFF consists of six successive neural layers. The first four layers extract characteristic features, and the last two perform the actual classification (face/non-face). The CFF is applied on several resized instances of the original image at several positions. The input of the system is a 32 × 36 window extracted from each rescaled image. The first step consists of convolving this input with 5 × 5 kernels and adding a bias; 4 kernel variants are applied, resulting in 4 different *feature maps*. The produced feature maps are then down-sampled by a factor of two, multiplied by a weight, and corrected by a bias before a sigmoid activation function is applied. Subsequently, this convolution/sub-sampling scheme is repeated with 3 × 3 masks resulting in 14 new feature maps which consist on the characteristic features extracted for classification. The last two layers, comprised of traditional neural processing units decide on the presence of a face.

This face detector is an example of a monolithic "brute force" approach for the problem of omnidirectional face detection.

2.2.4 Comparative Analysis of the Methods

It is important to note that the abovementioned methods achieve omnidirectional face detection only for a limited range of in-plane and out-of-plane rotations. Upside-down oriented faces, for instance will likely not be detected. To achieve true omnidirectionality, multiple detectors have to be combined.

As explained earlier, the objective of this research work is twofold: while the aim is to efficiently detect faces regardless of their pose, this should be achieved on video content in real-time with a reasonable amount of processing power.

The Schneiderman-Kanade detector achieves high detection rates (above 90% on the CMU frontal set); it performs especially well on difficult profile face images (similar rates on the CMU profile test set) when compared to other multiview systems. The drawback of this approach lies on its computational cost, unacceptable for the purpose at hand, even if the heuristics described in 9 are included. Based on experiments conducted during current research, it was found that processing of each image takes several seconds.

The CFF is able to detect frontal and difficult semi-profile faces with a high detection rate and a very low false alarm rate, without using a specific detector for a given viewpoint or without running a pose estimator. Garcia and Delakis 11 report detection rates on the CMU Frontal set of around 90%, with an execution speed of approximately 4 frames per second for 384 × 288 images on a 1.6GHz P4 processor. Consequently, it appears suitable for our scope in terms of processing speed but this detector does not perform well on full-profile faces, which is a considerable disadvantage.

Finally, the combination of the omni-directional Viola-Jones pose estimator (SU PE) / pose-specific face detector (SU FD), as described in 2.2.2, proved to be the fastest of the methods analyzed. A frontal Viola-Jones FD runs at approximately 15 frames per second on a 3.2GHz P4 processor in images with 720 × 576 resolution, so the combination of pose estimator followed by a face detector is estimated to run roughly at 7 frames per second; current experiments point towards this assumption. ***Figure 3*** compares the detectors on a qualitative speed vs. detection performance plot.

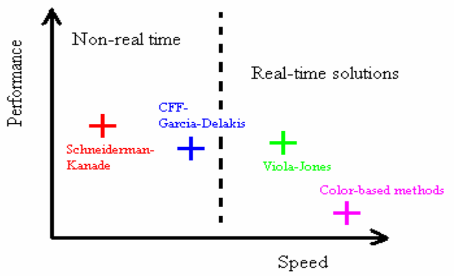


Figure 3. *Qualitative comparison of face detectors.*

Viola-Jones-based detectors (frontal, described in Section 2.2, or omni-directional, described in Section 2.3.2) exhibit the best trade-off between speed and performance. Skin-color based methods, like the compressed-domain method described in Section 2.1, are extremely fast, but have not proven to be sufficiently robust. On the other extreme, the

Schneiderman-Kanade method shows a good detection performance but with a relatively low speed performance.
The Schneiderman-Kanade detector achieves the best performance on full profile face images. The drawback with this approach is that two different detectors trained for different views are used. The image is then scanned three times (once for each profile and one for frontal view), which further slows down the process. An original approach would be to apply this method after a Viola-Jones detector with pose estimation. The use of heuristics such as skin color filtering could also significantly improve the speed performance on color video or image content.
Concerning the CFF, it appears to be very robust, while covering a wide range of views (especially for semi-profiles in the range -60º to +60º). Garcia and Delakis 11 evoke a more complex version with additive feature maps, in order to detect full profiles. Using two CFFs trained for frontal face and full profile could be a sound approach both in terms of execution time and detection performance. Another efficient procedure for Convolutional Neural Networks could be the combination of simultaneous pose estimator and face detector, which would also yield in a real-time system.
Finally, based on our experimentations and results reported in the literature, the conclusion is that an omnidirectional face detector should incorporate a pose estimator and a face detector, instead of consisting in several detectors applied separately on the image, if the objective is to achieve detection and speed performances suitable for the applications the algorithms are intended for.

2.3 SU Pose Estimation

As described in the previous section, pose estimation can be used as a valuable pre-processing step to face detection, being also very useful on its own. The pose of a face can be defined as one in-plane and two out-of-plane angles with a known low tolerance. The description of a face pose may provide useful semantic information. It may be used, for instance, to determine if people are facing one specific direction or if two persons are facing (possibly talking to) each other. This information can also facilitate the determination of facial points since it allows 3-D model fitting with the faces in the images. Pose estimation can thus aid in the determination of facial points of profile and non-upright faces; which in turn can help identifying and analyzing the expression of these faces.

2.4 SU Face Tracking

The previous section discussed several methods to detect faces in still images. However, to view a video as a collection of still images is a considerable naïve approach. Using the temporal dimension of video for object detection may lead to improvement in both localization and speed performances for two reasons, both related to the trivial observation that adjacent video frames are likely to share similar content:

1. False object detections and recognitions and wrong pose estimates may occur in single frames; by combining information from multiple frames, part of the false alarms can be removed and parameter accuracy can be increased without actually increasing computational complexity.
2. In frames that belong to the same shot, faces are unlikely to suddenly appear or disappear and objects do not change dramatically their position or size from frame to frame; this observation allows for a substantial reduction of the search window used for subsequent frames after initial detections have taken place.

Temporal localization of a face may also provide helpful cues for face identification.

3 Conclusions

In this paper, the potential of the Cassandra Framework's modular 3 approach – using SUs for individual services - in combination with face-related content analysis algorithms has been described. The framework provides an easy-to-use prototyping environment enabling the real-time execution of efficient and heterogeneous face-related algorithms, such as omnidirectional face detection, pose estimation and face tracking in a distributed environment. The high modularity of this real-time distributed system will trivially allow the addition of new face-based solutions, such as individual identification, facial expression recognition, or mood estimation. Current research on face detection was also discussed and some conclusions were drawn regarding the direction in which current work will proceed towards a robust efficient omnidirectional face detector.

References

1. J. Nesvadba, P. M. Fonseca, et. al. (2004), *Face Related Features in Consumer Electronic (CE) device environments*, Proc. Int'l Conf. on Systems, Man, and Cybernetics, pp 641-648, The Hague - Netherlands, October 2004.
2. MultimediaN: www.multimedian.nl/
Cassandra:www.research.philips.com/technologies/storage/cassandra/
Candela: www.hitech-projects.com/euprojects/candela/
3. J. Nesvadba, P. Fonseca, et al. (2005), *Real-Time and Distributed AV Content Analysis system for Consumer Electronics Networks*, Proc. IEEE Int'l Conf. for Multimedia and Expo, Amsterdam - The Netherlands, June 2005.
4. P. Fonseca, J. Nesvadba (2004). *Face Detection in the Compressed Domain*, Proc. IEEE Int'l Conf. on Image Processing 2004, pp. 2015-2018, Singapore – Singapore, October 2004.
5. P. Viola, M. Jones (2001). *Rapid Object Detection using a Boosted Cascade of Simple Features*, Proc. IEEE Computer Vision and Pattern Recognition, 2001.
6. R. Lienhart, J. Maydt (2002). *An Extended Set of Haarlike Features for Rapid Object Detection*, Proc. IEEE Int'l Conf. on Image Processing, Vol 1, pp. 900-903.
7. Philips Centre for Industrial Technology (2004). *Inca 311: Smart Firewire Camera with Rolling Shutter Sensor*, http://www.cft.philips.com/industrialvision.
8. R. Kleihorst et al. (2003). *An SIMD Smart Camera Architecture for Real-time Face Recognition*, Abstracts of the SAFE & ProRISC/IEEE Workshops on Semiconductors, Circuits and Systems and Signal Processing, Veldhoven - The Netherlands.
9. H. Schneiderman, T. Kanade (2000), *A statistical method for 3D object detection applied to faces and cars*. International Conference on Computer Vision.
10. M. Jones, P. Viola (2003). *Fast Multi-View Face Detection*, MERL, TR2003-96, July 2003.
11. C. Garcia, M. Delakis (2004). *Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection,* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11.

Learning spatio-temporal models of facial expressions

M. Pantic, I. Patras, M.F. Valstar
Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

Abstract
The human face is used to regulate the conversation by gazing or nodding, to interpret what has been said by lip reading, and to communicate and understand somebody's affective state and intentions on the basis of the shown facial expression. Machine understanding of human facial behavior could revolutionize human-machine interaction technologies and fields as diverse as security, behavioral science, medicine, communication, and education. Yet development of an automated system that detects and interprets human facial signals is rather difficult. This article summarizes our research efforts in meeting this challenge. It presents two systems for machine recognition of facial muscle actions (i.e., Action Units, AUs) in face video and a case-based reasoning system capable of classifying facial expressions (coded in terms of AUs) into the emotion categories learned from the user.

Keywords
Facial expression analysis, temporal templates, particle filtering, case-based reasoning, emotion.

1 Introduction
The human face provides a number of signals essential for interpersonal communication in our social life. The face houses the speech production apparatus and is used to identify other members of the species, to regulate the conversation by gazing or nodding, and to interpret what has been said by lip reading. It is our direct and naturally preeminent means of communicating and understanding somebody's affective state and intentions on the basis of the shown facial expression [5].
Automating the analysis of facial behavior would be highly beneficial for fields as diverse as security, medicine and education. In security contexts, facial expressions play a crucial role in establishing or detracting from credibility. In medicine, facial expressions are the direct means to identify when specific mental processes are occurring. In education, pupils' facial expressions inform the teacher of the need to adjust the instructional message. As far as interfaces between humans and computers (PCs / robots / machines) are concerned, facial expressions provide a way to communicate information about needs and demands to the machine. Where the user is looking (gaze tracking) can be effectively used to free computer users from the classic keyboard and mouse. Also, certain facial signals (e.g. a wink) can be associated with certain commands (e.g. a mouse click) offering an alternative to traditional mouse and keyboard commands. The human ability to read emotions from someone's facial expressions is the basis of facial affect processing that can lead to expanding interfaces with emotional communication and, in turn, to obtaining a more flexible, adaptable, and natural interaction between humans and machines.

2 Facial action coding
Most approaches to automatic facial expression analysis attempt to recognize a small set of prototypic emotional facial expressions, i.e., fear, sadness, disgust, happiness, anger, and surprise (for an exhaustive survey of the past work on this research topic, the reader is referred to [10]). This practice may follow from the work of Darwin and more recently Ekman [5], who suggested that basic emotions have corresponding prototypic expressions. In everyday life, however, such prototypic expressions occur relatively rarely; emotions are displayed more often by subtle changes in one or few discrete facial features, such as raising of the eyebrows in surprise. To detect such subtlety of human emotions and, in general, to make the information conveyed by facial expressions available for usage in various aforementioned applications, automatic recognition of facial muscle actions, such as the action units (AUs) of the FACS system [3], is needed. Facial Action Coding System (FACS) is designed for human observers to describe changes in facial expression in terms of observable facial muscle actions (AUs). FACS provides the rules for visual detection of 44 different AUs and their temporal segments (onset, apex, offset) in a face image sequence. Using these rules, a human coder decomposes a shown facial expression into the specific AUs that produced the expression.
Few approaches have been reported for automatic AU recognition in face images (for an exhaustive survey of the past work on this research topic, the reader is referred to [6]). These include automatic detection of 16 AUs from face video using lip tracking, template matching and neural networks [14], color and motion based detection of 20 AUs occurring alone or in combination in profile-view face video [8], and automatic detection of 18 AUs from face video using Gabor filters, AdaBoost and Support Vector Machines [1]. In contrast to these methods, which address mainly the problem of spatial modeling of facial expressions, the methods proposed in this article address the problem of temporal modeling of facial expressions as well. In other words, the methods proposed here are very suitable for encoding temporal activation patterns (onset → apex → offset) of AUs shown in an input face video.

2.1 AU detection using temporal templates
Figure 1 outlines our method for AU detection in face video using temporal templates. Temporal templates are

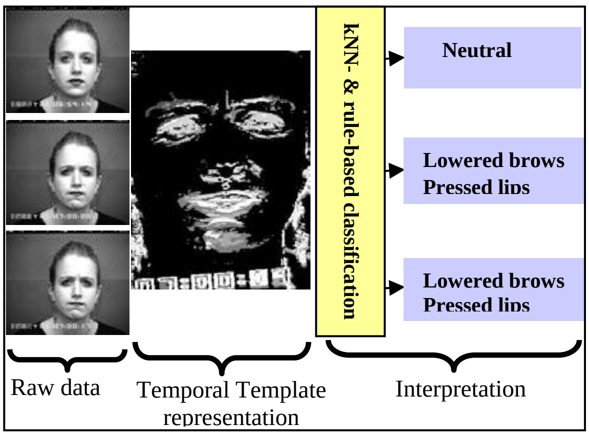


Figure 1. *AU detection using temporal templates*

2D images constructed from image sequences, which show motion history, that is, where and when motion in the input image sequence has occurred [2]. We employ Motion History Images (MHI), which in contrast to Motion Energy Images preserve not only spatial but also the temporal information. More specifically, the value of a pixel in an MHI image indicates where and when motion in the input image sequence has occurred. This value decays over time, so that a high intensity pixel denotes recent motion, a low intensity pixel denotes a motion that occurred earlier in time, and intensity zero denotes no motion at all at that specific location.

Before we can construct a MHI from an input video, the face present in the video needs to be registered in two ways. Intra registration removes all rigid head movements within the input video while the inter registration places the face at a predefined location in the scene. The inter registration process warps the face onto a predefined 'normal' face, eliminating inter-person variation of face shape and facilitating the comparison between the facial expression shown in the input video and template facial expressions. Under the assumption that each input image sequence begins and ends with a neutral facial expression, we downsample the number of frames to a fixed number of (n+1) frames. In this way our system becomes robust to the problem of varying duration of facial expressions.

After the registration and time warping of the input image sequence, the MHI is obtained as follows. Let *I(x, y, t)* be an image sequence of pixel intensities of *k* frames and let *D(x, y, t)* be the binary image that results from pixel intensity change detection, that is by thresholding $\left|I(x,y,t)-I(x,y,t-1)\right| > th$, where *x* and *y* are the spatial coordinates of picture elements and *th* is the minimal intensity difference between two images. In an MHI, say H_t, the pixel intensity is a function of the temporal history of motion at that point with *t* being a frame of the downsampled input video (with (n+1) frames). Using the known parameter *n*, H_t is defined as:

$$H_t(x,y,t) = \begin{cases} s * t & D(x,y,t)=1 \\ H_t(x,y,t-1) & otherwise \end{cases} \qquad (1)$$

where $s = (255/n)$ is the intensity step between two history levels and where $H_t(x,y,t) = 0$ for $t \le 0$. The final MHI, say *H(x, y)*, is found by iteratively computing equation *(1)* for $t = 1 \ldots n+1$.

For automatic detection of AU from MHI-represented face image sequences, we employ a combined kNN/rule-based classifier. The utilized kNN algorithm is straightforward: for a test sample it uses a distance metric to compute which *k* (labeled) training samples are “nearest” to the sample in question and then casts a majority vote on the labels of the nearest neighbors to decide the class of the test sample. Parameters of interest are the distance metric being used and *k*, the number of neighbors to consider. The optimal parameters were experimentally determined to be the simple Euclidian distance measure and *k* = 3 [15]. Although it gives a good indication about the AUs shown in an input video, the kNN algorithm can confuse AUs that have partially the same MHI-representation. To address this drawback, we created a set of rules. With these rules we can correctly reclassify samples that the kNN algorithm misclassifies at first. For instance, the kNN classifier often confuses AU4 and AU1+AU4. Both produce activity in the same part of the MHI, but AU4 causes the eyebrows to move inward and downward, while AU1+AU4 first causes an upward movement followed by an inward and downward movement of the eyebrows. This results in high activation between the brows and relatively low activation above the inner corners of the brows in the case of AU4 activation. Hence, the rules used to resolve the confusion in question have been defined based upon this kind of knowledge about the facial muscle anatomy.

When tested on the Cohn-Kanade Facial Expression Database [4] and the MMI Facial Expression Database [9], the proposed method achieved a recognition rate of 68%, respectively 61%, when detecting 21 AUs occurring alone or in combination in an input face image sequence (for details about this method see [15]).

2.2 AU detection using temporal rules

Figure 2 outlines our method for AU detection in face video using temporal rules. The method processes an input face image sequence in four steps: Face Detection, Facial Fiducial Points Detection, Point Tracking and AU Coding.

To detect the face region in the first frame of an input face video, we adopt a real-time face detector proposed in [1], which represents an adapted version of the original Viola-Jones face detector [16]. The Viola-Jones face detector consists of a cascade of classifiers trained by AdaBoost. Each classifier uses integral image filters, which remind of Haar Basis functions and can be computed very fast at any location and scale. For each stage in the cascade, a subset of features is chosen using a feature selection based on AdaBoost. The adapted version of the Viola-Jones face detector that we employ uses GentleBoost instead of AdaBoost and it uses a smart training procedure in which, after each single feature, the system can decide whether to test another feature or to make a decision. By this the system retains information about the continuous outputs of each feature detector rather than converting to binary decisions at each stage of the cascade.

The detected face region is then divided in 20 relevant Regions of Interest (ROIs), each one corresponding to one facial point to be detected. A combination of heuristic techniques based upon the analysis of the vertical and horizontal image histograms achieves this. The employed facial feature point detection method [17] uses individual feature patch templates to detect points in the relevant ROI. These feature models are 13×13 pixels GentleBoost templates built from both gray level intensities and Gabor wavelet features. In the training phase, the feature models are learned using a representative set of positive and negative examples, where the positive examples are image patches centered on a particular facial feature point and the negative examples are image patches randomly displaced a small distance from the same facial feature. In the testing phase, each ROI is filtered first by the same set

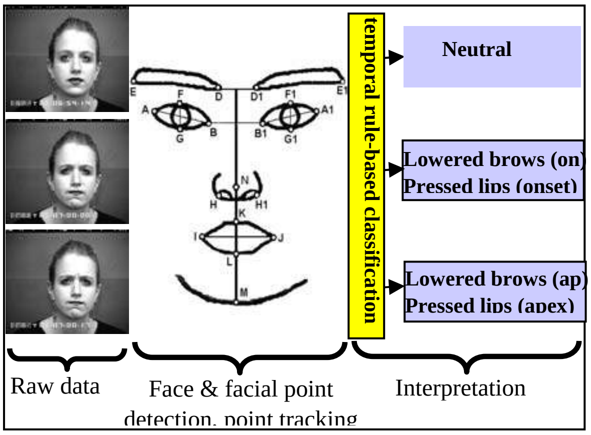


Figure 2. *AU detection using temporal rules*

of Gabor filters used in the training phase (in total, 48 Gabor filters are used). Then, for a certain facial point an input 13×13 pixels window (*sliding window*) is slid pixel by pixel across 49 representations of the relevant ROI (grayscale plus 48 Gabor filter representations). For each position of the sliding window, a GentleBoost classifier outputs a response depicting the similarity between the 49-dimensional representation of the sliding window and the learned feature point model. After scanning the entire ROI, the position with the highest response reveals the feature point in question.

After 20 fiducial points are localized in the first frame of the input face image sequence, windows positioned around each of the facial points, define a number of color templates. Let us denote such a color template with $\boldsymbol{o} = \{\boldsymbol{o_i}\}$ where $\boldsymbol{i}$ is the pixel subscript. We subsequently track each color template for the rest of the image sequence with the auxiliary particle filter that was introduced by Pitt and Shepard [12]. Particle filtering has become a dominant tracking paradigm due to its ability to deal successfully with noise, occlusion and clutter. In order to adapt it for the problem of color-based template tracking, we define an observation model that is based on a robust color-based distance between the color template $o = \{o_i \mid i = 1 \ldots M\}$ and a color template $c = \{c_i \mid i = 1 \ldots M\}$ at the current frame. We attempt to deal with shadows by compensating for the global intensity changes. We use the distance function d, see equation *(2)* below, where M is the number of pixels in each template, m_c (and m_o) is the average intensity of template $c = \{c_i\}$ (and, respectively, of template $o = \{o_i\}$), i is the pixel index and the robust function that we use is the absolute value.

$$d = \sum_{i=1}^{M} \rho\left(\left\| \frac{c_i}{m_c} - \frac{o_i}{m_o} \right\|_1 \mu_c \right) / M \rightarrow (2)$$

Based upon the changes in the position of the fiducial points, we measure changes in facial expression. Changes in the position of the fiducial points are transformed first into a set of mid-level parameters for AU recognition. We defined two parameters: ***up/down(P)*** and ***inc/dec(PP')***. Parameter ***up/down(P)*** $= y(P_{t1}) - y(P_t)$ describes upward and downward movements of point P and parameter ***inc/dec(PP')*** $= PP'_{t1} - PP'_t$ describes the increase or decrease of the distance between points P and P'. Based upon the temporal consistency of mid-level parameters, a rule-based method encodes temporal segments (onset, apex, offset) of 27 AUs occurring alone or in combination in the input face videos. For instance, to recognize the temporal segments of AU4, which pulls the eyebrows closer together, we exploit the following temporal rules:

IF ([*inc/dec*(DD1)]$_t$ > [*inc/dec*(DD1)]$_{t-1}$ + ε)
AND *inc/dec*(DD1) > ε THEN ***AU4-onset***
IF | [*inc/dec*(DD1)]$_t$ – [*inc/dec*(DD1)]$_{t-1}$ | ≤ ε
AND *inc/dec*(DD1) > ε THEN ***AU4-apex***
IF ([*inc/dec*(DD1)]$_t$ < [*inc/dec*(DD1)]$_{t-1}$ – ε)
AND *inc/dec*(DD1) > ε THEN ***AU4-offset***

When tested on the Cohn-Kanade Facial Expression Database [4] and the MMI Facial Expression Database [9], the proposed method achieved a recognition rate of 90% when detecting 27 AUs occurring alone or in combination in an input face image sequence (for details about this method see [7]).

3 User-profiled facial-afect recognition

As already noted above, virtually all systems for automatic facial affect analysis attempt to recognize a small set of universal/basic emotions [10]. However, pure expressions of "basic" emotions are seldom elicited; most of the time people show blends of emotional displays [5]. Hence, the classification of human non-verbal affective feedback into a single "basic"-emotion category is not realistic. Also, not all non-verbal affective cues can be classified as a combination of the "basic" emotion categories. Think for instance about the frustration, skepticism or boredom. Furthermore, it has been shown that the comprehension of a given emotion label and the ways of expressing the related affective state may differ from culture to culture and even from person to person [13]. Hence, pragmatic choices (user-profiled choices) must be made regarding the selection of affective states to be recognized by an automatic analyzer of human affective feedback.

The rest of this paper describes our case-based reasoning system that performs classification of AUs into the emotion categories learned from the user. The utilized case base is a dynamic, incrementally self-organizing event-content-addressable memory that allows fact retrieval and evaluation of encountered events based upon the user preferences and the generalizations formed from prior input. Each event (case) is one or more micro-events, each of which is a set of AUs. Micro-events related by the goal of communicating one specific affective state are grouped within the same dynamic memory chunk. In other words, each memory chunk represents a specific emotion category and contains all micro-events to which the user assigned the emotion label in question. The indexes associated with each dynamic memory chunk comprise individual AUs and AU combinations that are most characteristic for the emotion category in question. Finally, the micro-events of each dynamic memory chunk are hierarchically ordered according to their typicality: the larger the number of times a given micro-event occurred, the higher its hierarchical position within the given chunk. The initial endowment of the dynamic memory is achieved by asking the user to associate an interpretation (emotion) label to a set of 40 typical facial expressions (micro-events that might be hardwired to emotions according to [13]).

The classification of the AUs detected in an input face image into the emotion categories learned from the user is further accomplished by case-based reasoning about the content of the dynamic memory. To solve a new problem of classifying a set of input AUs into the user-defined interpretation categories, the following steps are taken:

1. Search the dynamic memory for similar cases, retrieve them, and interpret the input set of AUs using the interpretations suggested by the retrieved cases.
2. If the user is satisfied with the given interpretation, store the case in the dynamic memory. Otherwise, adapt the memory according to user-provided feedback on the interpretation he associates with the input facial expression.

The utilized retrieval and adaptation algorithms employ a pre-selection of cases that is based upon the clustered organization of the dynamic memory, the indexing structure of the memory, and the hierarchical organization of cases within the clusters/ chunks according to their typicality (for details about this method see [11]).

Two validation studies on a prototype system have been carried out. The question addressed by the 1st validation study was: How acceptable are the interpretations given by the system after it is trained to recognize 6 basic emotions? The question addressed by the 2nd validation study was: How acceptable are the interpretations given by the system, after it is trained to recognize an arbitrary number of user-defined interpretation categories? In the first case, a human FACS coder was asked to train the system. In the second case, a lay expert, without formal training in emotion signals recognition, was asked to train the system. The same expert used to train the system was used to evaluate its performance, that is, to judge the acceptability of interpretations returned by the system. For basic emotions, in 100% of test cases the expert approved of the interpretations generated by the system. For user-defined interpretation categories, in 83% of test cases the lay expert approved entirely of the interpretations and in 14% of test cases the expert approved of most but not of all the interpretation labels generated by the system for the pertinent cases.

4 Conclusion

In this paper, we presented two methods for AU detection in a nearly frontal view face video and a facial expression recognition system that performs classification of AUs into the emotion categories learned from the user.

The presented approaches extend the state of the art in automatic AU detection from face image sequences in two ways including temporal modeling of facial expressions and the number of AUs (21 and 27 AUs in total) handled. Namely, the automated systems for AU detection from face video that have been reported so far address mainly the problem of spatial modeling of facial expressions and, at best, can detect 16 to 18 AUs (from in total 44 AUs). Our methods also improve other aspects of automated AU detection compared to earlier works. The performance of both proposed methods is invariant to occlusions like glasses and facial hair as long as these do not entirely occlude facial points that are tracked (this is of importance for the second proposed AU detector). Also, the methods perform well independently of changes in the illumination intensity. As far as our method for automatic facial affect interpretation is concerned and given that the previously reported facial expression analyzers are able to classify facial displays only in one of the 6 basic emotion categories, the proposed method extends the state of the art in the field by enabling facial expression interpretation in a user-adaptive manner.

However, the proposed methods cannot recognize the full range of facial behavior (i.e. all 44 AUs defined in FACS). Furthermore, they assume that the input data are facial displays which are isolated or pre-segmented, showing a single temporal pattern (onset → apex → offset) of an expression that begins and ends with a neutral state. In reality, such segmentation is not available; human facial behavior is more complex and transitions from a facial (emotional) expression to another do not have to involve intermediate neutral state. Hence, our facial behavior analyzers cannot deal with spontaneously occurring facial displays. Further research efforts are necessary if the full range of human (spontaneous and posed) facial behavior is to be coded in an automatic way.

The work of M. Pantic is supported by the Netherlands Organization for Scientific Research Grant EW-639.021.202. The work of I. Patras and M.F. Valstar is supported by the Netherlands BSIK-MultimediaN-N2 Interaction project.

References:

1. Bartlett, M.S.; Littlewort, G.; Lainscsek, C.; Fasel, I.; Movellan, J. (2004). Machine learning methods for fully automatic recognition of facial expressions and facial actions. *Proc. Int'l Conf. Systems, Man and Cybernetics (SMC'04)*, 592-597.
2. Bobick, A.F.; Davis, J.W. (2001). The recognition of human movement using temporal templates. *IEEE Trans. Pattern Analysis & Machine Intelligence*, **23**, 257-267.
3. Ekman, P.; Friesen, W.V. (1978). *FACS Manual*. Consulting Psychologist Press, Palo Alto, USA.
4. Kanade, T.; Cohn, J.; Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proc. Int'l Conf. Face and Gesture Recognition (FGR'00)*, 46-53.
5. Keltner, D.; Ekman, P. (2000). Facial expression of emotion. *Handbook of Emotions*. Guilford Press, New York, USA, 236-249.
6. Pantic, M. (2005). Face for Interface. *The Encyclopedia of Multimedia Technology and Networking*. Idea Group Publishing, Hershey, USA.
7. Pantic, M.; Patras, I. (2005). Detecting facial actions and their temporal segments in nearly frontal view face image sequences. *Proc. Int'l Conf. System, Man and Cybernetics (SMC'05)*.
8. Pantic, M.; Patras, I.; Rothkrantz, L.J.M. (2002). Facial action recognition in face profile image sequences. *Proc. Int'l Conf. Multimedia and Expo (ICME'02)*, 37-40.
9. Pantic, M.; Valstar, M.F.; Rademaker, R.; Maat, L. (2005). Web-based database for facial expression analysis. In *Proc. Int'l Conf. Multimedia and Expo (ICME'05)*.
10. Pantic, M.; Rothkrantz, L.J.M. (2003). Toward an Affect-Sensitive Multimodal Human-Computer Interaction. *Proceedings of the IEEE*, **91**, 1370-1390.
11. Pantic, M.; Rothkrantz, L.J.M. (2004). Case-based reasoning for user-profiled recognition of emotions from face images. In *Proc. Int'l Conf. Multimedia and Expo (ICME'04)*, 391-394.
12. Pitt, M.K.; Shephard, N. (1999). Filtering via simulation: auxiliary particle filtering", *J. Amer. Stat. Assoc.*, **94**, 590-599.
13. Russell J.; Fernandez-Dols, J. (1997). *The psychology of facial expression*, Cambridge University Press, Cambridge, USA.
14. Tian, Y.; Kanade, T.; Cohn, J.F. (2001). Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Analysis & Machine Intelligence*, **23**, 97-115.
15. Valstar, M.F.; Pantic, M.; Patras, I. (2004). Motion History for Facial Action Detection from Face Video. *Proc. Int'l Conf. System, Man and Cybernetics (SMC'04)*, 635-640.
16. Viola, P.; Jones, M. (2001). Robust real-time object detection. *ICCV Workshop on Statistical and Computation Theories of Vision*.
17. Vukadinovic, D.; Pantic, M. (2005). Fully automatic facial feature point detection using Gabor feature based boosted classifiers. *Proc. Int'l Conf. System, Man and Cybernetics (SMC'05)*.

Bimodal Emotion Recognition

Nicu Sebe[1], Erwin Bakker[2], Ira Cohen[3], Theo Gevers[1], and Thomas Huang[4]
[1] *Faculty of Science, University of Amsterdam, The Netherlands,* [2] *LIACS Media Lab, Leiden University, The Netherlands,* [3] *HP Labs, USA,* [4] *University of Illinois at Urbana-Champaign, USA*

Abstract

Recent technological advances have enabled human users to interact with computers in ways previously unimaginable. Beyond the confines of the keyboard and mouse, new modalities for human-computer interaction such as voice, gesture, and force-feedback are emerging. Despite important advances, one necessary ingredient for natural interaction is still missing - emotions. This paper describes the challenging problem of bimodal emotion recognition and advocates the use of probabilistic graphical models when fusing the different modalities. We test our audio-visual emotion recognition approach on 38 subjects with 11 HCI-related affect states. The experimental results show that the average person-dependent emotion recognition accuracy is greatly improved when both visual and audio information is used in classification.

Keywords

Emotion recognition, bimodal approach, Bayesian networks, probabilistic graphical models

1 Introduction

In many important HCI applications such as computer aided tutoring and learning, it is highly desirable (even mandatory) that the response of the computer takes into account the emotional or cognitive state of the human user [7]. Computers today can recognize much of what is said, and to some extent, who said it but, they are almost completely in the dark when it comes to how things are said, the affective channel of information. Affective communication explicitly considers how emotions can be recognized and expressed during human-computer interaction. Addressing the problem of affective communication, Bianchi-Berthouze and Lisetti [1] identified three key points to be considered when developing systems that capture affective information: embodiment (experiencing physical reality), dynamics (mapping experience and emotional state with its label), and adaptive interaction (conveying emotive response, responding to a recognized emotional state).

Humans interact with each other mainly through speech, but also through body gestures, to emphasize a certain part of the speech and display of emotions. As a consequence, the new interface technologies are steadily driving toward accommodating information exchanges via the natural sensory modes of sight, sound, and touch. However, the roles of multiple modalities and their interplay remain to be quantified and scientifically understood. What is needed is a science of human-computer communication that establishes a framework for multimodal "language" and "dialog", much like the framework we have evolved for spoken exchange.

Mehrabian [8] indicated that when judging someone's affective state, people mainly rely on facial expressions and vocal intonations. Thus, affect recognition should inherently be related to multimodal analysis. From the perspective of engineering, even with improvements, the modalities of facial expression and prosody will undoubtedly be very fallible. The primary aim of our work in this paper is to combine cues from these modalities so that the affective state of a person can be inferred more accurately.

In this paper, we present our effort toward audio-visual HCI-related affect recognition. With HCI applications in mind, we take into account some special affective states that indicate user's cognitive/motivational states. We test our bimodal affect recognition approach on 38 subjects with 11 affective states. The experimental results show that the average person-dependent emotion recognition accuracy is greatly improved when both visual and audio information is used in classification.

2 Related work

Automatic affect analysis has attracted much interest from researchers in various research fields. The studies in facial expression recognition and vocal affect recognition have been done largely independent of each other. Most current works in facial expression recognition use still photographs or video sequences where the subject exhibits only facial expression without speaking any words. Similarly, the works on vocal emotion detection used only the audio information. There are situations where people would speak and exhibit facial expressions at the same time. For example, "he said hello with a smile." Pure facial expression recognizers may fail because the mouth movements may not fit the description of a pure "smile." For computers to be able to recognize emotional expression in practical scenarios, these cases must be handled.

According to the latest overview of automatic affect recognition [9], only a few reports [2][4][10][13][14] of bimodal affect recognition are found. Thus, while the recent research and technology advances make multimodal analysis of human affective states tractable, progress toward this direction is only just beginning. Compared with the previous reports of bimodal affect recognition listed in [9] and [11], the progress in this paper includes:

1) 11 affective states are analyzed, especially including 4 HCI-related affective states (confusion, interest, boredom, frustration). The other works (excepting [14]) only analyzed 5-6 basic emotions.
2) 38 subjects are tested. The numbers of subjects in [2][4] [13] are at most five. Thus, the generality of their algorithms is not guaranteed.
3) Bayesian networks are applied for bimodal fusion. [2][4] applied rule-based methods for combining two modalities; [13] used weighted intensity summation. It is not clear whether their rules or methods are suitable for more subjects.
4) Integrate a variable into the Bayesian network that indicates whether the person is speaking or not. In [14] a smoothing method is applied to reduce the detrimental influence of speech on the information provided by facial expression (see Figure 3); [2][4][13] did not take into account this problem.

3 Feature Extraction

The face tracker we use is based on a system developed by Tao and Huang [12] and described in detail in [3]. A snap shot of the system with the face tracking and the recognition result is shown in Figure 1.

Figure 1. *A snap shot of our facial expression recognition system. On the right side is a wireframe model overlayed on a face being tracked. On the left side the correct expression, Angry, is detected.*

This face tracker uses a model-based approach where an explicit 3D wireframe model of the face is constructed. In the first frame of the image sequence, landmark facial features such as the eye and mouth corners are selected interactively. A face model consisting of 16 surface patches embedded in Bezier volumes is then warped to fit the selected facial features. The surface patches defined this way are guaranteed to be continuous and smooth. The shape of the mesh can be changed by changing the locations of the control points in the Bezier volume.

Once the model is constructed and fitted, head motion and local deformations of the facial features such as the eyebrows, eyelids, and mouth can be tracked. First the 2D image motions are measured using template matching between frames at different resolutions. Image templates from the previous frame and from the very first frame are both used for more robust tracking. The measured 2D image motions are modeled as projections of the true 3D motions onto the image plane. From the 2D motions of many points on the mesh, the 3D motion can be estimated by solving an overdetermined system of equations of the projective motions in the least squared sense.

The recovered motions are represented in terms of magnitudes of some predefined motion of various facial features. Each feature motion corresponds to a simple deformation on the face, defined in terms of the Bezier volume control parameters. We refer to these motions vectors as Motion-Units (MU's). Note that they are similar but not equivalent to Ekman's AU's [5] and are numeric in nature, representing not only the activation of a facial region, but also the direction and intensity of the motion.

We use three kinds of prosody features for affect recognition: logarithm of energy, syllable rate and two pitch candidates and corresponding scores. The log energy is computed by $E = \log \sum_{i=1}^{N} x_i^2$ where N is the frame length and x_i is the i^{th} signal in that frame. For pitch extraction, an autocorrelation based pitch detector is used to extract two candidates of pitch frequency. The autocorrelation function is the correlation of a waveform with itself by delaying some time lag. The mathematical definition of the autocorrelation function is shown as follows: $Xor_p = \sum_{i=1}^{N} x_{i+p} x_i$ where x_{i+p} is the $(i+p)^{\text{th}}$ signal in that frame. The autocorrelation of periodic signal is also periodic. As the signal lags to the length of one period, the autocorrelation increases to the maximum; the first peak (excluding 0 tag) indicates the period of the signal. So the pitch is detected by

$$P_1 = \arg \max_{P\min \le p \le P\max} Xor_p$$

where *Pmin* is the possible minimum pitch and *Pmax* is the possible maximum pitch. In this paper, the search range for pitch is set to be 50~1000Hz. In addition to the pitch with the maximum autocorrelation score, the pitch P_2 with the second maximum of autocorrelation score is also chosen as a pitch candidate. Also, the autocorrelation scores are treated as features to detect whether the frame is a vowel. The syllable rate is computed by the formula: *#syllables/duration* where *duration* is the segment duration (0.5 second). To detect the numbers of syllables in the segments, a threshold-based speech detection method is used to detect the syllables in the signal. In detail, the frame is considered as speech if the following condition is satisfied:

$$E > EnThresh\ (50) \wedge Xor_{P_1} > XorThres\ (0.5)$$
$$\wedge\ Xor_{P_2} > XorThres\ (0.5) \wedge \left| \frac{P_{i1}}{P_{i2}} - 2 \right| < 0.2$$

where E the log energy of one frame, Xor_{P1} and Xor_{P2} are the autocorrelation scores of the two pitch candidates, P_{i1} and P_{i2} are the larger and the smaller values of P_1 and P_2, respectively. After speech detection, we can count the number of speech segments and compute the syllable rate as one dimension of prosody features. The prosody modality in our experiment can output 92 frames per second in real-time condition.

The features extracted are used as inputs to a classifying stage described in the next section.

4 Bayesian Networks and Fusion

A typical issue of multimodal data processing so far is that the multisensory data are typically processed separately and only combined at the end. Yet this is almost certainly incorrect; people display audio and visual communicative signals in a complementary and redundant manner. Chen and Huang [2] have shown this experimentally. In order to accomplish a human-like multimodal analysis of multiple input signals acquired by different sensors, the signals cannot be considered mutually independent and cannot be combined in a context-free manner at the end of the intended analysis but, on the contrary, the input data should be processed in a joint feature space and according to a context-dependent model. In practice, however, besides the problems of context sensing and developing context-dependent models for combining multisensory information, one should cope with the size of the required joint feature space, which can suffer from large dimensionality, different feature formats, and timing. Our approach to achieve the target tightly coupled multisensory data fusion is to develop context-dependent versions of a suitable method such as the Bayesian inference method [3].

If we consider the state of the art in audio and visual signal processing, noisy and partial input data should also be expected. A multimodal system should be able to deal with these imperfect data and generate its conclusion so that the certainty associated with it varies in accordance to the input data. Probabilistic graphical models, such as hidden Markov models (including their hierarchical variants), Bayesian networks, and dynamic Bayesian networks are very well suited for fusing such different sources of information. These models can handle noisy features, temporal information, and missing values of features all by probabilistic inference. Hierarchical HMM-based systems [3] have been shown to work well for facial expression recognition. Dynamic Bayesian networks and HMM variants have been shown to fuse various sources of information in recognizing user intent, office activity recognition, and event detection in video using both audio and visual information [6]. The success of these research efforts has shown that fusing audio and video for detection of discrete events using probabilistic graphical models is possible. Therefore, in this work we propose the Bayesian network topology for recognizing emotions from audio and facial expressions presented in Figure 2.

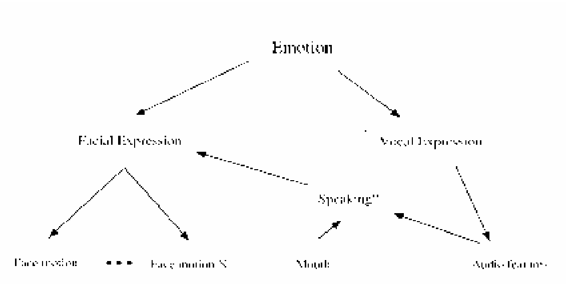


Figure 2. Bayesian network topology for bimodal emotion expression recognition.

The network topology combines the two modalities in a probabilistic manner. The top node is the class variable (recognized emotional expression). It is affected by recognized facial expressions, recognized vocal expressions, and by the context in which the system operates (if that is available). Vocal emotions are recognized from audio features extracted from the person's audio track. Facial expressions are recognized by facial features tracked using video, but the recognition is also affected by a variable that indicates whether the person is speaking or not. Recognizing whether a person is speaking uses both visual cues (mouth motion) and audio features. The parameters of the proposed network are learned from data. By using this framework, inferring the human emotional expression can be performed even when some pieces of information are missing, e.g., when audio is too noisy, or the face tracking loses the face.

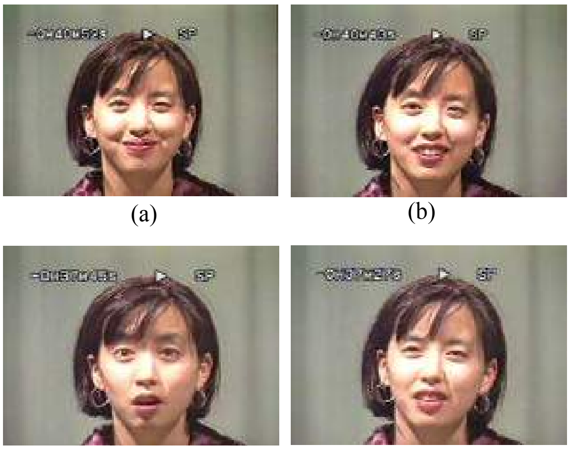
5 Experiments

In the previous reports [2][4][13], the datasets used were so small that the generality of their methods is not guaranteed. In addition, their methods only detected 5-6 basic emotional states that are not directly related to human computer interaction. However, we noticed that subjects facing a computer tutor seldom expressed these basic emotions except the neutral state. Actually, detecting some special affects, including interest, boredom, confusion, and frustration, is very important for the computer tutor to interact with users. These affects indicate the cognitive/motivational states of the subjects' learning. They provide information about whether the subject is engaged or whether the subject is having difficulties during the learning activities.

In our experiments we used a large-scale database [14] that is more related to the human-computer interaction; 11 affect categories were used which include 7 basic affects (i.e. happiness, sadness, fear, surprise, anger, disgust, and neutral), and 4 HCI-related affects (i.e. interest, boredom, confusion, and frustration). We tested our methods on 38 subjects (24 females and 14 males).

The subjects consist of mostly graduate and undergraduate students in various fields. Some staff and faculty members also volunteered to participate. Although the subjects displayed affect expression on request, minimal instruction was given to the subjects. In particular, no instruction on how to portray the affects was given. They were simply asked to display facial expressions and speak appropriate sentences. Each subject was required to pose a pure facial expression without speech three times, followed by a facial expression with speech three times, and then a pure facial expression three more times.

In the dataset in which subjects were facing the camera, we found that appropriate sentences made subjects more natural, and this way it was easier for them to express affects than without speech. In particular, without appropriate sentences, some subjects found it difficult to display subtle differences among the 4 HCI-related affects (confusion, interest, boredom, and frustration). On the other hand, speaking reduces the discriminability of the different facial expression. Figure 3 illustrates the comparison between pure facial expressions without speech and facial expression with speech. It shows that compared with pure facial expression in (a) and (c), the facial expressions with speech in (b) and (d) are more natural, but they are more difficult to be separated from each other.



(a) (b)

(c) (d)

***Figure 3**. Happy expression without speaking (a) and with speaking (b); Surprise expression without speaking (c) and with speaking (d).*

For every subject, the data was divided into two parts. One half of every subject's frames are selected as training data, and the other half as testing data. Our experimental results (Figure 4) show that the average recognition accuracy is about 56% for the face-only classifier, about 45% for the prosody-only classifier, but almost 90% for the bimodal classifier. The testing time was about 8.7 seconds for each of the three methods and for every affect.

The male-female average confusion matrix for the bimodal classifier is presented in Table 1. The analysis of the confusion matrix shows that the neutral state has the highest recognition accuracy, and is the state with which the other affective states are most confused. One reason behind this result is that every affective expression in the dataset started from the neutral state, and ended at the

Both		Detected										
		neut	hap	surp	ang	disg	fear	sad	frust	puzz	inter	bore
Desired	neut	98.66	0.27	0.27	0.00	0.00	0.00	0.54	0.00	0.00	0.27	0.00
	hap	4.07	89.15	0.68	0.68	0.68	0.68	1.36	0.00	0.00	1.02	1.69
	surp	4.70	0.31	88.40	1.57	0.00	1.57	1.88	0.00	0.94	0.00	0.63
	ang	1.16	0.23	0.47	94.65	1.40	0.23	1.16	0.00	0.47	0.23	0.00
	disg	2.31	0.77	0.77	3.08	88.72	0.77	0.77	1.03	0.51	0.51	0.77
	fear	1.81	0.30	0.30	3.93	1.51	89.12	1.21	0.00	1.51	0.30	0.00
	sad	4.60	0.31	0.31	3.07	1.53	0.00	88.04	0.61	0.31	0.31	0.92
	frust	3.31	0.37	2.21	3.31	0.37	2.57	2.21	83.46	1.10	0.37	0.74
	puzz	1.88	0.31	0.63	2.82	1.25	0.94	0.31	0.63	88.71	1.25	1.25
	inter	2.14	0.36	1.42	3.20	1.78	1.07	1.42	0.36	1.07	85.05	2.14
	bore	2.32	0.99	0.66	1.32	0.33	0.66	2.32	0.66	0.99	0.33	89.40

Table 1. *Male-female average confusion matrix for the bimodal classifier*

neutral state. Thus, the first and last few frames of each affect are very close to neutral states.

In addition, anger, boredom, and happiness have high recognition accuracies while sadness and interest have low accuracies. Contrary to the neutral state, the frustrated state is of lowest recognition accuracy, and is the state with which other affects are not likely to be confused.

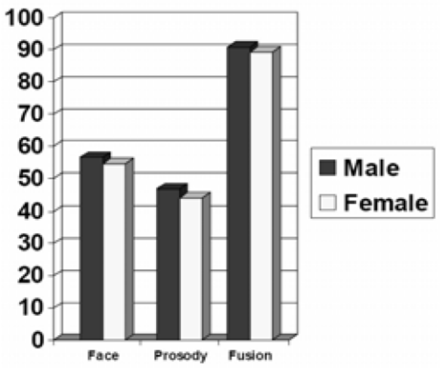


Figure 4. *Average affect classification accuracies*

6 Conclusion

The ability to detect and perceive a user's affective states could play an important role in human-computer intelligent interaction. With an automatic affect recognizer, the computer can respond appropriately to the user's affective state rather than simply responding to user commands. In this way, the computer would become more natural, persuasive, and friendly. In this paper, we introduce our effort toward multimodal affect recognition. Compared to previous studies our progress includes the use of a Bayesian network approach and an extensive testing on more subjects and more affective states. The experimental results show that the average person-dependent emotion recognition accuracy is greatly improved when both visual and audio information is used in classification.

References:

[1] Bianchi-Berthouze, N. and Lisetti, C., 2002. Modeling multimodal expression of user's affective subjective experience, *User Modeling and User-adapted Interaction,* **12**, 49-84.

[2] Chen, L. and Huang, T., 2000. Emotional expressions in audio-visual human computer interaction, in *Proc. ICME,* 423-426.

[3] Cohen, I., Sebe, N., Garg, A., Chen, L., and Huang, T., 2003. Facial Expression Recognition from Video Sequences: Temporal and Static Modeling, *Computer Vision and Image Understanding*, **91(1-2)**, 160-187.

[4] De Silva, L.C. and Ng, P.C., 2000. Bimodal emotion recognition, in Proc. *Face and Gesture Recogn.,* 332-335.

[5] Ekman, P. and Friesen, W., 1978. *Facial Action Coding System: Investigator's Guide,* Consulting Psychologist Press.

[6] Garg, A, Pavlovic, V., and Rehg, J., 2003. Boosted learning in dynamic Bayesian networks for multimodal speaker detection, *Proceedings of the IEEE,* **91(9),** 1355-1369.

[7] Goleman D., 1995. *Emotional intelligence,* Bantam Books.

[8] Mehrabian, A., 1968. Communication without words, *Psychology Today,* **2(4)**, 53-56.

[9] Pantic, M. and Rothkrantz, L., 2003. Toward an affect-sensitive multimodal human-computer interaction, *Proceedings of the IEEE,* **91(9)**, 1370-1390.

[10] Song, M., Bu. J., Chen. C. and Li, N., 2004. Audio-visual based emotion recognition: A new approach, in Proc. *CVPR,* 1020-1025.

[11] Sebe, N, Cohen, I., and Huang, T., 2005. Multimodal emotion recognition, *Handbook of Pattern Recognition and Computer Vision,* World Scientific.

[12] Tao, H. and Huang, T., 1998. Connected vibrations: A modal analysis approach to non-rigid motion tracking, in Proc. *CVPR,* 735-740.

[13] Yoshitomi, Y., Kim, S., Kawano, T., and Kitazoe, T., 2000. Effect of sensor fusion for recognition of emotional states using voice, face image, and thermal image of face, in Proc. *Int'l Workshop on Robot-Human Interaction,* 178-183.

[14] Zeng, Z., Tu, J., Liu, M., Zhang, T., Rizzolo, N., Zhang, Z., Huang, T., Roth, D., and Levinson, S., 2004.Bimodal HCI-related Affect Recognition, in Proc. *ICMI.*

A Robust Scheme for Facial Analysis and Expression Recognition

S. Ioannou, M. Wallace, K. Karpouzis and S. Kollias
Image, Video and Multimedia Systems Laboratory,
School of Electrical and Computer Engineering, National Technical University of Athens, GREECE

Abstract
Since facial expressions are a key modality in human communication, the automated analysis of facial images and video for the estimation of the displayed expression is central in the design of intuitive and human friendly human computer interaction systems. In this paper we present a robust integrated system able to consider issues such as uncertainty and lack of confidence in the process of feature extraction from image and video in the process of facial expression analysis and recognition. The proposed approach has been implemented in the framework of an EU funded R&D project

Keywords
Facial analysis, expression recognition, feature points, anthropometric validation.

1 Introduction
In recent years there has been a growing interest in improving all aspects of the interaction between humans and computers, providing a realization of the term "affective computing" [1]. Humans interact with each other in a multimodal manner to convey general messages; emphasis on certain parts of a message is given via speech and display of emotions by visual, vocal, and other physiological means, even instinctively (e.g. sweating) [2]. Everyday face-to-face communication utilizes many and diverse channels and modalities, increasing the flexibility of a communication scheme. In these situations, failure of one channel is usually recovered by another channel; this kind of behavior should actually be considered as a blueprint of the requirements of robust, natural and efficient multimodal HCI [3].

Despite common belief, social psychology research has shown that conversations are usually dominated by facial expressions, and not spoken words, indicating the speaker's predisposition towards the listener. Mehrabian indicated that the linguistic part of a message, that is the actual wording, contributes only for seven percent to the effect of the message as a whole; the paralinguistic part, that is how the specific passage is vocalized, contributes for thirty eight percent, while facial expression of the speaker contributes for fifty five percent to the effect of the spoken message [4]. This implies that the facial expressions form the major modality in human communication.

An overview of the methodologies used for automatic analysis of facial expression can be found in [5]. A usual approach to measuring deformation, fortified by the fact that there are inter-personal variations of facial action amplitude, is to refer to the neutral – expression face of a given person. An important parameter of this approach is the effectiveness of the image processing procedures. In actual situations, such as processing visual data from talk shows, many kinds of noise may hinder feature extraction: subjects turning their heads or moving their hands may lead to feature occlusion or bad and uneven lighting may hamper edge- or color-based feature extraction algorithms. As a result, the appearance and deformation of one or more features may not be available for a given frame of a video sequence; worse yet, an erroneous deformation estimate may be unknowingly fed into the knowledge representation infrastructure.

An ideal recognition system should be able to classify all visually distinguishable facial expressions; a robust and extensible face and facial action model is a vital requirement. Ideally, this would result in a particular face model setup uniquely describing a particular facial expression [6]. A usual reference point is provided by the 44 facial actions defined in FACS (Facial Action Coding System) whose combinations form a complete set of facial expressions and facial expressions with a similar facial appearance [7]. It has to be noted though, that some of the facial action tokens included in FACS may not appear in meaningful facial expressions, since the purpose of FACS is to describe any visually distinguishable facial action and not to concentrate on emotional expressions [8].

When put to practice, these principles typically suffer from the imperfection of the image or video processing components, that cannot always detect all the required facial features correctly (detected point on the face mapped to correct feature) and accurately (the exact position of the point on the face detected with absolute precision). Thus, errors, noise and uncertainty in general are inserted in the process of expression analysis from the very first step and are therefore inherent in the whole process.

2 Methodology outline
A very important requirement for an ideal facial expression architecture is that all of the processes therein have to be performed without any or with the least possible user intervention. This typically involves initial detection of the face, extraction and tracking of relevant facial information, and facial expression classification. In this framework, actual implementation and integration details are enforced by the particular application. For example, if the application domain of the integrated system is behavioral science, real-time performance may not an essential property of the system.

In the framework of MPEG-4 standard, parameters have been specified for Face and Body Animation (FBA) by defining specific Face and Body nodes in the scene graph; the initial goal of FBA definition is the animation of both realistic and cartoonist characters. Thus, MPEG-4 has defined a large set of parameters and the user can select subsets of these parameters according to the application.

MPEG-4 specifies 84 feature points on the neutral face, which provide spatial reference for Facial Animation Parameter (FAP) definition; these feature points are presented in Figure 1. FAPs are defined through the comparison of distances between pairs of feature points on the observed and the neutral face. Most of the techniques for facial animation are based the well-known system for describing "all visually distinguishable facial movements" FACS. FACS is an anatomically oriented coding system, based on the definition of "Action Units" (AU) of a face that cause facial movements. An Action Unit could combine the movement of two muscles or work in the reverse way, i.e., split into several muscle movement. The FACS model has inspired the derivation of facial animation and definition parameters in the framework of the ISO MPEG-4 standard [9]. In particular, the Facial Definition Parameter (FDP) and the Facial Animation Parameter set were designed in the MPEG-4 framework to allow the definition of a facial shape and texture through FDPs, thus eliminating the need for specifying the topology of the underlying geometry, and the animation of faces through FAPs, thus reproducing expressions, emotions and speech pronunciation.

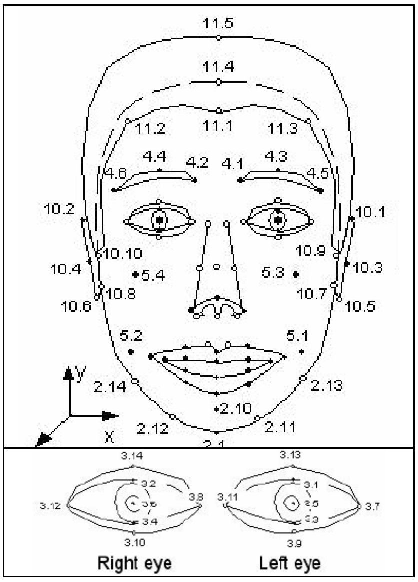


Figure 1. *MPEG-4 defined feature points on the neutral face.*

A long established tradition attempts to define facial expression in terms of qualitative targets, i.e. static positions capable of being displayed in a still photograph. The still image usually captures the apex of the expression, i.e. the instant at which the indicators of emotion are most marked. More recently emphasis, has switched towards descriptions that emphasize gestures, i.e. significant movements of facial features. Either way, analysis of the emotional expression of a human face requires a number of pre-processing steps. Following the most recent approach that emphasizes facial gestures, the required raw processing steps are to detect or track the face, to locate characteristic facial regions such as eyes, mouth and nose on it, to extract and follow the movement of facial features, such as characteristic points in these regions, or model facial gestures using anatomic information about the face. Continuing, extracted information needs to be combined with higher level knowledge, mapping detected facial feature movements to their corresponding facial expressions.

In our approach, we start by detecting the face, initially by using variance, mean color and number of skin color pixels in order to discard most candidate frames and applying more sophisticated face detection techniques on the rest. The detected face is pre-processed in order to roughly estimate regions of interest for i) the eyes and eyebrows and ii) the mouth. Each one of these ROIs is processed by a variety of methodologies in order to extract required facial features as accurately and certainly as possible. Distances between these feature points define the FAPs, which are combined in fuzzy rules in order to provide an estimation of the observed facial expression. The step in which uncertainty is most inherent, i.e. that of image processing for feature extraction, is analyzed in the following section.

3 Feature extraction

Besides expression representation, an important parameter of the expression analysis process is the effectiveness of the image processing procedures. Automatic analysis systems usually require good input to avoid misclassification or errors which is often ensured by the use of specific environment conditions such as in [13]. In actual situations, such as processing visual data from talk shows, many kinds of noise may hinder feature extraction: subjects turning their heads or moving their hands may lead to feature occlusion or bad and uneven lighting may hamper edge- or color-based feature extraction algorithms. As a result, the appearance and deformation of one or more features may not be available for a given frame of a video sequence; worse yet, an erroneous deformation estimate may be unknowingly provided as input to the subsequent expression analysis and classification procedures.

In this work, precise facial feature extraction is performed resulting in a set of masks, i.e. binary maps indicating the position and extent of each facial feature. The left, right, top and bottom–most coordinates of the eye and mouth masks, the left right and top coordinates of the eyebrow masks as well as the nose coordinates, are used to define the feature points. For the nose and each of the eyebrows, a single mask is created. On the other hand, since the detection of eyes and mouth can be problematic in low-quality images, a variety of methods is used, each resulting in a different mask. In total, we have four masks for each eye, three for the mouth and one for each one of the eyebrows. The methodologies applied in the extraction of these masks include:

- A feed-forward back propagation neural network trained to identify eye and non-eye facial area. The network has thirteen inputs; for each pixel on the facial region the NN inputs are luminance Y, chrominance values Cr & Cb and the ten most important DCT coefficients (with zigzag selection) of the neighboring 8x8 pixel area.
- A second neural network, with similar architecture to the first one, trained to identify mouth regions.
- Luminance based masks, which identify eyelid and sclera regions.
- Edge-based masks.
- A region growing approach based on standard deviation

Since, as we already mentioned, the detection of a mask using any of these applied methods can be problematic, all detected masks have to be validated against a set of criteria; of course, different criteria are applied to masks of different facial features. Each one of the criteria examines the masks in order to decide whether they have acceptable size and position for the feature they represent. This set of criteria consist of relative anthropometric measurements, such as the relation of the eye and eyebrow vertical positions, which when applied to the corresponding masks produce a value in the range [0,1] with zero denoting a totally invalid mask; in this manner, a validity confidence degree is generated for each one of the initial feature masks. For example, two criteria that can be used for the validation of the eye masks are the following:

$$M_{eye}^{1c} = 1 - \left| 1 - \frac{d_2 / d_6}{0.49} \right| \text{ and } M_{eye}^{2c} = 1 - \frac{|d_4|}{d_5}$$

where M_{eye}^{1c} and M_{eye}^{2c} are the confidence degrees acquired trough the application of each validation criterion on an eye mask. The former of the two criteria is based on [11], where the ration of eye width over bipupil breadth is reported as constant and equal to 0.49. In almost all cases these validation criteria, as well as the other criteria utilized in mask validation, produce confidence values in the [0,1] range. In the rare cases that the estimated value exceeds the limits, it is set to the closest extreme value, zero for negative values and one for values exceeding one..

For the features for which more than one masks have been detected using different methodologies, the multiple masks have then to be fused together to produce a final mask. The choice for mask fusion, rather than simple selection of the mask with the greatest validity confidence, is based on the observation that the methodologies applied in the initial masks' generation produce different error patterns from each other, since they rely on different image information or exploit the same information in fundamentally different ways. Thus, they provide independent information on the location on the mask; combining information from independent sources has the property of alleviating a portion of the uncertainty present in the individual information components. In other words, the final masks that are acquired via mask fusion are accompanied by lesser uncertainty than each one of the initial masks.

The fusion algorithm is based on a Dynamic Committee Machine structure that combines the masks based on their validity confidence, thus producing a final mask together with the corresponding estimated confidence. As already explained, this confidence degree is always higher than the degree of any of the considered initial masks. A final, more refined, confidence value can be acquired when also taking into account the temporal information from the video sequence. The final confidence for each feature mask is based on three parameters: absolute anthropometric measurements based on [11], face symmetry exploitation and examination of the facial feature size constancy over a period of ten frames. The outcome of this procedure is a set of final masks along with the final confidence of their validity.

A way to evaluate our feature extraction performance is Williams' Index (WI) [12], which compares the agreement of an observer with the joint agreement of other observers. An extended version of WI which deals with multivariate data can be found in [13]. The modified Williams' Index *I'* divides the average number of agreements (inverse disagreements, $D_{j,j'}$) between the computer (observer 0) and *n-1* human observers (*j*) by the average number of agreements between human observers:

$$WI = \frac{\frac{1}{n}\sum_{j=1}^{n}\frac{1}{D_{0,j}}}{\frac{2}{n(n-1)}\sum_{j}\sum_{j':j'>j}\frac{1}{D_{j,j'}}}$$

and in our case we define the average disagreement between two observers *j,j'* as

$$D_{j,j'} = \frac{1}{D_{bp}}\left\| M_j^x \copyright M_{j'}^x \right\|$$

where © denotes the pixel-wise xor operator, $\left\| M_j^x \right\|$ denotes the cardinality of feature mask *x* constructed by observer *j*, and D_{bp} (see table 1) is used as a normalization factor to compensate for camera zoom on video sequences.

These feature masks are used to extract the Feature Points (FPs) considered in the definition of the FAPs used in this work. Each FP inherits the confidence level of the final mask from which it derives; for example, the four FPs (top, bottom, left and right) of the left eye share the same confidence as the left eye final mask. Continuing, FAPs can be estimated via the comparison of the FPs of the examined frame to the FPs of a frame that is known to be neutral, i.e. a frame which is accepted by default as one displaying no facial deformations. For example, FAP F_{37} is estimated as:

$$F_{37} = \left\| FP_{4.5}^n - FP_{3.11}^n \right\| - \left\| FP_{4.5} - FP_{3.11} \right\|$$

where FP_i^n, FP_i are the locations of feature point i on the neutral and the observed face, respectively, and $\left\| FP_i - FP_j \right\|$ is the measured distance between feature points i and j. Obviously, the uncertainty in the detection of the feature points propagates in the estimation of the value of the FAP as well. Thus, the confidence in the value of the FAP, in the above example, is estimated as $F_{37}^c = \min(FP_{4.5}^c, FP_{3.11}^c)$. On the other hand, some FAPs may be estimated in different ways. For example, FAP F_{31} is estimated as:

$$F_{31}^1 = \left\| FP_{3.1}^n - FP_{3.3}^n \right\| - \left\| FP_{3.1} - FP_{3.3} \right\|$$

or as

$$F_{31}^2 = \left\| FP_{3.1}^n - FP_{9.1}^n \right\| - \left\| FP_{3.1} - FP_{9.1} \right\|$$

As argued above, considering both sources of information for the estimation of the value of the FAP alleviates some of the initial uncertainty in the output. Thus, for cases in which two distinct definitions exist for a FAP, the final value and confidence for the FAP are as follows: $F_i = \frac{F_i^1 + F_i^2}{2}$.

The amount of uncertainty contained in each one of the distinct initial FAP calculations can be estimated by $E_i^1 = 1 - F_i^{1c}$ for the first FAP and similarly for the other. The uncertainty present after combining the two can be given by some t-norm operation on the two

$E_i = t(E_i^1, E_i^2)$. The Yager t-norm with parameter $w = 5$ gives reasonable results for this operation:

$$E_i = 1 - \min\left(1, \left((1-E_i^1)^w + (1-E_i^2)^w\right)^w\right)$$

The overall confidence value for the final estimation of the FAP is then acquired as $F_i^c = 1 - E_i$. While evaluating the expression profiles, FAPs with greater uncertainty must influence less the profile evaluation outcome, thus each FAP must include a confidence value. This confidence value is computed from the corresponding FPs which participate in the estimation of each FAP.

Finally, FAP measurements are transformed to antecedent values x_j for the fuzzy rules using the fuzzy numbers defined for each FAP, and confidence degrees x_j^c are inherited from the FAP $x_j^c = F_i^c$ where F_i is the FAP based on which antecedent x_j is defined.

4 Expression recognition

In previous work we have defined expression vocabularies, i.e. the set of FAPs that each may be activated for each expression, and expression profiles, i.e. sets of FAP values, each profile representing a specific instance of the expression [14].

Each profile is easily transformed into a fuzzy rule, thus leading to the generation of a neurofuzzy classifier that, given the FAP values extracted from a still image as input, provides an estimation of the user expression as output.

In order to further reduce the uncertainty in the facial expression estimation, one may consider that although expression varies rapidly, emotion – and thus groups of emotion – do not vary equally rapidly. Based on this observation evidence theory can be used in order to combine the findings of the analysis, when applied on consecutive or almost consecutives frames of a shot [15].

5 Conclusions

Automatic analysis of facial pictures and video is an essential tool towards estimation of the human emotion in HCI. In this paper we have presented the abstract description of a system that is able to tackle the task, taking into account the uncertainty that is inherent in the comprising steps and utilizing anthropometric criteria and dynamic committee machines in order to alleviate a part of it.

The theory and methodologies presented herein have been developed in the framework of the ERMIS IST [16] project and are also being applied and extended in the HUMAIN European Network of Excellence [17], through the participation of the Image, Video and Multimedia Systems Laboratory of the National Technical University of Athens.

References:

1. Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J. (1999). Measuring Facial Expressions by Computer Image Analysis, *Psychophysiology*, **36**, 253-263.
2. Cowie, R., Douglas-Cowi,e E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J. (2001). Emotion Recognition in Human-Computer Interaction, *IEEE Signal Processing Magazine*.
3. Ekman, P. (1982). *Emotion in the Human Face*. Cambridge Univ. Press.
4. Ekman, P., Friesen, W.V. (1978) *Facial Action Coding System (FACS): A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, Palo Alto.
5. Mehrabian, A. (1968). Communication without Words, *Psychology Today*, **2(4)**, 53-56.
6. Picard, R.W. (1997). *Affective Computing*, MIT Press, Cambridge, MA.
7. Picard, R.W., Vyzas, E. (1999) Offline and Online Recognition of Emotion Expression from Physiological Data, *Emotion-Based Agent Architectures Workshop Notes, Int'l Conf. Autonomous Agents*, 135-142.
8. Raouzaiou, A., Tsapatsoulis, N., Karpouzis, K., Kollias, S. (2002). Parameterized facial expression synthesis based on MPEG-4, *Eurasip Journal on Applied Signal Processing*, **10**, 1021-1038.
9. Tekalp, A.M., Ostermann, J., (2000). Face and 2-D Mesh Animation in MPEG-4, *Signal Processing: Image Communication*, **15**, 387-421.
10. Young, J.W. (1993). Head and face anthropometry of adult U.S. civilians, *FAA Civil Aeromedical Institute*.
11. Williams, G.W. (1976). Comparing the joint agreement of several raters with another rater, *Biometrics*, **32**, 619-627.
12. Chalana, V., Kim, Y. (1997). A Methodology for Evaluation of Boundary Detection Algorithms on Medical Images, *IEEE Transactions on Medical Imaging*, **16**, 5.
13. Pantic, M. Rothkrantz, L.J.M. (2000). Expert system for automatic analysis of facial expressions, *Image and Vision Computing*, **18**, 881-905.
14. Pantic, M. Rothkrantz, L.J.M. (2000). Automatic Analysis of Facial Expressions: The State of the Art, Maja Pantic, Leon J.M. Rothkrantz, *IEEE Transactions on PAMI*, **22(12)**.
15. Wallace, M., Raouzaiou, A., Tsapatsoulis, N., Kollias, S. (2004). Facial Expression Classification based on MPEG-4 FAPs: The use of evidence and prior knowledge for uncertainty removal, *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Budabest, Hungary.
16. IST Project: *Emotionally Rich Man-Machine Interaction Systems* (ERMIS), 2001-2003. http://www.image.ntua.gr/ermis/
17. NoE: *Human-Machine Interaction Network on Emotion* (HUMAINE), 2004-2007. http://emotion-research.net/

On Combining the Facial Movements of a Talking Head

T.D. Bui[1], D. Heylen[2], A. Nijholt[2], M. Poel[2]
[1] *College of Technology, Vietnam National University, Hanoi, Vietnam*
[2] *Human Media Interaction, University of Twente, Enschede, The Netherlands*

Abstract
We present work on Obie, an embodied conversational agent framework. An embodied conversational agent, or talking head, consists of three main components. The *graphical* part consists of a face model and a facial muscle model. Besides the graphical part, we have implemented an *emotion model* and a mapping from emotions to facial expressions. The *animation* part of the framework focuses on the combination of different facial movements temporally. In this paper we propose a scheme of combining facial movements on a 3D talking head.

Keywords
Emotions, Emotional Facial Expressions, Facial Movement Combination, Embodied Conversational Agents.

1 Introduction
The combination of research in the field of computer graphics, autonomous agents, and speech and language technology has led to the development of embodied agents [2,3,20,21]. The emerging technology of embodied agents can realize different promising applications including human-like interfaces to improve the interaction between human and computer; simulated virtual characters for different applications such as entertainment, education, and the like; and believable animated characters to increase the interestingness of computer games.

In this paper, we present work on Obie, an embodied conversational agent framework (see ***Figure 4***). An embodied conversational agent, or talking head, consists of various components. The graphical part of the framework consists of a face model and a facial muscle model which allow realistic facial expressions to be produced in real-time on a standard pc. Besides the graphical part, we need a system that accounts for the actions (dialogue) and emotions of the agent. We have implemented an emotion model and a mapping from emotions to facial expressions. The animation part of the framework focuses on the combination of different facial movements temporally. We concentrate on the dynamic aspects of facial movements and the combination of facial expressions in different channels that are responsible for different tasks.

Section 2 discusses the emotion model and the mapping from emotions to emotional facial expressions in the framework. Section 3 focuses on the animation part of the framework. Some illustrations are presented in Section 4.

2 Emotions and emotional facial expressions
Emotions play an essential role in making embodied conversational more believable [19]. For our agents to experience emotions, we have implemented an emotion model in our framework [5]. Emotions can be expressed in different ways: facial expressions, gestures, speech, etc. Among these, facial movements play a very important role in interpreting emotions. We have therefore proposed a fuzzy rule based system to map an emotion state to emotional facial expressions [4].

Probably, the most popular descriptive work on how emotions are expressed on faces is done by Ekman and Friesen [10]. It discusses how several emotions as well as their blends are displayed on the face. We have based ourselves on this work to map emotion representations onto the contraction level of facial muscles. We focus on two aspects of generating emotional facial expressions. First, we take into account the continuous changes in expressions of an emotion depending on the intensity by which it is felt. Secondly, we propose a way to specify combinations of expressions due to more than one emotion, i.e., blends, in accordance with the literature mentioned. We have used a fuzzy rule-based system, which allows us to incorporate qualitative as well as quantitative information. Fuzzy rules can capture descriptions which are described in natural language as well as vague concepts like “slight sadness”, “more intense sadness”, etc. Moreover, the fuzzy rule-based approach can assure the smooth mapping between emotions and facial expressions.

Following Ekman and Friesen [10], we consider the following six so-called “basic “ emotions: **Sadness**, **Happiness**, **Anger**, **Fear**, **Disgust** and **Surprise**. These are claimed to be associated with prototypical facial expressions that are said to be universal in that research suggests that they are associated consistently with these across different cultures. Ekman and Friesen also describe in detail what the expressions for these emotions and certain blends look like. Emotion feelings may differ in intensity. In [10], it is pointed out how for each of the basic emotions the expression can differ depending on the intensity of the emotion. It is therefore important for us to build our system on a representation that takes intensities into account. We have used their descriptions as the basis for our fuzzy rules.

3 The facial animation
The Obie framework is mainly concerned with the facial expression behavior of an embodied agent. The expressions are displayed on a face model, which we developed to allow high quality and realistic facial expressions [6]. The face model is implemented together with a muscle model that produces realistic deformation of the facial surface, handles multiple muscle interactions correctly and produces bulges and wrinkles in real-time.

It is particularly important for the framework to deal with the problem of combining different facial movements temporally. Existing facial animation systems usually combine these movements by adding them up [1,8,15,17]. Unfortunately, when and how a movement appears and disappears, and how co-occurring movements are integrated (co-articulation effects, for instance) are difficult to quantify [12]. In addition, the problems of overlaying and blending facial movements in time, and the way felt emotions are expressed in facial activity during speech, have not been studied thoroughly [16]. Thus, the activity of human facial muscles is far from simply additive. A typical example would be smiling while speaking. The Zygomatic Major and Minor muscles contract to pull the corner of the lip outward, resulting in a smile. The viseme corresponding to the phoneme “@U” in the word “Hello” requires the contraction of the lip funneler Orbicularis Oris, which drives the lips into a tight, pursed shape. However, the activation of the Zygomatic Major and Minor muscles together with the lip funneler Orbicularis Oris would create an unnatural movement. We call these “conflicting” muscles. The activation of a muscle may require the deactivation of other muscles [10]. Depending on the priority of the tasks to be performed on the face, appropriate muscles are selected for activation. In most of the cases, the visual speech has higher priority than the smile. In some situations, the smile may have higher priority than the visual speech, for instance, when the subject is too happy to utter the speech naturally.

Based on psychological literature [10,9] and research in facial movement measurement and synthesis [13,18], we have proposed a scheme of combining facial movements on a 3D talking head. There are several types of movements, such as conversational signals, emotion display, etc. We call these channels of facial movement. We concentrate on the dynamic aspects of facial movements and the combination of facial expressions in different channels that are responsible for different tasks.

Although facial movements happen continuously, most of them are known from electromyography (EMG) studies to occur in distinct phases [13]. The flow of movements can then be broken up into so-called “atomic” movements. We define an “atomic” movement as a group of muscle contractions that share the same function (e.g., conversation signal, emotion display), start time, end time, and onset and offset duration. Each “atomic” facial movement belongs to a specific channel, which contains only non-conflicting movements. Atomic movements within a channel occur sequentially, although they may overlap each other at their beginning and ending. This classification is also based on the function of the movements [9].

In our system, we distinguish six channels:

- Channel 1 contains **manipulators**, which are movements to satisfy biological requirements of the face. In our system, we consider eye blinking to wet the eyes as manipulators. These movements are random rather than repeated with fixed rate as in [18]. The random eye blinking is generated based on the algorithm proposed in [14].
- Channel 2 contains **lip movements** when talking (represented as viseme segments). Lip movements are generated from the text that is going to be spoken by the talking head. The text is converted to phoneme segments (phoneme with temporal information -- starting and ending time). The phonemes are converted to corresponding visemes. Each viseme is equipped with a set of dominance functions of parameters participating in the articulation of the speech segment. We use dominance functions from [8] for each viseme segment.
- Channel 3 contains **conversational signals**. These are movements to accentuate or emphasize speech, or to provide feedback from a listener. They can occur on pauses due to hesitation or to signal punctuation marks (such as a comma or an exclamation mark). They are used to improve the interaction between the speaker and the listener. In some systems the generation of conversational signals has been done by analyzing the text [18] or speech [1].
- Channel 4 contains **emotion displays**, which are **emotional expressions** or **emotion emblems**. **Emotional expressions** are movements to express felt emotions of the speaker. On the other hands, **emotion emblems** express emotions that are being mentioned, or instance, a disgust expression when talking about something disgusting.
- Channel 5 contains **gaze movements** and Channel 6 contains **head movements**. Gaze and head movements are generated to support eye contact or to point to something during conversation. Head movements are also used to replace verbal content (e.g., nodding the head for saying yes). As the eyes and the head do not stay in the same place all the time, we use a noise generating function to create random subtle movements to make the talking head livelier.

We follow Pelechaud et al. [18] to synthesize facial movements in three phases: onset, apex, and offset. We used Essa's work [13] on analysis, identification and synthesis of facial expressions to design the temporal pattern of facial movements. Essa used exponential curves to fit the onset and offset portions of each parameter. Based on the suggested functions by Essa, we derived two functions for the onset and offset portion of a parameter activity. An example of the temporal pattern of a facial movement is shown in Figure 1.

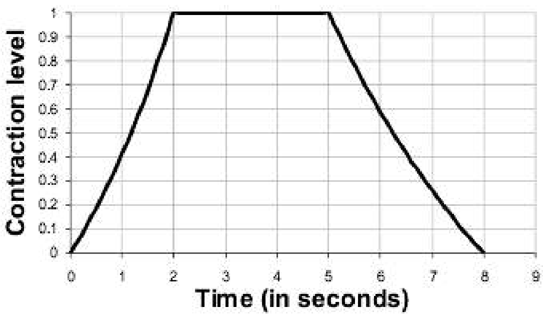


Figure 1. The activity function of a facial movement.

In order to combine facial movements, first, we concatenate the movements in the same channel. This is done by modulating the activity of each muscle involved in the movements in that channel, to create transition effects between movements. This combination only applies to individual muscles.

We use the dominance model [7] to create the co-articulation effect of lip movements when talking. Co-articulation is the blending effect that surrounding phonemes have on the current phonemes. For the combination of movements from other channels, we have proposed an algorithm to produce smooth transitions between movements. An example of combining the Jaw Rotation of two movements in the same channel is shown in Figure 2.

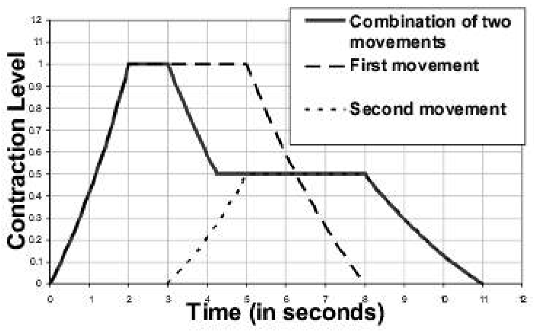


Figure 2. Combination of the Jaw Rotation of two movements in the same channel.

The movements from all channels are then combined taking into account the resolution of possible conflicting muscles. At a certain moment in time, when there is a conflict between parameters in different animation channels, the parameters involved in the movement with higher priority will dominate the ones with lower priority. The activity of those parameters around that time is also affected so that they cannot activate or release too fast. An example of combining the activity of Zygomatic major and Orbicularis Oris is shown in Figure 3.

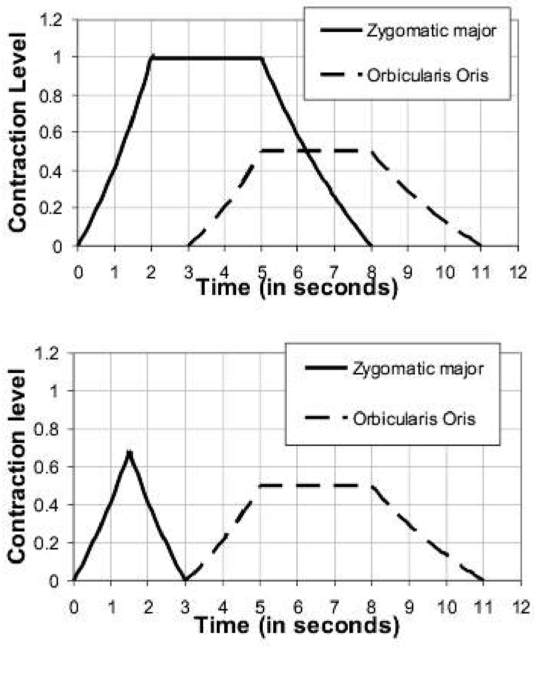


Figure 3. *The activity of Zygomatic major and Orbicularis Oris before (top) and after (bottom) applying the combination.*

4 Illustration

Figure 4 shows (frame by frame) our face model uttering the sentence “Oh, really?". An emotion display happens during the utterance, which is a full surprise display, starting at time 0 and lasting 2 seconds. The surprise display is mapped to the contraction of the parameter Jaw Rotation with the fuzzy rule based system. The lip movements during speech are calculated based on empirical data from Rutgers university [8]. The figure shows the smooth and natural combination of visual speech and the emotion display.

5 Conclusion

In this paper, we have discussed a framework for an embodied conversational agent. The framework is represented as a talking head. We focus on the emotion facial expressions of an agent as well as the dynamic in the combination of facial movements. Our framework has been built based on psychological literature as well as work on facial behavior measurement. With this framework, we can create smooth and natural animation for our embodied conversational agent.

In the future, we want to verify our framework with more empirical data from research on facial behavior measurement. We also want to do human judgment studies to evaluate our framework.

Dirk Heylen, Anton Nijholt, and Mannes Poel continue this research within the Humaine Network of Excellence.

REFERENCE

1. Albrecht, J. Haber, M. Schröder, and H. P. Seidel (2002). Automatic generation of non-verbal facial expressions from speech. In *Proceedings Computer Graphics International (CGI) 2002*, pp. 283–293.
2. E. André, G. Herzog, and T. Rist (1997). Generating multimedia presentations for robocup soccer games. In H. Kitano, ed., *RoboCup ’97: Robot Soccer World Cup I*, pp. 200–215. Springer-Verlag, New York.
3. B. Blumberg, M. Downie, Y. A. Ivanov, M. Berlin, M. P. Johnson, and B. Tomlinson (2002). Integrated learning for interactive synthetic characters. In *SIGGRAPH 2002*, pp. 417–426.
4. T. D. Bui, D. Heylen, M. Poel, and A. Nijholt (2001). Generation of facial expressions from emotion using a fuzzy rule based system. In *Proc. AI 2001*, pp. 83–95.
5. T. D. Bui, D. Heylen, M. Poel, and A. Nijholt (2002). Parlee: An adaptive plan-based event appraisal model of emotions. In *Proc. KI 2002*, pp. 129–143.
6. T. D. Bui, D. Heylen, and A. Nijholt (2003). Improvements on a simple muscle-based 3d face for realistic facial expressions. In *Proc. CASA-2003*, pp. 33–40.
7. Cohen, M. M. and Massaro, D. W. (1993). Modeling coarticulation in synthetic visual speech. In Magnenat Thalmann, N. and Thalmann, D., ed., *Models and Techniques in Computer Animation*, pp. 139-156. Springer, Tokyo.
8. D. C. DeCarlo, M. S. Revilla, and J. Venditti (2002). Making discourse visible: Coding and animating conversational facial displays. In *Computer Animation 2002*.
9. P. Ekman (1989). The argument and evidence about universals in facial expressions of emotion. In H. Wagner and A. Manstead, ed., *Handbook of Social Psychophysiology*. Wiley, Chichester, New York.

10. P. Ekman and Friesen, W. V. (1975). *Unmasking the Face: A Guide To Recognizing Emotions From Facial Clues*. Prentice-Hall, Englewood Cliffs, New Jersey.
11. P. Ekman and Friesen, W. V. (1978). *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, CA.
12. P. Ekman, T.S. Huang, T.J. Sejnowski, and J.C. Hager, ed. (1993). *Final Report to NSF of the Planning Workshop on Facial Expression Understanding*. Human Interaction Laboratory, University of California, San Francisco, 1993.
13. Essa, I. A. (1994). *Analysis, Interpretation, and Synthesis of Facial Expressions*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.
14. L. Itti, N. Dhavale, and F. Pighin (2003). Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proc. SPIE 48th Annual International Symposium on Optical Science and Technology*.
15. S. A. King, R. E. Parent, and B. Olsafsky (2000). An anatomically-based 3d parametric lip model to support facial animation and synchronized speech. In *Proceedings of Deform 2000*, pp. 7–19.
16. C. Latta, N. Alvarado, S. S. Adams, and S. Burbeck (2002). An expressive system for animating characters or endowing robots with affective displays. In *AISB 2002 Annual Conference, Symposium on Animating Expressive Characters for Social Interactions*.
17. C. Pelachaud, M. L. Viaud, and H. Yahia (1993). Rule structured facial animation system. *IJCAI*, 2:1610–1615.
18. C. Pelechaud, N. I. Badler, and M. Steedman (1996). Generating facial expressions for speech. *Cognitive Science*, 20:1–46.
19. Picard, R. (1997). *Affective Computing*. MIT Press, Cambridge, MA.
20. J. Rickel and W. L. Johnson (1998). Steve: A pedagogical agent for virtual reality. In *Proceedings of the Second International Conference on Autonomous Agents*.
21. M. Theune, S. Faas, D. Heylen, and A. Nijholt (2003). The virtual storyteller: Story creation by intelligent agents. In *Proceedings TIDSE 03*, pp. 204–215. Fraunhofer IRB Verlag.

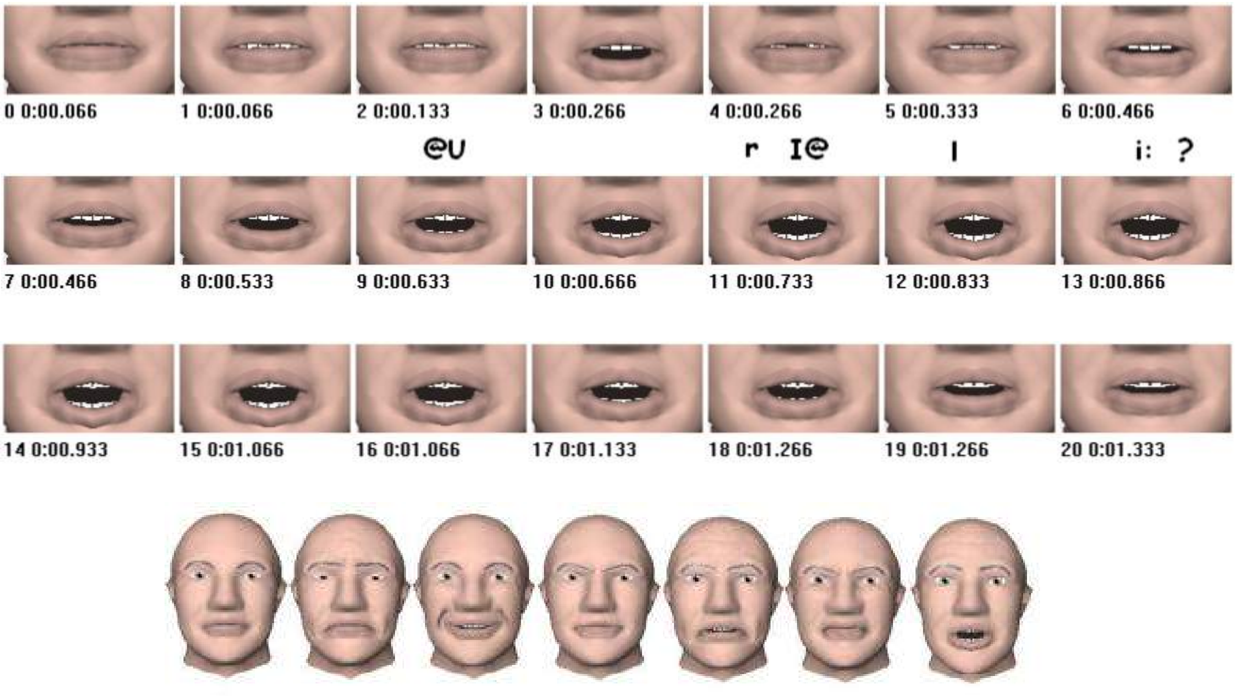


***Figure 4**. Our talking head utters the sentence “Oh! Really?” while displaying surprise and a sample of expressions.*

Fast facial animation design for emotional virtual humans

S. Garchery, A. Egges, N. Magnenat-Thalmann
MIRALab, University of Geneva, Geneva, Switzerland

Abstract

Designing facial animation parameters according to a specific model can be time consuming. In this paper we present a fast approach to design facial animations based on minimal information (only feature points). All facial deformations are automatically computed from MPEG-4 feature points. We also present an extension of this approach that allows to personalize or to customize the deformations according to different characteristics. We will describe different prototypes of the facial animation system, available on different platforms. Then, we demonstrate how emotions and expression can be incorporated into the facial animation system.

Keywords

Facial Animation, parameterization, personalization, emotion, MPEG4.

1 Introduction

A literature survey on different approaches on facial animation system reveals the following characteristics that should be found in an ideal facial animation system:

Easy to use: A facial animation system should be easy to use and simply to implement. This means:

- Be able to work with any kind of face model (male, female, children, cartoon like …)
- Require a minimum of time to set up a model for animation
- Allow for creative freedom of the animator to define specific deformations if necessary
- Get realistic results
- Be able to precisely control the animation

Integration: Using the system should be simple, fast and it should work in any kind of environment (PC, web, mobile…)
Generality: The possibility to reuse previous work (deformation data or animation) with a new model presents a big advantage and reduces the resources needed to develop new animations or applications.
Visual quality: finally, the result should look realistic with a cartoon like model or a cloned one. The quality should also be taken into account during the design process.

In order to achieve a maximum of these goals, it is crucial to properly define which parameters are used in the model. By proposing a parameterization system (FACS), Paul Ekman [4] started the definition of a standardization system to be used for facial synthesis. In 1999, MPEG-4 defined an interesting standard using facial animation parameters. This standard proposed to deform the face model directly by manipulating feature points of the face and presented a novel animation structure specifically optimized for e.g. networking applications. These parameters are completely model-independent, based on very few information and leave open the adaptation of animations for each face model according to the facial engine that is used. A lot of research has been done in order to develop facial animation engines based on this parameterization system. Commonly, a piece-wise linear interpolation function for each animation parameter [8,14] is used to produce the desired result (see section 2.1). This approach is easy to deploy but not easy to implement. It is useful for animating a face model, but it requires the preparation of deformation data. This data could be designed manually but in order to obtain good results a huge effort is required. Some research [12,15] has been done in order to simplify this work or to propose semi-automatic approaches.

In this paper we will briefly describe some aspects of the MPEG-4 standard for facial animation and present how it is applied in our facial deformation engine. Additionally, we will show a method to easily combine a quick preparation of a 3D model, together with a personalization of deformation. We will describe how different automatic and semi-automatic approaches can be applied to a talking head system, including a means to automatically personalize the facial animation in real-time. We will present different models of emotion that are used to control the expressivity of the face. Finally, we will present some results and conclusions.

2 Fast Facial Deformation Design

2.1 MPEG-4 overview and description

In order to understand facial animation based on MPEG-4 parameters system, we should describe some keywords of the standard and the pipeline in order to animate compliant face models.

- FAPU (Facial Animation Parameters Units): all animation parameters are described in FAPU units. This unit is based on face model proportions and computed based on a few key points of the face (like eye distance or mouth size).
- FDP (Facial Definition Parameters): this acronym describes a set of 88 feature points of the face model. FAPU and facial animation parameters are based on these feature points. These points could be also used in order to morph a face model according to specific characteristics.
- FAP (Facial Animation Parameters): it is a set of values decomposed in high level and low level parameters that represent the displacement of some features points (FDP) according to a specific direction. Two special values (FAP 1 and 2) are used to represent visemes and expressions. All 66 low level FAP values are used to represent the displacement of some FDPs according to a specific direction (see ***Figure 1***). The combination of all deformations resulting from these displacements forms the final expression. A facial animation then is a variation of these expressions over time.

Another aspect of MPEG-4 for facial animation, like Facial Interpolation Tables, could be applied to simplify the quantity of data needed to represent an expression or animation. With this approach, an animation can be represented by a small set of parameters, which is an efficient approach for network applications (less than 2 Kb for each frame).

The FAP stream does not provide any information on the displacement of neighboring vertices; therefore, for each FAP, we use a method that defines each displacement, i.e. which vertices are influenced and in which weight and direction according to FAP intensities. MPEG-4 provides a referencing method called Face Definition Tables, based on a piece-wise linear interpolation in order to animate the face model [7]. These tables (also referred to as Facial Animation Tables or FAT) provide information about which vertices should be translated or rotated for each

FAP displacement. More information can be found in Magnenat-Thalmann & Thalmann [11].

The Face Definition Tables representation is optional: the information it provides can be created by an animator by defining each influence on each FAP manually. However areas such as the lips (shown in *Figure 1*) make this work very tedious due to the close proximity of the 21 FAP values in a very small region.

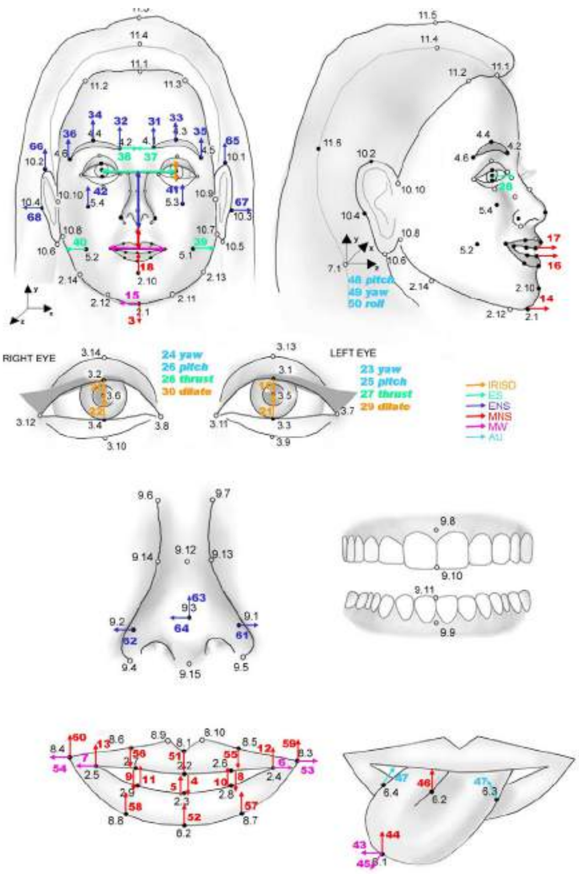


Figure 1. *Partial FDP position and FAP orientation*

2.2 Automatic and generic influence definition

In this paper we propose a generic approach to compute each region of influence. The main idea is to keep coherence and a maximum of simplicity in order to adapt this approach to different platforms or environments, but in the same time also to be able to produce realistic expressions. Exactly the same process is used for each FDP in order to define the influence area. In order to obtain a maximum of simplicity, we have developed an approach to compute deformations from simple FDP information.

Computation description

The main problem is to find a correct definition of the influence area for each FAP according to its neighboring vertices. We propose an approach based on the following process:

- Compute the distribution of FDPs on the model and a relation of distance between them.
- For each vertex of the mesh:
  1. Define which control point is able to influence it
  2. Define the ratio of influence of each control point

A 3D face model is composed of different meshes for eyes, teeth, tongue and skin. The skin mesh is mainly defined by the holes for the eyes and the mouth. Also, often a face model has a vertex distribution that is not uniform over the mesh. In order to develop a model-independent approach we have to take into account these specificities, and then we should define an appropriate distance measurement. A measurement based on Euclidian distance is efficient to manage the variations in mesh density but it does not take into account problems like holes in the model. A measurement based on the topology like the number of edges between vertices takes the holes into account, but it is not efficient for the mesh density variations.

We propose to use a metric based on both aspects: Euclidian distance and mesh topology. The metric is computed following this rule: "the distance is equal to the sum of the edge distances along the shortest path between two vertices". Using this metric, we are able to manage in the same time, holes and mesh density variation on the face model.

Our approach is based on the definition of a list of influencing feature points for each vertex in the face mesh. Initially, all of the feature points are in the influence list. First, we find the closest feature point to the vertex (according to the previously defined metric). Then, we remove all the feature points from the list that are in the plane perpendicular to the vector between the vertex and the selected feature point (see *Figure 2*). Then, we select the next closest feature point in the remaining list and we apply the same procedure until all feature points have been taken into account.

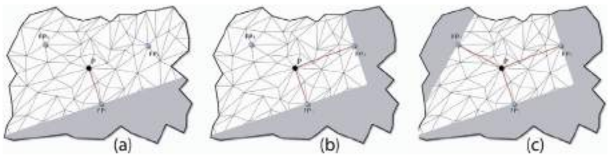


Figure 2. *3 steps to define potential feature point*

When the list of influencing feature points is established, we compute the influence of each of them to the current vertex by using a rule based on the metric explained above. If *P* is a vertex of the mesh, we compute a balanced sum *d*:

$$d = \frac{\sum_{i=1}^{n} d_i * \cos\theta_i}{\sum_{i=1}^{n} \cos\theta_i}$$

where *n* equals the number of influencing feature points for *P*, θ_i the angle between *P* and the feature point and d_i the distance between the feature point and the vertex. The weight associated to a specific feature point for *P* is computed as:

$$W_{i,P} = \sin\left(\frac{\pi}{2}\left(1 - \frac{d_i}{d}\right)\right)$$

By applying this weight computation for each vertex of the mesh, we obtain for each feature point a list of the vertices that are influenced by it, with an associated weight.

This approach takes into account the problem of overlapping regions and the diversity of the mesh (variation of density and holes).

Application to the face by defining simple constraint

This generic approach is applied on the 3D face model 3 times, one time for each direction of deformation. When we look at the FAP repartition from directional point of view, we can see a big variation in the feature points (see *Figure 3*).

In order to take into account this diversity and produce a realistic deformation, we compute a different influence area according to each displacement direction. We obtain then 3 different lists of vertices influenced for each feature point for each region. We use this list during the animation in order to deform the mesh according to feature point's displacements.

This information can easily be represented in the FaceDefMesh format and be used in an MPEG-4 compliant face system.

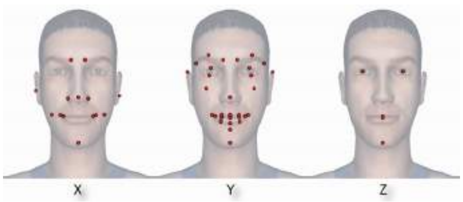


Figure 3. *Directional repartition of Facial Animation Parameters*

This initialization step of influence computation is done only once or is already saved in the MPEG-4 format. This approach is computationally very efficient (see *Figure 4*). We only need a few seconds to compute all this information for a model composed with more than 10K polygons on a standard PC.

Model Name	Nb Vertices	Nb Polygons	Influence process (sec.)	
Seb	431	774	PC1	0.50
			PC2	0.37
			PC3	0.23
Linda	2020	3849	PC1	2.48
			PC2	2.38
			PC3	1.47
Inam	5781	11017	PC1	6.71
			PC2	5.48
			PC3	3.34

Figure 4. *Computation of influence area on different face model. PC1 = AMD Athlon 1.7+, PC2 = PIV 2 Ghz, PC3 = PIV 3 Ghz.*

The advantages of our approach are multiple: we find a quick way to simplify the number of influent point, the same computation applies to all vertices independent of the part of face. Therefore we can add or remove feature points easily without changing the computation process. This approach works not only with MPEG-4 but with any feature point-based deformation approach.

2.3 Manual personalization

After computing the influences, a designer can still edit this information and modify it if necessary or if the designer would like to control a specific deformation. Starting for the previous process (automatic definition of influence) has big advantages. Defining all influence areas in a part of the face composed of many vertices and feature points is a lot of work. Starting from already calculated influence areas is easier for the animator to modify and it saves a lot of manual work.

3 Animation Design

Different approaches are possible in order to produce a facial animation stream more or less in real-time depending on the intended application. In this section, we will briefly present some of these approaches.

3.1 Text-to-visual approach

When starting from written text, we use a Text-to-Speech engine to produce the phoneme data and the audio. For defining the visemes and expressions, we use the Principal Components (PCs) as described by Kshirsagar et al. [9]. The PCs are derived from the statistical analysis of the facial motion data and reflect independent facial movements observed during fluent speech. The main steps incorporated in the visual front-end are the following:

1. Generation of FAPs from text
2. Expression blending: Each expression is associated with an intensity value and it is blended with the previously calculated co-articulated phoneme trajectories.
3. Periodic facial movements: Periodic eye-blinks and minor head movements are applied to the face for increased believability.

3.2 Optical capture

In order to capture a more realistic trajectory of feature points on the face, we are using a commercial optical tracking system to capture the facial motions, with 6 cameras and 27 markers corresponding to MPEG-4 FDPs. In the parameterization system, a total of 40 FDPs are animatable, but since markers are difficult to be set up on the tongue and lips, we use a subset of 27. We obtain 3D trajectories for each of the marker points as the output of the tracking system, suitable as well for 2D animation. During the data capture, head movements are not restricted and thus, a compensation process is required to obtain the local deformation of the marker [1].

Once we extract the head movements, the global movement component is removed from the motion trajectories of all the feature point markers, resulting in the absolute local displacements. The MPEG-4 FAP values are then easily calculated from these displacements. For example, FAP 3 open jaw is defined as the normalized displacement of FDP 2.1 from the neutral position, scaled by the FAPU MNS. As FAP values are defined as normalized displacements for the feature points from the neutral position, it is trivial to compute the FAP value given the neutral position and the displacement from this position [10]. The algorithm is based on a general purpose feature point based mesh deformation, which is extended for a facial mesh using MPEG-4 facial feature points.

3.3 Automatic personalization of animation

As presented above, an optical tracking system can be used in order to capture spatial motion of marker placed on the face. This information is converted in MPEG-4 FAPs that can be interpreted by any MPEG4 compliant facial animation engine. But during the conversion of motion capture data to FAP, we lose some information due of FAP restrictions (see section 2.1). In other words, with the standard format we cannot fully recover the same displacement as we had from the motion capture. The displacement in the directions which are not stored will be lost. We will present a solution to restore this information during the synthesis process.

As described above, our system is able to deform a face model according to the FDP position. This deformation is normally computed for FAP values in term of FAPU units. The system could move any FDP point and not only those points that are FAP values. In other words, we compute the deformation simply from the spatial position of control points. Thus we are able to deform a face model according to any the FDP position coming from FAP values or randomly. We propose then to apply a spring mass network in order to recalibrate the spatial position of control points. We do not want to connect the points linearly, because than it would be less dynamic.

In order to reach this goal, the first step is to define which FDP to connect and in which direction. For this we have studied different optical capture data and analyzed which displacements are lost during MPEG-4 conversion. We have obtained a relation between some feature points like eyebrows and head border for Z displacements or mouth displacement with cheek and the chin. With this mass spring correction of the FDP position before the deformation computation we are able to reproduce more degrees of freedom from the same data.

Another interesting idea with this approach is the possibility to personalize the value of the mass springs according to a specific person. The ultimate goal is to be able to automatically set these spring mass parameters, and use these parameters in order to reproduce a more realistic animation with the same number of facial animation parameters. This research is ongoing.

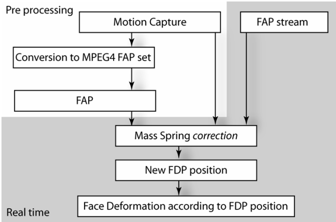


Figure 5. *Pipeline of facial animation with optional Mass spring correction of FDP position.*

4 Emotional System/Model

The concept of emotion has been widely researched in the psychology field. Many approaches and theories exist, but according to Cornelius [2], they can be broadly organized in four different theoretical perspectives: the Darwinian, Jamesian, cognitive, and social constructivist perspectives. The central organizing idea of the Darwinian perspective is the notion that emotions are phenomena that have evolved from important survival functions. This means that we should see more or less the same emotions (and expressions) in all humans. Theory and research from the Jamesian perspective views emotions from the perspective that it would be impossible to have emotions without bodily changes and bodily changes always come first. The cognitive approach to the study of emotions started with the work of Arnold [1]. In this approach, the central assumption is that thought and emotions are inseparable. More specifically, all emotions are seen within this perspective as being dependent on appraisal, the process by which events in the environment are judged as good or bad for us. Finally, Cornelius identifies the social constructivist approach to emotion study. Social constructivists believe that emotions are cultural products that owe their meaning and coherence to learned social rules.

In each of these perspectives, it is generally assumed that an emotional state can be viewed as a set of dimensions. The number of these dimensions varies among different researchers. Ekman [5] has identified six common expressions of emotion: fear, disgust, anger, sadness, surprise and joy, Plutchik [16] proposes eight, and in the OCC appraisal model, there are 22 emotions. The OCC model of appraisal has a computational counterpart that was developed by Elliott [6]. We propose a dynamic emotion simulation system that allows for any amount of different dimensions. The facial expressions are generated automatically from the emotional state [3].

5 Results and discussion

The techniques presented above have been tested and implemented on different platform and for different kind of applications. We present here briefly some result.

5.1 Multi-platform results

The same facial animation engine based on FDP position and manual adaptation in real time has been developed for different platforms such as PC, web or mobile.

Figure 6 (a) (b). *Java web application (without plugin) for facial and body animation. (c) Mobile facial animation system. (d) PC application sample.*

5.2 Application of emotional system

Using the system for emotion simulation, we can create a talking head which integrated speech, facial animations and emotional expressions. ***Figure 7*** shows some different expressions on a virtual human.

Figure 7. *Some expressions displayed by different virtual humans.*

This work is partially supported by the European Project HUMAINE (IST NoE 507422). A special thanks to Lionel Egger for his help in emotional design and to Edwin Tangenberg for work done on MPEG4 automatic recalibration data.

References

1. Arnold, M.B. (1960). Emotion and personality. *Columbia University Press*, New York.
2. Blostein, S.; Huang, T. (1988). Motion Understanding Robot and Human Vision. *Kluwer Academic Publishers*, pp. 329-352.
3. Cornelius, R.R. (1996). The science of emotion. *Research and tradition in the psychology of emotion*. Prentice-Hall, Upper Saddle River (NJ).
4. Egges, A.; Kshirsagar, S.; Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Computer Animation and Virtual Worlds*, **15(1),** 1-13.
5. Ekman, P.; Friesen, W.V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. *Consulting Psychologists Pres*s, Palo Alto, California.
6. Ekman, P. (1982). Emotion in the human face. *Cambridge University Press*, New York.
7. Elliott, C.D. (1992). The Affective Reasoner: A process model of emotions in a multiagent system. *PhD thesis*, Northwestern University.
8. Garchery, S.; Magnenat-Thalmann, N. (2001). Designing mpeg-4 facial animation tables for web applications. *In Multimedia Modeling 2001*, Amsterdam, pp. 39-59.
9. Kim, J.W.; Song, M.; Kim, I.J; Kwon, Y.M; Kim, H.G.; Ahn, S.C. (2000). Automatic fdp/fap generation from an image sequence. *In ISCAS 2000 - IEEE International Symposium on Circuits and Systems*.
10. Kshirsagar, S.; Molet, T.; Magnenat-Thalmann, N. (2001). Principal components of expressive speech animation. *In: Proceedings Computer Graphics International*, pp. 59-69.
11. Kshirsagar, S.; Garchery, S.; Magnenat-Thalmann, N. (2001). Feature Point Based Mesh Deformation Applied to MPEG-4 Facial Animation. *Kluwer Academic Publishers*, pp. 33-43.
12. Magnenat-Thalmann, N.; Thalmann, D. (2004). *Handbook of Virtual Human,* Wiley & Sons, Ltd.
13. Noh, J.Y.; Fidaleo, D.; Neumann, U. (2000). Animated deformations with radial basis functions. *In: VRST*, pp. 166-174.
14. Ortony, A.; Clore, G.L.; Collins A. (1988). The Cognitive Structure of Emotions. *Cambridge University Press*.
15. Ostermann, J. (1998). Animation of synthetic faces in mpeg-4. *In: Computer Animation.* Held in Philadelphia, Pennsylvania, USA.
16. Pasquariello, S. ; Pelachaud, C. (2001). Greta: A simple facial animation engine. *6th Online World Conference on Soft Computing in Industrial Applications*, Session on Soft Computing for Intelligent 3D Agents.
17. Plutchik, R. (1980). Emotion: *A psychoevolutionary synthesis.* Harper & Row, New York.

The CMU/Pitt Automated Facial Image Analysis System

T. Kanade[1], J.F. Cohn[2]
[1]*The Robotics Institute, Carnegie Mellon University, Pittsburgh, USA*
[2]*Department of Psychology, University of Pittsburgh, Pittsburgh, USA*

Abstract
To investigate the timing and configuration of facial actions, an interdisciplinary group of behavioral and computer scientists developed a computer-vision based approach, the CMU/Pitt Automated Facial Image Analysis (AFA) System. AFA has progressed through several versions and is capable of automatically recognizing facial action units and analyzing their timing in facial behavior. The latest version is based on Active Appearance Models (AAMs). These are generative, parametric models and consist of a shape component and an appearance component. The shape component is a triangulated mesh that deforms in response to changes in the parameters corresponding to a face undergoing both rigid motion (head pose variation) and non-rigid motion (expression). The appearance component of the AAM is an image of the face, which itself can vary under the control of the parameters. As the parameters are varied, the appearance varies so as to model effects such as the emergence of furrows and wrinkles and the visibility of the teeth as the mouth opens. The system extracts and separates head orientation, 3D shape deformation, and appearance change of the face, which are then input to a facial action recognizer. AFA has demonstrated concurrent validity with human-observer based facial expression recognition and both human-observer and EMG based analysis of timing.

Keywords
Automatic facial image analysis, facial expression, timing.

1 Introduction
People are highly sensitive to the timing of facial actions (Edwards, 1998). For example, slower facial actions, within limits, appear more genuine [18, 22], as do those that are more synchronous in their movement [14].

The timing of facial actions has been difficult to study in part because manual measurement of timing is relatively coarse and labor intensive [1]. Until recently, the only automatic method of measuring facial action has been facial electromyography (EMG). Facial EMG effectively quantifies the timing of covert muscle action [11], but the need to attach sensors to the face restricts its use to controlled settings, and may in itself affect facial behavior [2]. For more natural and less obtrusive measurements of facial actions (both timing and configurations), an interdisciplinary group of behavioral and computer scientists developed a computer-vision based approach, the CMU/Pitt Automated Facial Image Analysis (AFA) System. AFA is capable of unobtrusively recognizing facial action units and analyzing their timing in facial behavior.

a
b
c

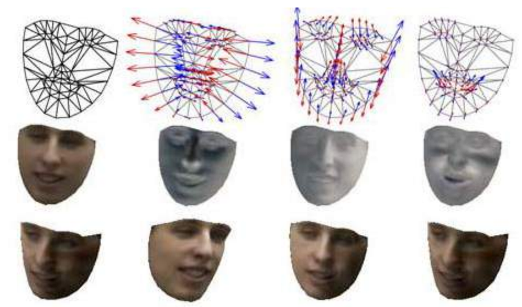


Figure 1. (a) The mean face shape (left) and 1st 3 shape modes. (b) The mean appearance (left) and 1st 3 appearance modes. (c) Four example faces generated with the AAM in (a) and (b).

2 CMU/Pitt Automated Facial Image Analysis

The latest version of the CMU/Pitt Automated Facial Image Analysis (AFA) system is based on Active Appearance Models (AAMs) [10, 20]. These are generative, parametric models that consist of a shape component and an appearance component. The shape component is a triangulated mesh that deforms in response to changes in the parameters corresponding to a face undergoing both rigid motion (head pose variation) and non-rigid motion (expression). The appearance component is an image of the face, which itself can vary under the control of the parameters. As the parameters are varied, the appearance varies so as to model effects such as the emergence of furrows and wrinkles and the visibility of the teeth as the mouth opens.

The complete AAM in which shape and appearance components are combined can generate faces undergoing a wide range of expression changes. For example, Figure 1(c) contains face images generated using the AAM in Figure 5(a-b) by setting appropriate model parameters. The key technique that allows an AAM to be used for facial analysis is a fitting algorithm. Given an input image of the face, a fitting algorithm searches for the model parameters that best match the input image.

2.1 Recovery of 3D head motion and shape

Traditional AAMs are actually "2D" in the sense that rigid head motion and non-rigid facial motion are confounded in the 2D-mesh shape model [20]. Thus, a 2D-based AAM does not handled larger head motions. To address these problems, we use an extension to AAMs that augments the usual 2D mesh model with an actual 3D shape model, thus separately and explicitly modeling the 3D rigid motion of the head and 3D non-rigid facial expression into two disjoint sets of parameters [25]. This advancement allows us to extract and separate the pose, 3D shape deformation, and appearance change of the face. We refer this approach as a 2D+3D AAM because it estimates both 2D and 3D parameters.

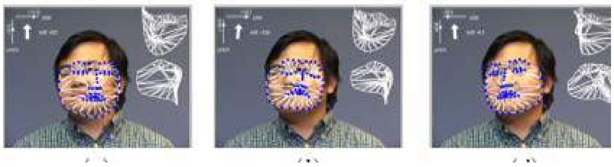


Figure 2. An example of fitting a 2D+3D AAM to a single image. Shown are the initial fit, intermediate result, and final result after the algorithm has converged. The 3D shape estimate (white) is projected onto the original image and also from two novel viewpoints (top right). Estimates of 3D pose (top left) and 2D shape (blue dots) are displayed as well. Estimated head orientation (pitch, roll, and yaw) is shown in the top left-hand corner of each image. From [25]

Figure 2 shows an example of fitting a 2D+3D AAM to a single image. The first image shows the initial estimate of 2D and 3D shape parameters. The former appear as blue dots, and the latter as a white mesh. The fitting process is iterative. The final result is shown in the last figure.

This same approach can be applied to an image sequence with change in pose (rigid head motion) and expression (non-rigid head motion). This is illustrated in Figure 3. A 2D+3D AAM was fit to each frame in an image sequence. The face mesh tracked the rigid and non-rigid motion of the face. By separating rigid and non-rigid motion, the timing of facial actions, such as smiling, is not confounded by rigid head motion.

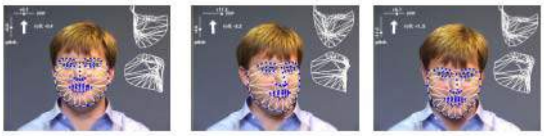


Figure 3. An example of tracking an expressive face in a video by fitting the model successively to each frame. The results in 3 frames from a longer image sequence are shown here. As the model is fit to the image, the 3D shape (white mesh), 2D shape (blue dots), and the 3D pose are estimated. Estimated head orientation (pitch, roll, and yaw) is shown in the top left-hand corner of each image.

2.2 Parameter estimation using multiple cameras As out-of-plane head motion becomes large, automatic recovery of 3D shape becomes increasingly difficult because of self-occlusion. To solve this problem, we fit a single AAM to multiple images captured simultaneously from multiple cameras. An example of this approach is shown in Figure 4. The first row shows the initialization of the mesh at the first iteration. The initial mesh at first does not correspond to its correct position on the face, but converges to the right one. The bottom row shows the result after the algorithm converges and the correct result is obtained.

3 Capabilities and Limitations

This version of AFA is capable of fully automatic initialization and recovery of 3D face shape and appearance. By using an individualized 3D face model, it more accurately extracts change in facial features than previous versions [3]. Processing speed exceeds video frame rate (30 Hz) and long image sequences can be processed. The system is not limited to offline use. It can be used in real-time applications, such as gaze estimation while driving [16, 17].

The system in its current state has some limitations. One, before use with video of a new person, it must first be trained on images of the person's face. This training may require hand labeling 10 to 20 or more images and is time consuming. Efforts are underway to make training more efficient. Two, facial features extraction is based on a set of 68 points, only a subset of which is informative of facial expression. Many of the points are for head tracking and feature registration. Increasing the number of feature points may be informative for facial feature analysis. A final limitation is the need for multiple synchronized video streams when out-of-plane head motion becomes moderate to large. A possible solution is to integrate the 3D cylindrical head tracker that had been used in the previous version [26] with the 2D+3D AAM-based head tracking.

4 Use of AFA to Investigate the Configuration and Timing of Facial Actions

4.1 Configuration of facial actions

We have used this and previous versions of AFA to recognize facial action units, make comparisons with criterion measures of facial dynamics, and investigate the timing of facial behavior and head motion. In directed facial action tasks, AFA has shown high agreement with manual FACS coding for approximately 20 action units.[8, 19, 23, 24] The action units recognized include most of those that have been a focus in the literature on facial expression and emotion [13]. In the more challenging case of spontaneous facial behavior with non-frontal pose, out-of-plane head motion, and occlusion from glasses or facial jewelry, AFA achieved 98% agreement with manual FACS [12] coding of blinks

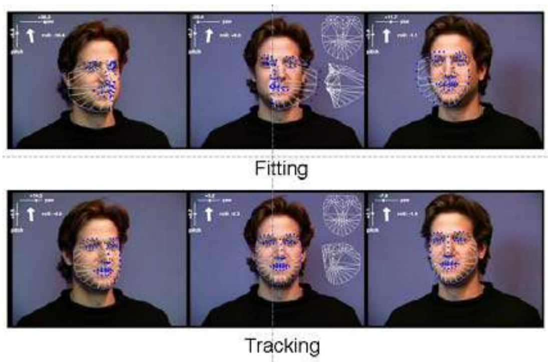


Figure 4. An example of multi-view 3D AAM fitting. Each image is overlaid with the corresponding 2D shape for that image in blue dots. The head pose is displayed in the top left of each image as pitch, roll, and yaw. The single 3D shape is displayed in the top right of the center image. This 3D shape is also overlaid in each image as a white mesh. From [15].

[7] and 89% accuracy for eyebrow raising and lowering [4].

4.2 Comparison with facial EMG measurements

To evaluate the temporal precision of AFA, we compared it with facial EMG. Facial EMG is considered a gold standard for measurement of facial muscle activity. AFA measurement of lip-corner displacement and *Zygomaticus major* EMG were compared [5, 21]. Facial EMG was recorded while subjects watched a film clip of a comedy routine intended to elicit spontaneous smiles (AU 12 in FACS). We analyzed an 11-second interval beginning 1-second prior to the punch line of a joke and continuing for 10 seconds. This interval was intended to capture the onset, peak, and offset of each smile. Smile onsets were highly correlated ($r > 0.90$), with EMG onset preceding AFA onset by about a quarter second. (See Figure 5 for an example). The amount of lip corner motion was also in agreement in 72% of cases with distinct EMG onset, where the amount was quantified in EMG by its amplitude, and in AFA by the visual displacement $\Delta d = \sqrt{\Delta x^2 + \Delta y^2}$. Note that EMG can detect occult changes in muscle activation which may not result in visible motion (AU12) that AFA can detect, so we consider this a good agreement. Relating physiological measurement (i.e., EMG) and visible behavior (lip motion) at this level of precision has not been possible using manual coding, which lacks sufficient temporal resolution [1].

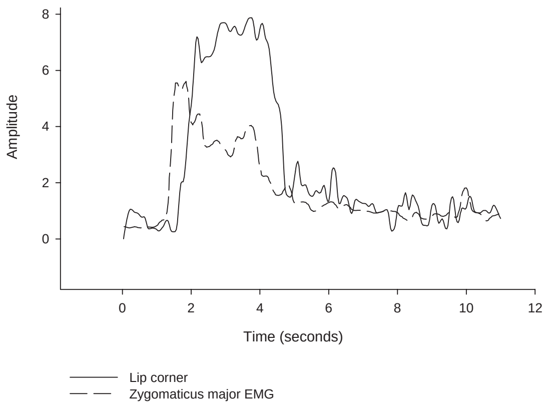


Figure 5. Relation between zygotmatic major EMG and lip-corner displacement in a spontaneous smile (AU 12).

4.3 Dynamics of interpersonal behavior

In work with Daniel Messinger at the University of Miami, we have used AFA to track changes in facial expression of mothers and infants during face-to-face interaction. Figure 6 shows an example from one of two mother-infant dyads time series for lip-corner displacement, a measure of smile intensity, and mouth opening. These and related measurements enable us to more rigorously test parent-infant bidirectional influence and synchrony than was previously possible [6], and provide new capability to investigate features of emotion regulation [9].

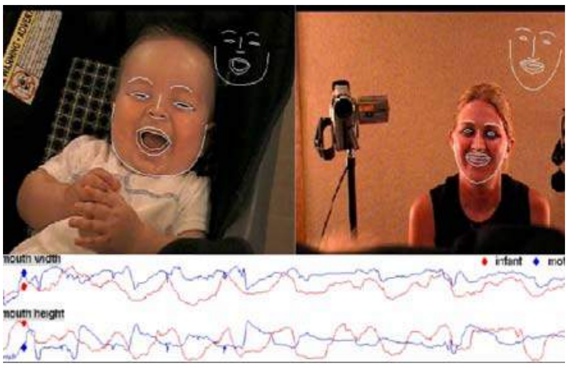


Figure 6. Intra- and interpersonal coordination of lip-corner displacement (smile intensity) and mouth opening in mother and infant.

5. Conclusions

The CMU/Pitt Automated Facial Image Analysis System has progressed through several versions. The current version uses generative, parametric shape and appearance models and estimates both 2D and 3D parameters. For large out-of-plane head motion, a multi-camera version may be used. The system is able to extract and separate head orientation, 3D shape deformation, and appearance change of the face. Previous versions have demonstrated concurrent validity for action unit recognition in both deliberate and spontaneous facial behavior with out-of-plane head motion and occlusion. The current version is more robust to long image sequences, has greater ability to represent the 3D structure of the face, and shows concurrent validity with facial EMG for the timing of facial actions. Unlike facial EMG, it is unobtrusive and has high specificity for observable facial action.

Supported by grants R01 MH051436 from the National Institute of Mental Health and BCS-0418001from the National Science Foundation.

References

1. Cohn, J. F., Ambadar, Z., & Ekman, P. (In press). Observer-based measurement of facial expression with the Facial Action Coding System. In J. A. Coan & J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science*. New York, NY: Oxford University.
2. Cohn, J. F., & Ekman, P. (in press). Measuring facial action by manual coding, facial EMG, and automatic facial image analysis. In J. A. Harrigan, R. Rosenthal & K. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences*. New York: Oxford.
3. Cohn, J. F., & Kanade, T. (in press). Use of automated facial image analysis for measurement of emotion expression. In J. A. Coan & J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment*. New York, NY: Oxford.
4. Cohn, J. F., Reed, L. I., Ambadar, Z., Xiao, J., & Moriyama, T. (2004). Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior. *IEEE Conference on Systems, Man, and Cybernetics,* the Hague, the Netherlands, 210-216.
5. Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing,* **2**, 1-12.
6. Cohn, J. F., & Tronick, E. Z. (1988). Mother-Infant face-to-face interaction: Influence is bidirectional and unrelated to periodic cycles in either partner's behavior. *Developmental Psychology, 34,* 386-392.
7. Cohn, J. F., Xiao, J., Moriyama, T., Ambadar, Z., & Kanade, T. (2003). Automatic recognition of eye blinking in spontaneous facial behavior. *Behavior Research Methods, Instruments, and Computers,* **35**, 420-428.
8. Cohn, J. F., Zlochower, A. J., Lien, J. J. J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology, 36,* 35-43.
9. Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development,* **75**, 317-333.
10. Cootes, T., Edwards, G., Taylor, C. (2001). Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**, 681-685.
11. Dimberg, U., Thunberg, M., & Grunedal, S. (2002). Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition and Emotion,* **16**, 449–471.

12. Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.
13 Ekman, P., & Rosenberg, E. (Eds.). (2005). *What the face reveals* (2nd ed.). New York, NY: Oxford.
14. Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stakes lies. *Journal of Personality and Social Psychology,* **72**, 1429-1439.
15. Hu, C., Xiao, J., Matthews, I., Baker, S., Cohn, J. F., & Kanade, T. (2004). Fitting a single active appearance model simultaneously to multiple images. *British Machine Vision Conference*, London, United Kingdom, 437-446.
16. Ishikawa, T., Baker, S., Matthews, I., & Kanade, T. (2004). Passive driver gaze tracking with active appearance models. *World Congress on Intelligent Transportation Systems*, Nagoya, Aichi, Japan.
17. Ishikawa, T., Baker, S., Matthews, I., & Kanade, T. (2004). Passive driver gaze tracking with active appearance models. *Robotics Institute Technical Report CMU-RI-TR-04-08*.
18. Krumhuber, E., & Kappas, A. (2005). Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior,* **29**, 3-24.
19. Lien, J. J. J., Kanade, T., Cohn, J. F., & Li, C. C. (2000). Detection, tracking, and classification of subtle changes in facial expression. *Journal of Robotics & Autonomous Systems,* **31**, 131-146.
20. Matthews, I., & Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision,* **60**, 135-164.
21. Reed, L. I., & Cohn, J. F. (Eds.). (In preparation). *Comparsion of AFA and facial EMG*. Pittsburgh: University of Pittsburgh.
22. Schmidt, K. L., Ambadar, Z., & Cohn, J. F. (Submitted). Timing of lip corner movement affects perceived genuineness of spontaneous smiles.
23. Tian, Y., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **23**, 97-115.
24. Tian, Y., Kanade, T., & Cohn, J. F. (2002, May). Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. *IEEE Conference on Automatic Face and Gesture Recognition,* Washington, D.C., *229-234.*
25. Xiao, J., Baker, S., Matthews, I., & Kanade, T. (2004). *Real-time combined 2D+3D active appearance models. IEEE Conference on Computer Vision and Pattern Recognition*, Washington, D.C.
26. Xiao, J., Kanade, T., & Cohn, J. F. (2003). Robust full motion recovery of head by dynamic templates and re-registration techniques. *International Journal of Imaging Systems and Technology,* **13,** 85-94.

Defining and measuring dominant-submissive behavior

E. Malatynska
Johnson & Johnson Pharmaceutical Research and Development, Spring House, PA, U.S.A.

Neuroscientists make every effort to automate behavioral tests to minimize subjective judgment and manual labor, and to maximize repeatability of the results between laboratories while increasing the throughput of experimental endpoints. In recent years, these efforts have accelerated due to the development of many inbred or transgenic mice strains that needed systematic behavioral phenotyping in order to find the functions delineated to strain differences or specific genes. At the same time, the development of in vitro biochemical high-throughput drug screening tests highlighted similar needs in behavioral neurosciences where the intact organism is a subject of study.

This symposium focuses on dominant-subordinate behavior across species. Social hierarchy is common in many animal phyla including fish, reptiles, birds and mammals. This extends to different species of a phylum. For example, within mammals this behavior is found among mice, rats, dogs, and most primates as well as humans. The premise of this symposium is that dominance and subordination observed in humans can be related to mania and depression and that important elements of both mania and depression can be modeled in animals based on observation of dominant and submissive behavior exhibited under well-defined conditions.

Human dominant-subordinate social behavior can be viewed as abnormal (e.g., active non-participant 'manic' and passive non-participant 'sad') or normal social behavior (active participant 'warm' and passive participant 'shy'). The interaction of subjects under these four conditions and their subsequent behavior on a dyadic game will be described in terms of the behavioral mechanisms of social adaptation. Investigators have used various endpoints to measure dominance in animals. This includes observation of hierarchy in groups of animals (two or more) by denoting their communication through different body postures, competition for priority of access to different resources using distinct types of scoring, and social defeat based on animal territorial instinct. These distinct approaches to measure dominance and submissiveness will be discussed by different speakers during the symposium, with the focus on the criteria used to define dominance and submissiveness and efforts to automate measurements.

Just as affective disorders are mood illnesses with two opposite poles, melancholia (depression) and mania that are expressed to different degrees in affected individuals, dominance and submissiveness are also two contrasting behavioral poles distributed as a continuum along an axis with less or more dominant or submissive animals. The importance of the selection process for these behavioral extremes will be underlined as a necessary factor for successful modeling of mania by dominance and depression by submissiveness.

Divergent patterns of social behavior result in rejection and reduced social reinforcement

A.J. Bond and W.S. Tse

Social skills are important for integration into society. These skills can be defined as the ability to acquire and use behaviors necessary for effective and satisfying interpersonal functioning and include both non-verbal and verbal elements. Substantial evidence has found that impaired social functioning is correlated with current depression. Two psychosocial models of depression, Coyne's social interaction model, and Lewinsohn's social skill model, have been proposed to explain the mechanism of depression on social integration. In general, both models agree that depression weakens social skills, resulting in reduced social support. To elucidate these mechanisms, non-verbal and verbal behaviors involved in interpersonal interaction can be studied. Untreated depressed patients are less active in engaging in social interaction, as indicated by paucity of speech and increased response latency and they participate less, as indicated by lack of eye contact in interpersonal exchanges. This behavior can lead to rejection by others. Much less work has examined the consequences of manic behavior on social integration. Therefore these two behaviors, depression which can be viewed as passive or submissive and mania which can be viewed as active or dominant were modeled in the laboratory. A standard behavioral measurement protocol in which interpersonal interaction style can be classified was constructed and resulted in 4 different styles. Two of the roles portrayed abnormal social behavior, active non-participant 'manic' and passive non-participant 'sad', and two portrayed normal social behavior, active participant 'warm' and passive participant 'shy'. The interaction of subjects with a confederate acting these 4 roles and their subsequent behavior on a dyadic game was recorded. Subjects were more likely to reject confederates in the manic or sad roles. This was shown by both non-verbal behavior and verbal report. They also behaved more punitively on the dyadic game. Thus displaying depressed or manic behavior not only results in social rejection but also in the reduction of social reinforcement.

Ethological analysis of rodent behaviour: elucidation of the behavioural effects of psychotropic drugs

P.J. Mitchell and P.H. Redfern

A wide diversity of animal models has been used to examine psychotropic drug activity. In recent years antidepressant drug research has focused on the search for new therapies with a rapid onset of action. It follows that, to be relevant, animal models must have the ability to measure the time course of drug-induced changes in behavior. Highly sophisticated animal models have been developed which yield a positive behavioral response to prolonged, chronic, drug treatment.

Two 'ethologically-relevant' animal models, the resident-intruder and social hierarchy paradigms, have been especially useful in elucidating the behavioral effects of antidepressant drugs. In the resident-intruder paradigm, male Wistar resident rats are housed in isolation for a minimum of 3 days before being exposed to an unfamiliar conspecific intruder. During the ensuing social encounter, control resident rats exhibit a wide range of non-social, social and conflict-related (i.e. agonistic) behaviors which are quantified during subsequent ethological analysis. In the social hierarchy model male Wistar rats are housed in triads. All group members are routinely involved in intense levels of social and agonistic behavior at the onset of the dark phase of the light:dark cycle. Ethological analysis of such behavior (where the 'winner' and 'loser' of each social encounter is identified) reveals the relative social position of each group member (the most successful group member during these encounters indicates the dominant animal).

Together these models of rodent social and agonistic behavior have demonstrated that chronic treatment with antidepressant drugs (irrespective of their acute pharmacological activity) increases rodent aggressive behavior which, in turn, results in increased hierarchical status in closed social groups. Furthermore, the increased rodent aggression is most likely a behavioral manifestation of increased assertive behavior and arguably reflects similar changes in human behavior (including the externalization of emotions) expressed during the recovery from depressive illness.

Monitoring the effects of social defeat in mice by automated observation in the home-cage and observer-based scoring during a social-interest test

J.E. van der Harst, M. Lubbers, M. Eijkhoudt and B.M. Spruijt

Social defeat in mice has been applied in several different paradigms and for several different reasons. However, as a model for depression it has only been studied on a few occasions. Mostly a long period of daily defeats is applied of which the effects are investigated immediately afterwards. Therefore, little is known about the time of onset and development of the effects of this chronic social stress. This study aimed to investigate the effects of a particular defeat-paradigm in mice during the long-term social-stress period of 20 days. For this, home cage behavior was automatically and continuously recorded during both the light and the dark phase in specially designed cages (PhenoTyper®, Noldus Information Technology bv, The Netherlands). Several parameters that may reveal depressive-like symptoms such as altered locomotor activity and sleep/wake cycle were analyzed.

It became apparent that social defeat resulted in several acute effects (thus, at day 1) on activity, velocity and use of the shelter that were not all persistent over time. Other measures, such as a decreased frequency of movement and increased time spent in the shelter during the first part of the dark-phase seemed to become more persistent over time indicating chronic stress effects, possibly related to the development of depressive-like symptoms.

To further validate the applied defeat-paradigm as a model for depression, a so-called Partition-test was conducted at the end of the long-term defeat-period. During this test the defeated mouse was confronted with another unfamiliar mouse at the other side of a perforated partition-wall. This test is used in several depression-studies to investigate anxiety-related behavior and social interest. Using observer-based scoring (The Observer®, Noldus Information Technology bv, The Netherlands), 2 classes were investigated in this test: zone and behavior.

It appeared that defeated animals displayed a significant decrease in social interest, activity and exploration and an increased alertness.

The results are discussed in the light of the success of the applied social-stress paradigm as a model for depression in mice and the development and onset of the symptoms.

The impact of continuous variation in heritable personalities on the social structure in the great tit (*Parus major*)

P.J. Drent

Great tits of both sexes show continuous variation in consistent phenotypically individual differences in exploration of a standard new environment (a gradual variation from fast to slow explorer). Although the absolute values of repeated tests varied with the year cycle, the inter-individual differences persist across time. This ‘exploration score’ is phenotypically correlated with many other behavioral traits related to coping with (environmental) challenges (e.g. boldness, risk-taking, aggressiveness, routine-formation, foraging patterns). Bi-directional selection and crossings experiments using a cross fostering design with guest-pairs show that these different behavioral traits are strongly genetically correlated. This all indicates a more general behavioral syndrome or coping strategy within the life history of the species, comparable with the variation in human personality.

Hand-reared and wild birds were used in an array of experiments to study the impact of these personalities on the composition, structure and hierarchy in winter flocks with a scrounger producer character. The dominant-submission interactions between the members of a group were standard scored on and around a feeding table. Males dominated females. Males with territorial status and for females mating with a territorial male have the highest position in the rank. In the hierarchy of territorial males the nearby the territory the higher the position in rank and within that: faster explorer dominated slower ones. In mixed groups of age and status classes the time of presence and the personality determined the rank between non-territorial males whereby in contrast to territorial birds long present slow explores dominated fast ones. This is caused by impact of actions of (old) territorial males on the non territorial juveniles that is different for the different personalities.

Reduction of dominant or submissive behavior as models for antimanic or antidepressant drug testing: technical considerations

E. Malatynska, A. Pinhasov, J. Crooke and D.E. Brenneman

Using observer-based scoring we have previously shown that dominant behavior measured in a food competition test, can serve as a model of mania and submissive behavior as a model of depression. These two models are based on a selection of animal pairs where one animal shows the behavioral trait of dominance and another submissiveness. Three criteria have to be achieved to assign dominant or submissive status to the animal. First, there has to be a significant difference between the average daily drinking scores of both animals in a pair.

Second, the dominant animal score has to be at least 25% greater than the submissive animal's score. Third, there must not be any 'reversals' during the pair selection week, where the putative submissive rat out-scores its dominant partner on isolated occasions. Twenty-five to thirty-three percent of the initial animal pairs achieve these criteria. The importance of the application of these criteria to the selection process as a reflection of the experimental outcomes will be discussed in this presentation.
Recently we have used automatic scoring of the time spent by rats in the feeder zone done by a multiple subject video-tracking system. We have noticed a similar reduction of rat submissive behavior after treatment with antidepressants, imipramine or fluoxetine as we have done previously with observer-based scoring. It is possible to observe four pairs of rats during each five-minute experimental session (one set). A duplicate parallel set enables the immediate switch to the observation of the next four pairs. The multiple video-tracking system increases the capacity of antidepressant drug testing and reduces the variability between observations.
The onset time of the drugs in the two models was delayed. The mechanism of delayed therapeutic activity of antimanic and antidepressant drugs is not known and animal models reflecting this clinical feature will enable progress in this area. The application of the reduction of submissive behavior model to such studies is also discussed.

Divergent patterns of social behavior result in rejection and reduced social reinforcement

A.J. Bond [a], W.S. Tse [b]
[a] Division of Psychological Medicine, Institute of Psychiatry, Kings College London, UK
[b] Department of Applied Social Studies, City University of Hong Kong, Hong Kong

Abstract
In this paper we describe a laboratory technique to model social behavior typical of mania and depression and present the responses of healthy subjects to this behavior in dyadic interactions. Interpersonal interaction style can be characterized by 2 dimensions, participation and activity. Both mania and depression involve low participation but they differ on activity. Active or dominant behavior is characteristic of mania and passive or submissive behavior of depression. Two normal interaction styles involving high participation but either active 'warm' or passive 'shy' behavior were constructed as comparisons. The non-verbal responses of the subjects to the 4 different interaction styles were measured during the interaction and their mood, rejection rating and behavior on a cooperative game after it. Manic or sad behavior did not affect the mood of healthy subjects but resulted in non-verbal behavior and ratings indicative of rejection as well as reduced social reinforcement on the game. This model can be used to test the effects of drug treatments.

Keywords
Social interaction, non-verbal behavior, participation, activity, rejection.

1 Introduction
Both verbal and non-verbal behavior are important in social interactions. It is not just what people say but the whole manner in which they interact which can lead to social acceptance or rejection. Non-verbal behaviors such as a flat tone of voice, a sad facial expression and avoiding eye contact can all increase the likelihood of social rejection [1,3]. These behaviors are typical of depression or submissive behavior and healthy volunteers have reported being more likely to avoid future contact with depressed subjects [2]. However, another pattern of behavior consisting of talking loudly, interrupting and not looking is associated with mania or dominance and this may also result in rejection. Rejection has generally been measured by rated desire for future contact but these patterns of behavior may also have consequences for social reinforcement during subsequent unavoidable interactions.

The present study was planned to measure the reactions of healthy volunteer subjects to a confederate trained to enact 4 distinct roles both during and subsequent to a dyadic social interaction. Two dimensions of behavior were studied, participation and activity. The two normal roles were both high on participation but differed on activity, low activity indicating shyness and high activity indicating warmth. The two abnormal roles were both low on participation but again differed on activity, low activity representing depression and high activity representing mania. Thus the active-passive dimension could be seen as a human representation of preclinical dominant-submissive behavior. Behavior during the interaction was videoed and immediately after the interaction, subjects took part in a cooperative task. Rejection was measured by rated desire for future contact and social reinforcement was measured by behavior during the game. It was hypothesized that subjects would reject the confederates enacting the non-participant roles and would display this not only by their reduced desire for future contact but also by both their non-verbal responses during the interaction and their subsequent behavior on the game.

2 Methods

2.1 Subjects
Sixty-three healthy volunteers (32 men, 31 women) aged between 18 and 60 years were recruited from the local community. The study was approved by the institutional ethical committee and subjects gave written informed consent on arrival. Subjects were randomly assigned to interact with a same-sex confederate in one of 4 experimental conditions.

2.2 Social interaction
Subjects were told they were going to play a game with another volunteer (confederate). They were first taken to a waiting room, equipped with various items of furniture but only 2 available seats. These were positioned adjacent to each other but separated by a small table. The confederate was already sitting on the chair furthest from the door. A small camera was hidden in one of a number of box files, on top of a filing cabinet on the opposite side of the room. None of the subjects detected it. Video recording started when the subject entered the room and continued for 3 minutes of interaction. The experimenter was able to monitor the equipment and observe the interaction on a large screen in an adjacent room. The sex-matched confederates enacted one of the 4 roles.

2.3 Confederate role
One male and 1 female confederate aged 22 years were recruited from a University Drama Department. They were blind to the experimental hypotheses. There were 4 roles, based on the 2 dimensions of activity and participation, in which they were trained for 6 hours.

- Active participant (warm) role – the confederate was instructed to respond to the subject's conversation quickly with relevant topics; to smile frequently; to engage in a lot of eye contact during conversation; to speak freely and initiate topics relevant to the subject.
- Passive participant (shy) role - the confederate was instructed to respond slowly with relevant topics; to smile; to engage in eye contact during conversation; to

speak little; to rarely initiate conversation but to keep topics relevant to the subject.

- Active non-participant (manic) role - the confederate was instructed to respond to the subject's conversation quickly with irrelevant topics; to smile little; to avoid eye contact during conversation; to speak a lot on topics relevant to themselves; to speak little on topics relevant to the subject.
- Passive non-participant (sad) role – the confederate was instructed to respond to the subject's conversation slowly; not to smile; to avoid eye contact during conversation; to speak very little (single words); never to initiate conversation.

The validity of the role performance was checked by 3 independent raters. Interactions where 1 rater failed to identify the correct role were discarded (N=4).

2.4 Social Behavior Variables

Five behavioral variables were measured from the videos:

- SPEECH: The total duration of speech in seconds was measured.
- GAZE: The proportion of speech by the subject with eye contact was calculated by measuring the total duration of eye contact while the subject was speaking divided by the total duration of speech.
- NOGAZE: The proportion of speech by the subject without any eye contact was calculated by summing the duration of each sentence spoken by the subject during which the subject showed no eye contact at all divided by the total duration of speech.
- LOOK: The proportion of subject eye contact during confederate speech was calculated by measuring the total duration of eye contact while the confederate was speaking divided by the total duration of his/her speech.
- NOLOOK: The proportion of no eye contact during confederate speech was calculated by summing the duration of each sentence spoken by the confederate during which the subject showed no eye contact at all divided by the total duration of speech by the confederate.

2.5 Mixed-Motive Game

This is a modified version of the Prisoner Dilemma game [4]. There were 60 trials divided into 4 blocks. Between each block, subjects first gave and then received feedback from their partner via a communication checklist. Each trial involved selecting a particular scheme of pay-off, which indicated how points were being distributed to each of two players. The 1st choice was determined by Player 1 (confederate) who had to choose 1 of 3 possible pay-off schemes, each of which had 3 possible options. The 3 schemes were in fact selected equally over the whole game and so were constant for all participants. The 2nd choice was determined by Player 2 who selected one option from the set of 3 chosen by Player 1. Player 2 had the higher power to determine how many marks to allocate to Player 1 and him/herself. The subject was thus able to give the confederate more, less or the same number of points as him/herself.

2.6 Scales

The Positive and Negative Affect Scale (PANAS) measures current mood. It was completed pre and post the social interaction. The Post Encounter Scale (PES) is a 7-item measure of desire for future contact with a specified person. It was completed after the interaction. The Communication Checklist (CCL) is a 24-item inventory, which measures 6 different types of communication: extrapunitiveness, cooperativeness, ingratiation, sadness, blaming partner, helplessness. Subjects completed the CCL online to send to their partner on 3 occasions during the game. They received a CCL completed by their partner (confederate) immediately afterwards each time. This feedback was kept neutral and constant for all subjects.

2.7 Procedure

On arrival, subjects gave written informed consent and filled in the PANAS. They were then told they were going to play a game with another volunteer (confederate). They were taken to a waiting room where they were asked to wait while the experimenter checked everything was ready. The confederate was already sitting in the room and they were left to interact. After 3 minutes the experimenter returned and took the subject to the game room, which was equipped with a computer. The subject filled in the PANAS and the PES. They were then given instructions for the game and were told that they would be playing the game with the volunteer they had just met (confederate). At the end they were debriefed and written permission was sought to use the video recording.

3 Results

3.1 Subjects

The final number of subjects was 59 with a mean age of 32.6 ± 10.1 years. Each experimental group contained 7-8 females and 7-8 males. There were no significant differences between males and females or between groups on age or mood pre interaction.

3.2 Mood

Confederate role had no effect on the positive or negative mood of the subjects.

3.3 Social behavior

The effects of the confederate role on the subjects' behavior were examined with ANOVA. The mean values are shown in Table 1. There were significant effects on SPEECH ($F_{3,55} = 41.3$, $p<0.001$) and NOGAZE ($F_{3,55} = 4.4$, $p<0.01$). In multiple comparison analysis, subjects paired with a warm confederate spoke more ($t_{55} = 9.6$, $p<0.01$) than subjects in any of the other conditions. They also looked away less when they were speaking ($t_{55} = 3.6$, $p<0.01$) than either the active or passive non-participant confederates (manic and sad). Subjects paired with a manic confederate spoke less ($t_{55} = 5.2$, $p<0.01$) and looked away more while they were speaking ($t_{55} = 2.8$, $p<0.01$) than subjects in the participant conditions (warm and shy). Subjects paired with a sad confederate spoke less ($t_{55} = 12.3$, $p<0.01$) than subjects in all the other conditions.

	Confederate role			
Behavior	**Warm** N=15	**Manic** N=15	**Shy** N=15	**Sad** N=14
SPEECH*	111.5 (26.3)	30.8 (15.5)	68.7 (44.0)	6.9 (9.2)
GAZE	0.53 (0.18)	0.39 (0.20)	0.51 (0.20)	0.52 (0.43)
NOGAZE*	0.12 (0.14)	0.43 (0.21)	0.19 (0.25)	0.39 (0.45)
LOOK	0.54 (0.18)	0.38 (0.27)	0.52 (0.24)	0.60 (0.47)
NOLOOK	0.15 (0.15)	0.11 (0.13)	0.22 (0.27)	0.31 (0.46)

***Table 1**. Means (SD) for the 5 types of social behavior of the subjects interacting with the confederate in the 4 different roles.*

3.4 Rejection

There was a significant effect of confederate role on the Post Encounter Scale ($F_{3,55}$ = 6.3, p<0.001). Rejection scores were higher for the non-participant (abnormal) roles. The subjects were significantly more likely to reject the confederate in the manic and sad roles than in the warm role (p<0.01) and showed a similar trend compared to the shy role (p<0.06). Rejection scores did not differ between the warm and shy roles, nor between the manic and sad roles.

3.5 Social behavior and Rejection

Removal linear multiple regression was conducted to investigate the possible role of social behavior as an indicator of rejection. The 5 social behavior variables were selected as independent factors and the PES as the dependent factor. Subjects who had a high proportion of speech without any eye contact (β = 0.5; t_{53} =2.0, p<0.05) or who spoke little (β = -0.4; t_{53} = -2.5, p<0.05) were more likely to reject the confederate.

3.6 Consequences of Rejection

Generalized linear modeling (GLM) was used to test the hypothesis that rejection led to punitive behavior on the Mixed Motive game. The PES was selected as the independent factor and scores awarded to player 1 (confederate) and player 2 (subject) on the game as the dependent factors. The PES was associated negatively with points given to the confederate (β = -0.4; $F_{1,54}$ = 11.9, p<0.001) but not with points earned by the subject. Rejection was associated with awarding the confederate lower points on the game.

GLM was also used to test the relationship between rejection and communication during the game. The PES was the independent factor and the 6 dimensions of the 3 communication checklists (CCL) were the dependent factors. On all 3 occasions, cooperativeness and ingratiation showed significant negative relationships with the PES (p<0.001). Rejection was associated with reduced cooperative and ingratiative communication.

4 Discussion

This paper describes a new experimental model. The stranger-dyadic social interaction paradigm was used to examine the effects of two dimensions of social behavior, participation and activity, on the responses of healthy human subjects. Participation divided normal from abnormal behavior and activity denoted different levels of these two behavior types. This resulted in 4 roles enacted by a confederate; warm, shy, manic and sad. The effects of the interaction were examined on non-verbal behavior during the interaction, affect immediately afterwards, and attitudes and behavior towards the confederate subsequent to the interaction.

Divergent social behavior displayed by a confederate during a brief dyadic interaction had no effects on mood but profound effects on the social behavior of healthy volunteers. Not only did they reject the confederate enacting these roles by expressing less desire for future contact but this rejection was also shown in their non-verbal behavior during the interaction and the reduced social reinforcement which they gave on a subsequent game.

These results applied to both non-participant roles, sad and manic. Previous experiments which have used a stranger-dyadic social interaction paradigm to study social behavior have concentrated on only a sad condition [9]. The sad role can be equated with submissive behavior, which has been used as a model of depression in preclinical work. However depression, although common, is only one type of abnormal behavior, characterised by non-participation and passivity. Mania is another type of abnormal behavior, which also involves little participation but is high on activity. The manic role may be similar to dominant behavior in preclinical work. The sad role reduced subjects' speech during the interaction but the manic role had more effect as it not only resulted in less speech but also in reduced eye contact. These results therefore extend the interpersonal consequences of non-verbal behavior typical of depression to other distinct forms of abnormal behavior e.g. hypomania. It was not possible to detect any differences in the consequences for depressed and manic behavior in this experiment but future experiments with larger sample sizes may be able to do this.

Previously used role enactment paradigms have measured rejection using a rating scale (PES). The present study confirmed previous results [8] in showing rejection of the sad confederate. In addition, the manic confederate was also rejected, indicating that non-participation is a key factor as predicted. Rejection was mediated by abnormal social behavior during the interaction. Previous studies have not attempted to examine behavior subsequent to the interaction. However, people may not always be able to avoid others displaying abnormal behavior and so it is important to know how they might behave towards them and the consequences for the 'patient'. The current study therefore examined the behavior of the subjects towards the confederate during a cooperative game following the interaction. Rejection was associated with both awarding the confederate fewer points on the game and sending them less cooperative and ingratiating messages. Cooperative communication might be seen as affiliative and a sign of acceptance and whereas ingratiating communication might be seen as flattery and indicate submissiveness. The subjects were therefore more punitive and less prepared to accept or encourage someone behaving differently.

The methodology described in this paper has also been used to examine the effects of antidepressants in healthy volunteers. Citalopram (20mg/d) and placebo were administered for 2 weeks in a double-blind, crossover trial [10]. The social interaction, with a warm confederate, took place on the final day of treatment and the subjects' social behavior was evaluated by their flatmate before and at the end of treatment. On citalopram, subjects increased eye contact during their own speech and decreased it during confederate speech during the interaction. This represents a dominant pattern of eye contact. They also reduced the number of points they awarded themselves on the game and sent more cooperative messages to the confederate. In addition, they were also rated as less submissive by their flatmates. This result lends support to the link found between serotonin and dominance in preclinical work [7] and confirms other findings showing that tryptophan can increase self-rated dominance during daily interactions [5] and paroxtine can increase affiliative behavior on a collaborative dyadic puzzle task [6]. However, in a second study, reboxetine (8mg/d) and placebo were administered for 2 weeks in a double-blind, crossover trial using exactly the same methodology [11]. When on reboxetine, subjects showed less eye contact when the confederate was speaking and sent fewer helplessness messages during the game. They were also rated as less submissive by their flatmates. Therefore antidepressants acting on both serotonin and noradrenaline may improve submissive behavior and increase assertiveness but their actions on specific social behaviors may differ. This model can help to delineate these effects.

This experimental model shows that interpersonal style of behavior is important for social acceptance and, in particular, that behaving in a manic or depressed way increases the likelihood of rejection and subsequent reduced social reinforcement. This may have important implications for the social integration of psychiatric patients in the community. The model could be adapted to measure other styles of interaction.

Supported by the UK Medical Research Council.

References

1. Argyle, M.; Lefebvre, L.; Cook, M. (1972). The meaning of five patterns of gaze. *European Journal of Social Psychology*, **4**, 125-136.
2. Coyne, J.C. (1976). Depression and the response of others. *Journal of Abnormal Psychology*, **85**, 186-193.
3. Gurtman, M.B. (1987). Depressive affect and disclosures as factors in interpersonal rejection. *Cognitive Therapy and Research*, **11**, 87-99.
4. Hokanson, J.E.; Sacco, W.P.; Blumberg, S.R.; Landrum, G.C. (1980). Interpersonal behavior in depressive individuals in a mixed-motive game. *Journal of Abnormal Psychology*, **89**, 320-332.
5. Moscowitz, D.S.; Pinard, G.; Zuroff, D.C.; Annable, L.; Young, S.N. (2001). The effect of tryptophan on social interaction in everyday life: a placebo-controlled study. *Neuropsychopharmacology*, **25**, 277-289.
6. Knutson, B.; Wolkowitz, W.M.; Cole, S.W.; Chan, T.; Moore, E.A.; Johnson, R.C.; Terpstra, J.; Turner, R.A.; Reus, V.I. (1998). Selective alteration of personality and social behavior by serotonergic intervention. *American Journal of Psychiatry*, **155**, 373-379.
7. Raleigh, M.J.; McGuire, M.T.; Brammer, G.L.; Pollack, D.B.; Yuwiler, A. (1991). Serotonergic mechanisms promote dominance acquisition in adult vervet monkeys. *Brain Research*, **559**, 181-190
8. Sacks, C.H.; Bugental, D.B. (1987). Attributions as moderators of affective and behavioral responses to social failure. *Journal of Personality and Social Psychology*, **53**, 939-947.
9. Segrin, C.; Dillard, J.P. (1992). The interactional theory of depression: A meta-analysis of the research literature. *Journal of Social and Clinical Psychology*, **11**, 43-70.
10. Tse, W.S.; Bond, A.J. (2002). Serotonergic intervention affects both social dominance and affiliative behaviour. *Psychopharmacology*, **161**, 324-333.
11. Tse, W.S.; Bond, A.J. (2003). Reboxetine promotes social bonding in healthy volunteers. *Journal of Psychopharmacology*, **17**, 65-71.

.

Ethological analysis of rodent behavior: elucidation of the behavioral effects of psychotropic drugs

P.J. Mitchell, P.H. Redfern
Department of Pharmacy & Pharmacology, University of Bath, Bath, UK

A wide diversity of animal models has been used to examine psychotropic drug activity. In recent years antidepressant drug research has focused on the search for new therapies with a rapid onset of action. It follows that, to be relevant, animal models must have the ability to measure the time course of drug-induced changes in behavior. Highly sophisticated animal models have been developed which yield a positive behavioral response to prolonged, chronic, drug treatment [6]. In particular, two 'ethologically-relevant' animal models, the resident-intruder and social hierarchy paradigms, have been especially useful in elucidating the behavioral effects of antidepressant drugs [5].

In the resident-intruder paradigm [2], male Wistar resident rats are housed in isolation for a minimum of 3 days before being exposed to an unfamiliar conspecific intruder. During the ensuing social encounter, control resident rats exhibit a wide range of non-social, social and conflict-related (i.e. agonistic) behaviors [1] which are quantified during subsequent ethological analysis [2]. During ethological analysis of rodent behavior the occurrence of each behavior or posture exhibited by the resident rats during each social encounter are identified and recorded during video playback. The scores for each behavior/posture are grouped according to their motivational category [5] for each animal and the total score for each category expressed as a percentage of the total number of behaviors observed for that subject. Data from the two groups of four resident rats are collated and the mean and standard error of the mean are calculated for both the percentage values of each motivational category and the total number of behaviors/postures observed within each treatment group. In a typical 10 min social encounter resident rats will exhibit between 1200 and 1500 behavioral elements. Invariably resident rats exhibit high levels of investigatory behavior (about 50-60% of all behaviors scored) directed at the intruder conspecific following introduction of the latter into the resident rat's home cage. In some cases such intense conspecific investigation leads to the expression of aggressive behavior directed at the conspecific (about 5-8% of total behaviors) or flight behavior (total flight behavior equals about 15-20% of total behaviors scored). Flight behavior in the rat may be subdivided into two flight pathways; flight submit and flight escape. In these studies the occurrence of flight escape behavior invariably far exceeds flight submit behavior (12-14 % c.f. 1-2%) and reflects the opportunity to flee afforded by the size of the cage in which the social encounter takes place. The remaining categories of behavior (maintenance and sexual behavior) occur very infrequently (about 1% of total behaviors scored in both cases). In most studies performed in our laboratory resident rats are exposed to an unfamiliar conspecific on four occasions at weekly intervals irrespective of any acute or chronic drug treatment regime used. However, by programming daily dyadic encounters, the resident-intruder paradigm can also be used to compare the rate at which chronic treatment with psychotropic drugs induce changes in agonistic behavior and to assess the ability of potential adjuvant treatment to accelerate any desirable change in behavior indicative of potential clinical utility [4].

The resident-intruder paradigm coupled with ethological analysis of the resulting non-social, social and agonistic behavior quite clearly shows that antidepressant drugs have very specific effects on rodent behavior that are dependent on the duration of drug treatment [5]. Thus, while acute treatment with antidepressant drugs selectively reduces aggressive behavior, chronic treatment increases aggression. In both cases such changes in aggressive behavior are associated with reciprocal changes in flight behavior.

Observations of the social behavior of grouped rats throughout their light-dark cycle indicate that intense levels of social and agonistic behavior, involving all group members to an equal extent, routinely occur at the onset of the dark phase and appear to indicate a fixed pattern of behavior that enable the re-establishment of the hierarchical structure. Such behavior precedes any grooming or consummatory behavior and may be analyzed using ethological techniques. In the social hierarchy model [3] male Wistar rats are housed in triads (i.e. n=3). Each social interaction between the grouped rats during the initial 30 min of the dark-phase is carefully monitored and the 'winner' and 'loser' of each social interaction recorded. The 'loser' is identified as the rat which adopted the final posture(s) of either the Flight Submit or Flight Escape behavioral pathways (i.e. Submit posture, or Flag and Evade or Retreat, respectively) in response to the aggressive posturing of a cage partner (which was therefore identified as the 'winner'). In some instances the social interaction is terminated by the loser exhibiting Flight Escape behavior prior to any Approach by a more dominant cage partner. In these situations the result of the social encounter is recorded provided the winner and loser may be clearly identified. Interactions where no winner or loser is unequivocally identified are ignored. The relative success level attained by each group member during social encounters is then calculated by expressing the total number of wins as a percentage of the total number of encounters in which that animal was involved. The highest success value indicates the dominant animal, the next highest the subdominant animal, and the lowest success value indicates the subordinate.

The ability of chronic antidepressant treatment to increase aggression in rats, as observed in the resident-intruder paradigm, is a measure of increased assertiveness in this species. This may reflect the increased assertiveness and associated externalization of emotions expressed during recovery from depressive illness in the clinic. Such increased assertive/aggressive behavior is consistent with the effects of such treatment in the social hierarchy model [3,5]. However, it must be acknowledged that the face validity of both the resident-intruder and social hierarchy models is slightly reduced by the fact that these tests use normal animals, whereas non-depressed people do not respond to antidepressant treatment. Together these models of rodent social and agonistic behavior have demonstrated that chronic treatment with antidepressant drugs (irrespective of their acute pharmacological activity) increases rodent aggressive behavior which, in turn, results in increased hierarchical status in closed social groups. Furthermore, the increased rodent aggression is most likely a behavioral manifestation of increased assertive behavior and arguably reflects similar changes in human behavior (including the externalization of emotions) expressed during the recovery from depressive illness [5,6].

Of particular interest is the time course of response to chronic antidepressant treatment in both models. The importance of this feature should not be underestimated since only when a model shows a gradual response that reflects a drug's gradual onset of action is it possible to detect a more rapid onset. In this respect the resident-intruder and social hierarchy paradigms can legitimately claim improved face and predictive validity, since the increase in aggressive behavior during chronic treatment observed in these models can be seen as reflecting the increased assertiveness and drive in patients when the symptoms of endogenous depression begin to lift. Equally important is the observation that these behavioral effects are seen at clinically relevant doses that do not produce other, potentially confounding effects on behavior [5,6].

References

1. Grant, E.C. (1963). An analysis of the social behaviour of the male laboratory rat. *Behaviour*, **21**, 260-281.
2. Mitchell, P.J.; Redfern, P.H. (1992a). Acute and chronic antidepressant drug treatments induce opposite effects in the social behaviour of rats. *J. Psychopharmacol.*, **6**, 241-257.
3. Mitchell, P.J.; Redfern, P.H. (1992b). Chronic treatment with clomipramine and mianserin increases the hierarchical position of subdominant rats housed in triads. *Behav. Pharmacol.*, **3**, 239-247.
4. Mitchell, P.J.; Redfern, P.H. (1997). Potentiation of the time-dependent, antidepressant-induced changes in the agonistic behaviour of resident rats by the 5-HT1A receptor antagonist, WAY-100635. *Behav. Pharmacol.*, **8**, 585-606.
5. Mitchell, P.J.; Redfern, P.H. (2005). Animal models of depressive illness: The importance of chronic drug treatment. *Curr. Pharamaceut. Des.*, **11**, 171-203.
6. Willner, P.; Mitchell, P.J. (2002). Animal models of depression: A diathesis/stress approach. In: D'Haenen, H.A.H.; Den Boer, J.A.; Willner, P. (Eds.), *Biological Psychiatry*, Vol. 2. Wiley, Chichester, pp. 703-726.

The impact of continuous variation in heritable personalities on the social structure in great tit (*Parus major)*

P.J. Drent
Netherlands Institute of Ecology (NIOO-KNAW), Heteren, The Netherlands

Abstract

Great tits of both sexes show continuous variation in consistent phenotipically individual differences in exploration of a standard new environment (a gradual variation from fast to slow explorer). Although the absolute values of repeated tests varied with the year cycle, the inter-individual differences persist across time. This 'exploration score' is phenotipically correlated with many other behavioural traits related to coping with (environmental) challenges (e.g. boldness, risk-taking, aggressiveness, routine-formation, foraging patterns). Bi-directional selection and crossings experiments using a cross fostering design with guest-pairs show that these different behavioural traits are strongly phonotypical and genetically correlated. This all indicates general differences in behavioural rules between individuals resulting in differences in behavioural syndromes or coping strategies that can play an important rule on individual variation in life-history traits within the species, comparable with the variation in human personality.

Hand-reared and wild birds were used in an array of experiments to study the impact of these personalities on the composition, structure and hierarchy in winter flocks with a scrounger producer character. The dominant-subordinate interactions between the members of a group were standard scored on and around a feeding table. Males dominated females. Males with territorial status and for females mating with a territorial male have the highest position in the rank. In the hierarchy of territorial males the nearby the territory the higher the position in rank and within that: faster explorer dominated slower ones. In mixed groups of age and status classes the time of presence and the personality determined the rank between non-territorial males whereby in contrast to territorial birds long present slow explores dominated fast ones. This is caused by the impact of actions of (old) territorial males on the non-territorial juveniles that is different for the different personalities

Keywords

Personality, dominant-subordinated relationship, territorial status, resident-intruder interactions, social hierarchy, great tit

1 Introduction

Social dominance-subordinate behavior and by experience ritualized dominance-subordinate relationships and hierarchies is common in many animal phyla including fish, reptiles, birds and mammals including humans. First the question is, how clearly defined and measured dominance and subordination in animals as our model species the great tit, a small territorial, hole breeding songbird from wooden areas in Europe, that outside the breeding season regularly lived in flocks.

Dominance and submissiveness are relative terms that define the outcome of encounters between individuals and the social position or status of one individual in a society toward others. Van Kreveld [1] defined dominance as "mutually respected rights one group member has over another". This definition is general and does not clarify whether the dominance of one individual over other individual(s) occur on one, few or in all contexts and locations.

2 Dominant - submissive behaviour in the great tit

2.1 Measuring dominant-subordinate behaviour

Dominance–subordination relationships measured on clumped food resources are the results of the outcomes of a series of encounters between two individuals in terms of victories and defeats. Sex, age, morphological and physiological characteristics [3,4,5,10], personality of the birds [5,8,9], status and earlier experiences [9] and the interference of other birds played [5,9,10] an important role in the final relationship of normally "peck-right type [8,9,10]. If the differences between the two birds in these characteristics were the same over the whole used area the dominance-subordinated relationship was found everywhere. However, if the differences in these characteristics between the two individuals differ between locations (differences in context as property of a roosting site or a territory and different birds that interfere in encounters), the dominance relationship can change with location. As long as the context remained the same the relationship was stable [10, 11]. In case of changes of characteristics in time (for instance change in territorial status or shifts in the location of the territory) the relationships can also change taken into account a transition period [10]. The relationships tended to be more clear cut and 'peck-right' if only encounters were used in which both birds showed agonistic behaviour than in the case based on avoidance. Differences in the personality that comes about in differences in exploration and aggressiveness also strongly influences the outcome of encounters and thereby the development of dominant – subordination relationships. In a dyadic tournaments "fast explorers" (FE) dominated almost always "slow ones" (SE) [8] of the same sex, age and experiences" When the age and experiences were markedly different this personality effect can be overruled by these differences [9].

2.2 dominant- subordination hierarchies

These relationships between two individuals were used in computing hierarchical structures. On each location these hierarchies mostly are linear at least in the upper segment. Inconsistencies and triangular relationships sometimes occurred and could be related to characteristics of the birds concerned [10]. In principle all males defeated all females while adult birds normally defeated young ones in their first period of life [9,10]. Moreover individuals already presence over longer time (more experienced) dominated recent arrived ones of the same sex and age. In flocks in captivity the above effects overruled differences in morphological characteristic and personality. In flocks of the same age and sex but different in genetic determined personality [6,7] a stable relationship established after a dynamic phase of several days, in which many reversals in dominance relationships occurred. During the first days, the situation was similar to that observed in tests with pair- wise confrontations [8]. After some days when the hierarchy had become stabilized, most of the FE birds were not find at the top but at the bottom of the hierarchy with SE birds above them [11]. So the initial advantage of the first blow by their higher aggressiveness is reversed by the higher alertness and faster increasing spatial knowledge of SE birds. They took advantage of the relative long recovery period of FE birds after a defeat. Only fast birds that win continually held their high rank.

In the field rank position in the flock on a certain location is strongly dependent on the property of a territory and roosting hole and the distance between location of observations and these properties. Territory owners dominated non territory-owners including birds that only own a roosting-site, the latter ones dominated non-owners [9]. In the first two categories how nearby the property is to the place of determining their rank, how higher the place in that rank. When in category of territorial males the difference in the distance to the territory is small and the birds came from different directions: “fast explorer” dominated “slower ones” [12]. In contrast to territorial birds long present “slow explores” in status mixed flocks dominated fast ones. This is caused by the impact of actions of (old) territorial males on the non-territorial juveniles that is different for the different personalities.

3 Summary

All Interpersonal Dyadic relationships are the consequence of social defeat in both Dominant-Submissive Behaviour influenced by sex, age, morphological and physiological differences and personality, and in Resident-Intruder Interactions determined by differences in status as consequence of difference in territory ownership and (local) experiences as described for rodents [2]. Particularly attention is focused on individual heritable behavioral rules or traits as part of animal personality that determines in a context dependent way dominant –submissive relationships. Using selection lines on components of the personality that an individual inherits the possibility of specific types of behavioral traits, which develops as the individual matures. There is also a certain amount of interplay between genetic and environmental factors later in life. The outcome of these behavioral rules in interactions and thus finally in the development of dominance-submissive relationships and ranks in flocks is phase (sex, age and territorial status) and context dependent. These hierarchies tended to be linear but the rank on different locations and in time are context dependent [10]. The mode of interaction may also influence physiological parameters and are under control of natural selection.

References

1. Van Kreveld, D. (1970). A selective review of dominance-subordination relations in animals. *Gen. Psychol. Monographs*, **81**, 143-173.
2. Blanchard, R.J.; Flannelly, K.J.; Blanchard, D.C. (1988). Life-span studies of dominance and aggression in established colonies of laboratory rats. *Physiol. Behav.*, **43**(1), 1-7.
3. Koolhaas, J.M.; Schuurman, T.; Wiepkema, P.R. (1980). The organization of intraspecific agonistic behaviour in the rat. *Prog. Neurobiol.*, **15**(3), 247-268.
4. Koolhaas, J.M., et al. (1990). Single social defeat in male rats induces a gradual but long lasting behavioral change: a model of depression? *Neurosci. Res. Comm.*, **7**, 35-4.
5. Carere, C., et al. (2001). Effect of social defeat in a territorial bird (Parus major) selected for different coping styles. *Physiol. Behav.*, **73**(3), 427-433.
6. Van Oers, K., et al. (2004). A genetic analysis of avian personality traits: correlated, response to artificial selection. *Behav. Genet.*, **34**(6), 611-619.
7. Drent, P.J.; van Oers, K.; van Noordwijk, A.J. (2003). Realized heritability of personalities in the great tit (Parus major). *Proc. Biol. Sci.*, **270**(1510), 45-51.
8. Verbeek, M.E.M., et al. (1996). Exploration, aggressive behaviour and dominance in pair-wise confrontations of juvenile male great tits. *Behaviour*, **133**, 945-963.
9. Drent, P.J., et al. (1999). Individuality, exploration and foraging in hand raised juvenile great tits. In: *Proceedings of the 22nd International Ornithological Congress, Durban*, Adams, N.J.; Slotow, R.H. (Eds.). Johannesburg: Birdlife South Africa, pp. 896-914.
10. Dent, P.J. (1983a). The functional ethology of the great tit Parus major L. (PhD) dissertation, Groningen: University of Groningen.
11. Verbeek, M.E.M, et al. (1999). Individual behavioural characteristics and dominance in aviary groups of great tits. *Behaviour*, **136**, 23-39.
12. Dingemanse, N.J.; De Goede, P. (2004). The relation between dominance and exploratory behaviour is context-dependent in wild great tits. *Behav. Ecol.*, **15**, 1023-1030.

Reduction of Dominant or Submissive Behavior as Models for Antimanic or Antidepressant Drug Testing: Technical Considerations.

E. Malatynska, A. Pinhasov, J. Crooke, and D.E. Brenneman.

Johnson & Johnson Pharmaceutical Research & Development, L.L.C, Spring House, PA 19477, U.S.A

Abstract

Using observer-based scoring we, have previously shown that dominant behavior measured in a food competition test can serve as a model of mania and submissive behavior as a model of depression. These two models are based on a selection of animal pairs where one animal develops the behavioral trait of dominance and another submissiveness after repeated interactions. The importance of the application of the three criteria to the selection process was reflected in the experimental results and discussed in this presentation. In automatic scoring the time spent by rats in the feeder zone is done by a multiple subject video-tracking system. We have shown that controls maintain established behavioral relation in both systems and that a reduction of rat submissive behavior after treatment with antidepressant, fluoxetine was similar to that measured previously with observer-based scoring. It is possible to score four pairs of rats during each five-minute experimental session (one set) by video tracking. The multiple video-tracking systems increase the capacity of antidepressant drug testing and reduce the variability between observations.

Keywords

Dominance, submissiveness, mania, depression, competition test.

1 Introduction

The Reduction of Submissive Behavior Model (RSBM) was developed as an animal model of depression based on previous work with the Clonidine-Reversal of Dominance Model (CRDM) [1-5]. Both models result from the hypothesis that submissive behavior in social animals is related to human depression. In the earlier CRDM, submissive behavior was induced by treatment with the α_2-adrenergic receptor agonist clonidine. The induction of submissive behavior by clonidine could be reversed by treatment with a wide range of antidepressant drugs, while many non-antidepressants did not have this effect [1, 5]. Thus the model has good predictive validity. However, the principle weakness of the CRDM is a lack of construct validity. This is seen in the subacute nature of the effect of clonidine and antidepressants in the CRDM. The CRDM is only as good a model of depression as clonidine treatment is. The CRDM was subsequently abandoned in favor of the RSBM. The RSBM does not require pretreatment with drugs to show antidepressant effects and provides a better model of time-dependent phenomena associated with depression [6]. We have shown that the antidepressants, fluoxetine, imiprmine, desipramine and maprotyline increased competitiveness of submissive rats while diazepam, naltrindole, and amphetamine did not.

The Reduction of Dominant Behavior model (RDBM) also originated from the CRDM. Clonidine, which is used in the clinic to alleviate episodes of acute mania, reduced dominant behavior in the CRDM. This observation prompted us to test the effects of other drugs used in the clinic to treat mania such as lithum, sodium valproate and carbamazepine. We have shown that all of these drugs reduced dominant behavior [7].

Dominance and submissiveness are two opposite poles of extreme behavior that otherwise are distributed as a gradual continuum along an axis with less and more dominant or less and more submissive animals. Less than half of animals studied in our tests form clear dominant-submissive relationships (DSR). The majority of animals form flexible relationships without dominant or submissive behavior. The process of animal selection is very important for the isolation of the relatively homogenous groups of dominant, submissive or neutral animals. In this paper we, have reviewed evolution of methods leading to our current formula defining behaviorally distinct groups.

2 Basic principles of the DSR method

2.1 Apparatus.

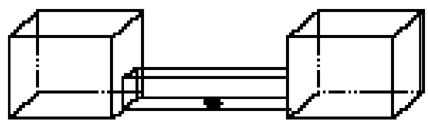


Figure 1. RSBM Apparatus. The RSBM apparatus consists of two plexiglass chambers connected by a passage having a small feeder dish with milk in the center from which only one rat can drink at a time.

A DSR is established and observed in a specially designed apparatus (figure 1) that was first constructed in the Institute of Psychiatry and Neurology (Warsaw, Poland) as a part of Dr. Malatynska's graduate work [1]. The apparatus is described in several publications [1-5]. This apparatus consists of two transparent plexiglass boxes (20 x 15x 12 cm) connected with a narrow hollway (8 x 8 x 40 cm). A feeder is placed in the middle of a hallway. The small top opening is located in the hallway ceiling above the feeder allows for easy milk filling.

The basic apparatus is the same through all changes in the study techniques with the exception of couple minor modifications that facilitate the study. First, the tunnel now has narrow holes cut on both sides of the feeder for easy gate insertion on the end and removal on the beginning of experiment. In this way, paired rats have an equal starting position at the beginning of the experimental session. Second, the 10 ml beacker (about 2 cm diameter) feeder was replaced with specially constructed self-refilling feeder with smaller diameter (1 cm).

2.2 Animals.

Either Wistar or Spraque-Dawley rats weighting 140-200 g were used in these experiments. Animals were housed in groups of 4 (RSBM) or 5 (CRDM) rats per cage. Animals in competing pairs were always housed separately. They only met each workday during the 5-min testing period. At the end of the 5-min. period, the animals were separated to individual cages and given free access to food for one hour. Thus, the food restriction time was about 23 hours a day during weekdays. The animals were given free access to food from Friday afternoon following testing to Sunday afternoon when they were once again food-deprived. Rats showed normal weight gain while on this feeding schedule.

2.3 Procedure used in the CRDM

On the beginning of the experiment, rats were individually placed in the apparatus for 15 min during first two experimental days (on Thursday and on Friday) for habituation purposes. On Monday of the subsequent week, rats were randomly paired and their time spent on the feeder was recorded during the 5-min. daily session. A point was assigned for each 5-sec period when a rat was drinking milk. At the end of the week, pairs of rats that differ significantly ($p < 0.05$ Student t-test) in their scores were judged to have established dominant-submissive relations. Using this method, about 80% of all pairs were selected as having dominant-submissive relations. In the following week (5 days), one group of dominant rats were treated with vehicle, a second group with clonidine and a third group with clonidine and an antidepressant or non-antidepressant drug. The percentage of dominant submissive pairs in these experiments was very high and many of them would not maintain this relationship through chronic experiments. When the protocol was changed to use the submissive rat as a model of depression that required chronic treatment, we had to extend the time for pair formation as is described in the next subsection.

2.4 Procedure used in early RSBM and RDBM

Similarly to previous experimental conditions animals were randomly assigned to same sex pairs. Behavioral testing was performed once a day for a 5-min period on weekdays. Testing was suspended on weekends but drug administrations were continued. The behavioral test consisted of placing each animal from a pair into opposite chambers of the apparatus after placing fresh milk in the feeder. The 5-minute period was divided into 5-second intervals. Animals observed drinking milk during each interval received one point. Drinking scores were tallied for each animal on each day of testing. Scores for the first five days (week 1) seldom showed a clear pattern and were not used. During the second week of testing, about half the pairs tested developed a pattern of behavior where one animal consistently out-scored the other. The second-week data from each member of a pair were tested for a significant difference using the two-tailed t-test. The member of a pair having a significantly lower drinking score ($p < 0.05$) was defined as submissive and his/her partner as dominant. Pairs showing this relationship were continued in the study while pairs not showing this difference were dropped from the study. Drug treatments started on Saturday after second week of testing. One member of a dominant-submissive pair was treated with the drug (dominant for testing antimanic drugs or submissive for testing antidepressant drugs). The partner of a drug-treated animal was always injected with vehicle. In a control pairs, both animals were injected with vehicle. The phases of experiment and a number of animals used in each phase (in manual scoring) are listed in the table 1.

Table 1. Timetable for basic experimental unit

procedure	time	N° of animals	N° of animals selected	N° pairs with D/S relation
habituation	5 days	16		
selection	5 days	16	10 - 14	5 - 7
drug dosing	3-6 weeks	5 - 7		5 - 7

In this experimental setting, we still used point scores instead of time scorse and the only selection criterion was significant difference between paired animals during the second week of study. This changed in the subsequent studies

2.5 Selection criteria for DSR

In further experiments, we have simplified data gathering by direct measurement of time spent on the feeder by paired animals without converting it to the point system. During the process of the experiments we, realized that control pairs did not hold stable relations as consistently as we have predicted from the beginning. Thus, we have designed experiments comparing less stringent to more stringent criteria for the selection of animal pairs with a dominant-submissive relationship. The scores of rat pairs selected with less stringent criteria are in the Table 2. Pairs of rats with scores presented in this table were borderline in statistical significance of difference in time spend on

Table 2. Pairs of rats selected according less stringent criteria

name / score, sec	Rd1	Y1	Rd2	Y2	Rd4	Y4	Rd3	Y3	Rd1	Y1
Day 1	155	240	217	192	215	282	100	239	240	286
Day 2	156	200	271	178	204	274	160	243	242	293
Day 3	210	222	260	190	257	285	249	281	264	300
Day 4										
Day 5	185	187	266	249	271	286	252	269	256	300
AVG	176.5	212.3	253.5	202.3	236.8	281.8	190.3	258.0	250.5	294.8
SUM	706	849	1014	809	947	1127	761	1032	1002	1179
P		0.09		0.05		0.07		0.2		0.001
AVG DL		35.8		51.3		45.0		67.8		44.3
SUM DL		143.0		205.0		180.0		271.0		177.0
%		16.8		20.2		16.0		26.3		15.0

Table 3. Pairs of rats selected according more stringent criteria

name / score, sec	Rd1	Y1	Rd1	Y1	Rd2	Y2	Rd3	Y3	Rd2	Y2	Rd4	Y4	Rd3	Y3
Day 1	212	151	158	203	155	238	240	157	254	189	230	49	221	192
Day 2	225	135	148	236	180	291	276	162	280	210	249	152	237	138
Day 3	209	177	177	224	126	277	262	181	298	240	250	119	250	147
Day 4	230	162	128	237	187	288	202	111	300	204	237	152	243	131
Day 5	230	169	184	202	166	286	246	170	300	231	178	132	252	147
AVG	221.2	158.8	159.0	220.4	162.8	276.0	245.2	156.2	286.4	214.8	228.8	120.8	240.6	151.0
SUM	1106	794	795	1102	814	1380	1226	781	1432	1074	1144	604	1203	755
P		0.0002		0.0015		0.0001		0.0009		0.0005		0.0022		0.0003
AVG		62.4		61.4		113.2		89.0		71.6		108.0		89.6
SUM		312.0		307.0		566.0		445.0		358.0		540.0		448.0
%		28.2		27.9		41.0		36.3		33.3		89.4		59.3

feeder by dominant and submissive rats. They were also well below the 40% value of difference between dominant and submissive rat in time spend on feeder as a percentage of dominant rat feeder time. In contrast, data presented in Table 3 are fulfilling these criteria. The results of this

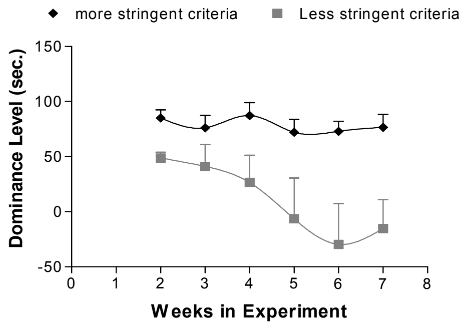


Figure 2. Shows stability of the dominance level in pairs of rats during seven weeks of experiment under application of selection criteria with different stringency (see Table 2 & 3). Pairs of rats selected using more stringent criteria (Table 3) are represented by black diamonds and pairs of rats selected using less stringent criteria (see Table 2) are depicted by gray squares.

experiment are presented in Figure 2. The dominance level in pairs of rats selected using less stringent criteria decreased while dominance level in pairs selected using more stringent criteria was stable for the experiment duration. Such condition would not obscure the effect of drug administered to test either RDBM or RSBM. The final criteria are summarized Table 4.

Table 4. Criteria used to select dominant, submissive and neutral animals

Characteristic of DL value *	Dominant – Submissive Pairs	Neutral Pairs
confidence level (two tail t-test)	P < 0.05	P > 0.6
% difference of higher scoring animal	at least 40%	up to 8%
Reversal of dominance **	no	yes

*Difference in time spend on the feeder by paired animals DL = $T_{A1} - T_{A2}$; T_{A1} time spend on feeder by animal #1 T_{A2} time spend on feeder by animal #2.

**Indicates reversals of daily success as expressed by longer and shorter time spent on the feeder by an animal from the pair (occasions when submissive animal outscores its partner) during first two weeks.

3 Endpoints and data analysis

Two endpoints were used in data analysis. First, the feeder time (FT) measured for each animal of a pair during a 5-min daily session. The units recorded were seconds. There was a maximum of 300 seconds of time on the feeder possible for one rat. Significant differences between time spent on the feeder by dominant and submissive rats were determined by ANOVA using GraphPad Prism software (GraphPad Software, Inc., San Diego, CA) followed by a two-tailed t-test ($P < 0.05$). Usually the weekly (5 days) average from daily feeding time was calculated. Second, we used dominance level values to measure the social relation between paired subjects. Dominance level (DL) = FTD – FST where FTD was the feeder time of dominant rats and FTS was the feeder time of submissive rats. To enable comparisons between treatment groups, data were normalized to the initial (second experimental) week DL value The normalization was conducted according to the formula below.

$$DL_{week\ n}\ (\%) = \frac{DL_{week\ n}}{DL_{week\ 2}}$$

The statistical significance of the difference in dominance level between the control group (pairs of rats were both dominant and submissive animals were treated with vehicle) and the treatment group (submissive rat was treated with drug and dominant rats with vehicle) was determined by ANOVA, followed by a t-test.

2.6 Automation of DSR observation and scoring and its validation in the RSBM

The basic testing apparatus did not change and was used as described in previous sections. The observation system was automated so, that during the 5-minute daily sessions, the time spent in the feeder zone by each rat was recorded by the video tracking software (PanLab, San Diego Instruments, CA) (Figure 3). The camera can distinguish rats by different colors. Thus, the rats' heads were colored for the purpose of video tracking, red in one cage and yellow in the other cage. The data were automatically

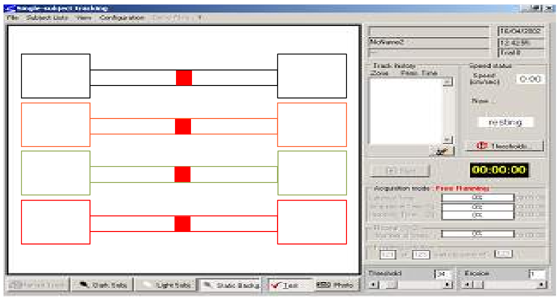


Figure 3. The computer display from the PanLab software (San Diego Instruments, CA) depicting the RSBM apparatus with defined zones. Solid squares represent the feeder areas. The time of the presence of a rat's head in this zone is recorded during each 5-minute experimental session.

saved into a text file and then pasted into an Excel extraction file through cell references. The basic apparatus was replicated for a second system. A total of four pairs of rats can be video tracked simultaneously on one table,

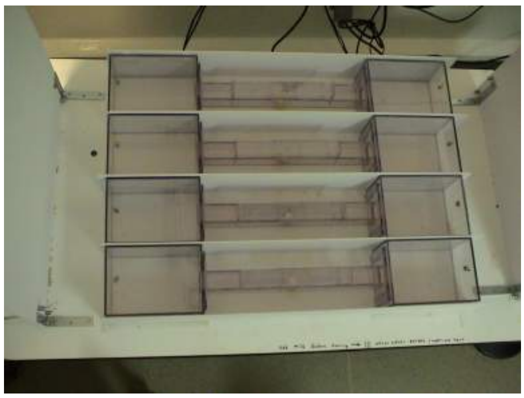


Figure 4. DSR test apparatus: the experimental set-up as viewed by the overhead video-camera. A unit of the apparatus consists of two plexiglass chambers connected by a passage having a small feeder dish with milk in the center from which only one rat can drink sweetened milk at a time. Four units are placed in parallel. The time spent at the feeder is scored for each of eight rats (four pairs) during one 5-minute experimental session.

using one camera (Figure 4). Similarly a second table with four apparati and a camera were set in the same room. This enabled faster experimental manipulation with first trial rats were put away while the second trial was running, and the third trial rats were being put in place. The experimental procedure and data analysis were conducted as described in the previous sections. We have compared stability of the controls (when both animals were injected by the vehicle for five weeks) and the effects of the SSRI inhibitor fluoxetine using the manual and automatic scoring methods. When both dominant and submissive rats were treated with sterile water for the period of five weeks, there was a statistically significant difference between FTD and FTS in these two behaviorally distinct groups of rats at each week (ANOVA followed by t-test). This was true for data generated by both manual and automatic scoring (Figure 5 A & B). The lines depicting submissive and dominant rat performance were parallel and there was no statistically significant difference in the performance of dominant or submissive rats between the selection week and the following five weeks of vehicle injections (Figure 5 A & B) in manually and automatically scored groups.

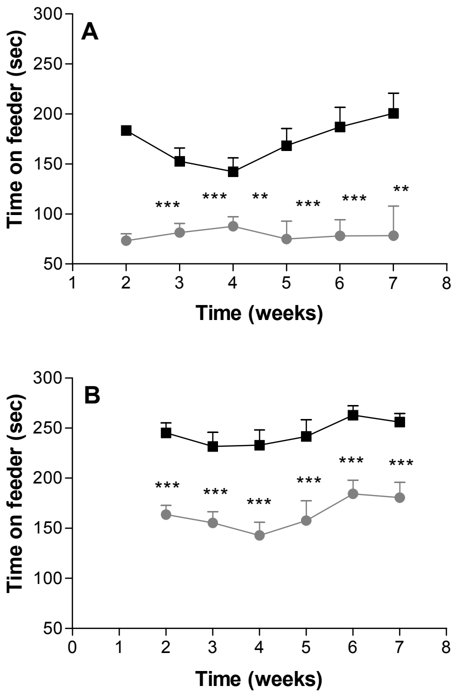


Figure5. Stability and dynamics of dominant-submissive relations in pairs of rats during seven weeks study. (A) Data scored manually by observer. (B) Data collected by video-tracking system (San Diego Instruments, Inc., San Diego, CA). Dominant and submissive animals were treated with sterile water for the five weeks. The difference between dominant and submissive rats scores marked ** at $P < 0.01$ and *** at $P < 0.001$.

The fluoxetine effect of increasing the competitiveness of submissive rats was similarly observed using manual and automatic scoring (Figure 6A&B).

4 Summary and Conclusions

One of the most significant clinical findings of treatments for mania and depression concern their time course, both in the development of the disease and response to treatment. Submissive behavior develops gradually over time as depression develops in humans. We have shown in the RSBM and RDBM that antidepressant or antimanic drugs need to be administered chronically to submissive or dominant rats to alter submissive behavior or dominant behavior. Thus, the response of submissive animals to antidepressant treatment and dominant animals to

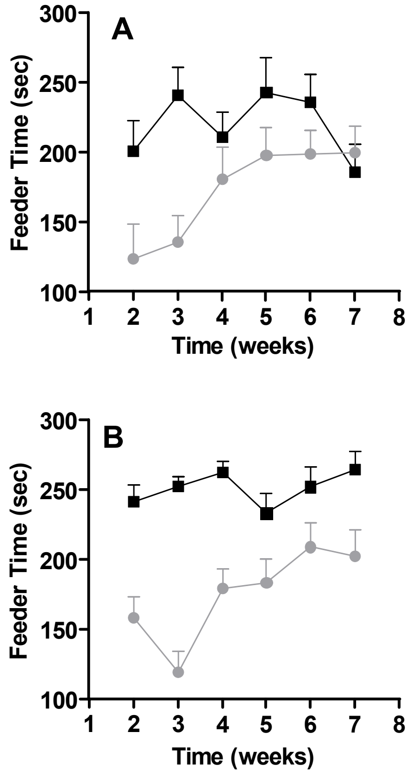


Figure 6. Effects of fluoxetine 10 mg/kg (A) Data scored manually by observer. (B) Data collected by video-tracking system (San Diego Instruments, Inc., San Diego, CA). Dominant animals (black squares) were treated with sterile water and submissive animals (grey circles) were treated with fluoxetine (10 mg/kg) for the five weeks.

antimanic drug treatment requires chronic administration as seen for human patients. With this similarity and the development of automatic version of the model, the described tests are ready for screening new medication candidates in search for one with faster onset of activity.

References:

1. Malatynska, E. (1985). Clonidine model of depression in laboratory animals, in *Departmant of Pharmacology and Physiology of CNS* Institut of Psychiatry and Neurology, College of Medicine: Warsaw. p. 137.
2. Malatynska, E. and Kostowski, W. (1988) Desipramine antagonizes clonidine-induced suppression of dominance in rats: possible involvement of amygdaloid nuclei. *Pol. J. Pharmacol. Pharm*., **40**(4), 357-364.
3. Danysz, W., et al. (1988) Comparison of desipramine, amitriptyline, zimeldine and alaproclate in six animal models used to investigate antidepressant drugs. *Pharmacol. Toxicol*., **62**(1), 42-50.
4. Kostowski, W., et al. (1986) Comparative studies on antidepressant action of alprazolam in different animal models. *Pol. J. Pharmacol. Pharm*., **38**(5-6), 471-481.
5. Malatynska, E. and Kostowski, W. (1984) The effect of antidepressant drugs on dominance behavior in rats competing for food. *Pol. J. Pharmacol. Pharm.,* **36**(5), 531-540.
6. Malatynska, E., et al. (2002) Reduction of submissive behavior in rats: a test for antidepressant drug activity. *Pharmacology,* **64** (1), 8-17.
7. Malatynska, E., Rapp, R., and Crites, G. (2002) Dominant behavior measured in a competition test as a model of mania. in *International Behavioral Neuroscience Society meeting*. Capri, Italy.

Innovation in pain research: development and validation of novel behavioral assays for the prediction of analgesic efficacy

E.J. Bilsky
University of New England, Biddeford, ME, USA

A significant unmet medical need exists for analgesics having better efficacy/tolerability than those currently available. The development of such drugs relies heavily on preclinical animal models of pain that typically involve the delivery of a presumably painful stimulus and measurement of a pain-evoked behavior (e.g. withdrawal responses). Candidate analgesics are tested for their ability to decrease pain-evoked behaviors. There are two general limitations to a reliance on these types of behavior. First, pain is associated not only with an increase in some categories of behaviors, but also with decreases in other categories of normally adaptive behaviors (e.g., feeding, locomotion). Consequently, an exclusive focus on pain-evoked behavior may neglect important consequences of pain. Indeed, patients with chronic pain often report spontaneous pain, which by definition does not involve a response to external stimulation and which cannot be assessed using standard stimulus-response approaches. A second limitation to reliance on pain-evoked behaviors to study candidate analgesics is that drugs may decrease pain-evoked behaviors by producing motor effects that impair the animal's ability to emit the pain-associated behavior. This necessitates the use of additional behavioral assays to measure drug effects on motor behavior. In view of these limitations, new behavioral procedures are being developed for the preclinical assessment of pain and analgesia. This symposium features presentations from academic and industry scientists and clinicians who are developing alternative and complementary behavioral paradigms for assessing analgesic efficacy. The speakers highlight how these models can be combined with existing procedures to provide a more comprehensive assessment of the impact of pain and analgesia on the organism. The symposium concludes with a presentation on how basic scientists and clinicians can work together to develop better assessments of analgesic efficacy and to highlight novel strategies for assessing efficacy in humans.

Pain-like behaviors in animals: How human are they?

G. Blackburn-Munro

The use of genetically-manipulated animals in conjunction with classical physiological and biochemical measurement have unravelled many of the pathological changes that occur in animal models of chronic pain, and these bear some striking similarities to those described in various human chronic pain conditions. In this presentation several possible limitations in the validation of animal models of chronic pain, with emphasis placed on neuropathic pain models, and the methods used for assessing pain-like or nociceptive behaviors in these models were discussed. The majority of preclinical pain researchers invariably measure drug effects on evoked nociceptive behaviors such as hyperalgesia and allodynia, in response to thermal or mechanical stimulation of an injured hindpaw. In contrast, drug effects on spontaneous or evoked behaviors as typified by hindpaw flinching or hindlimb weight bearing are far less frequently reported. And yet, patients suffering from chronic pain, especially that which is neuropathic in origin frequently present with a diverse constellation of symptoms inclusive of spontaneous pain and hypersensitivity to stimulation. Chronic pain patients may also suffer from comorbid psychiatric illness as typified by depression. It is now well accepted that long term stress contributes to the pathophysiology of depression, an effect which is mediated by inappropriate hypothalamo-pituitary adrenal (HPA) stress axis function. Surprisingly, very little information exists pertaining to the effects of chronic pain (a known physical stressor) on HPA function in relevant preclinical animal pain models. Based on these and other observations, alternative methods for assessing pain and stress in animals, that may better reflect the diverse symptomotology of chronic pain in humans were proposed.

Preclinical assessment of analgesic drugs in assays of acute pain

S.S. Negus

Pain is a pervasive public health problem, and analgesic drugs play a central role in its treatment. Currently, the most widely used analgesics include mu opioid agonists such as morphine and non-steroidal anti-inflammatory drugs (NSAIDS) such aspirin. Although these drugs are useful across a wide range of conditions, they are not uniformly ef.cacious, and undesirable side-effects often limit their use. Consequently, one long-standing focus of drug discovery has been the search for novel analgesics. Meaningful research on pain and analgesia depends on the development of validated procedures for identifying the presence of pain and quantifying its magnitude. Pain has been defined as 'an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.' As such, pain is essentially a subjective experience, and its existence in humans is typically assessed using some sort of verbal report. Verbal reports are obviously not suitable for measuring pain in animals, and this presents a challenge to preclinical assessment of candidate analgesics. How does one tell if an animal is in pain, and how does one tell if a putative analgesic is effective in reducing that pain? Historically, the most widely used assays for assessment of pain and analgesia in animals have involved three components:

1. Acute delivery of a thermal, mechanical, chemical or electrical noxious stimulus, which is defined as a stimulus capable of producing tissue damage.
2. Measurement of a noxious stimulus-evoked behavior, such as tail withdrawal following immersion of the tail in hot water, or abdominal constrictions/writhing after intraperitoneal administration of dilute acetic acid.

3 Assessment of the ability of drugs to decrease the rate, frequency or intensity of the noxious stimulus-evoked behavior.

In this approach the occurrence of noxious stimulus-evoked behaviors are interpreted as evidence of 'nociception', and nociception in animal is thought to be related to pain in humans. Drug-induced decreases in nociception are interpreted as evidence of 'antinociception', and antinociception in animals is thought to be related to analgesia in humans. Advantages and disadvantages of this type of assay are reviewed.

Targeting pain-suppressed behaviors in preclinical assays of pain and analgesia

E.J. Bilsky

Effective management of pain continues to be a clinical challenge. Industry and academic laboratories have invested significant resources into identifying novel targets, synthesizing small molecule drug candidates, and performing preclinical/clinical assessments of lead candidates. The introduction of novel pharmacotherapies has, however, been rather limited, with several advanced-stage candidates failing in phase III clinical trials. One potential limitation in the drug discovery process is the manner in which drug candidates are tested in preclinical models of nociception, with an emphasis being placed on pain-evoked behaviors in rodents. Our laboratories are developing assays that focus on pain-suppressed behaviors that may complement traditional methods. Here we present results that compare the effects of i.p. acetic acid on a pain-evoked behavior (writhing) and a pain-suppressed behavior (locomotor activity-LMA and feeding) in two strains of mice (ICR and C57BL/6J). Acetic acid (0.18-0.56%) produced concentration- and time-dependent increases in writhing and decreases in LMA and feeding. The potency of acetic acid was similar for producing these effects, but decreases in LMA and feeding lasted longer than increases in writhing. Interestingly, a high concentration of acetic acid (1% acid) produced less writhing than lower concentrations, whereas the highest concentrations of acetic acid tested produced maximal suppression of LMA and feeding. Morphine produced dose-related restoration of feeding at doses that did not significantly affect feeding in control animals. This effect was selectively reversed by the opioid antagonist naltrexone. For the LMA measure, the effects of morphine were confounded by the drug's propensity to stimulate LMA by itself. When comparing baseline levels of activity for the respective controls, morphine produced an upward and leftward shift in the LMA dose-response curve under pain conditions. The doses of morphine tested also reduced the number of abdominal writhes in a dose-dependent fashion. We are currently testing a number of other analgesic drugs, along with drugs that produce false positives, in this assay. These studies of pain-suppressed behaviors may help improve the predicative reliability of preclinical antinociceptive assays.

Analysis of pain related to movement in monoarthritis

K. Hygge Blakeman, S. Eriksson, K. Ängeby Möller and O.-G. Berge

Behavioral response to pain in monoarthritic rats can be measured by gait and stance analysis. We have investigated the effect of movement on pain behavior and pharmacological treatment by comparing two analysis paradigms; the box recording and the Paw Print walkway.

In the box recording, rats are placed in a 30x14 cm chamber, allowing only limited locomotion. Under the chambers glass floor a video camera is placed to record the rat's movements. The Paw Print walkway consists of a 100x10 cm path on which the rats are trained to make a swift, continuous passage. Light is projected into the long margin of the walkway's glass floor and internally reflected within the glass floor. When the rat's paw touches the floor, it lights up at the point of contact and a wide-angle camera placed under the walkway records the resulting footprint. In both analysis paradigms, the rat's behavior is manually given a pain score between one and three. In the Paw Print setup, there is also a computer-assisted analysis of the rat's gait pattern and weight-bearing.

Subjects were male Sprague-Dawley rats. Monoarthritis was induced by a tibio-tarsal injection of Freund's complete adjuvant and all animals were then sequentially scored in both setups for up to 8 days after induction. Morphine, dexamethasone and naproxen were administered at different time-points.

There was a strong positive correlation between the scores from the Paw Print walkway and box recording (tau=0.68). Linear regression showed a logarithmic relationship (R2=0.98) where the walkway tended to yield higher scores than box recording. Morphine and dexamethasone were less efficacious in reducing pain-scores in the walkway paradigm and this was also the case with naproxen.

There is a difference between the two analysis setups in both the behavioral endpoint and pharmacological sensitivity. The data suggest a higher level of pain in the Paw Print walkway, which may be a useful paradigm for studying pain related to movement.

Measuring signs and surrogate markers of spontaneous pain-like behaviours in animal models of nociception.

G.Blackburn-Munro, S.F.Bomholt, L.H.Pedersen
Department of Pharmacology, NeuroSearch A/S, Ballerup, Denmark

Abstract
The use of genetically-manipulated animals in conjunction with classical physiological and biochemical measurement have unravelled many of the pathological changes that occur in animal models of chronic pain, and these bear some striking similarities to those described in various human chronic pain conditions. In the following presentation, we will discuss several possible limitations in the validation of animal models of chronic pain, with emphasis placed on neuropathic pain models, and the methods used for assessing pain-like or nociceptive behaviours in these models. Based on these and other observations, alternative methods for assessing pain and stress in animals, that may better reflect the diverse symptomotology of chronic pain in humans shall be proposed.

Keywords
Allodynia, Hyperalgesia, Nerve injury, Spontaneous pain, Stress.

1 Introduction
Nociceptive information is relayed through multiple parallel pathways to higher brain structures to initiate the appraisal of pain. The sensory or discriminatory dimension of pain originates within the spinal dorsal horn and ascends by the spinothalamic pathway to terminate within the ventroposterior and ventrobasal thalamus. Neurons in these nuclei project to the somatosensory cortex, which rapidly discriminates the pain response with regard to its temporal encoding properties. The affective or cognitive dimension of pain originates within the superficial dorsal horn and ascends through the parabrachial area to project to structures in the brain associated with unpleasantness, including the hypothalamus, amygdala, and insular and anterior cingulate cortices. In addition to mediating the sympathoadrenal and neuroendocrine aspects of the pain response, this dimension comprises emotional feelings that pertain to the present and future consequences of coping with pain and ultimately reinforces the desire to terminate, reduce, or escape its presence [8]. Other pathways also contribute to the parallel processing of the discriminatory and affective dimensions of pain. Functional imaging studies in humans have consistently shown activation of these limbic and prefrontal cortical structures in response to various noxious stimuli [13]. Similarly, neuroimaging studies also have revealed multiple abnormalities in these same structures in patients with mood disorders such as depression [7].

Many of the pathophysiological changes that occur in animal models of chronic pain bear striking similarities to those described in various human chronic pain conditions. Patients suffering from chronic pain, especially that which is neuropathic in origin frequently present with a diverse constellation of symptoms inclusive of spontaneous pain and hypersensitivity to sensory stimulation. Recently, Backonja and Stacey (2004) have reported in a study of 309 patients with neuropathic pain of mixed etiology that almost all had signs of ongoing pain, two thirds were hypersensitive to mechanical stimulation and one third to thermal stimulation [1]. It is somewhat surprising therefore that the vast majority of animal inflammatory and neuropathic pain studies utilise sensory testing methods, which reflect hypersensitivity, rather than spontaneous pain measures [11].

The chronic, unremitting nature of pain likely contributes to altered mood status in chronic pain patients. Animal models of nerve injury that incorporate a number of the key sensory abnormalities observed in human neuropathic pain patients might also be associated with altered mood status, thereby reflecting the presence of spontaneous pain. To date, the only study to address this issue failed to show any behavioural differences between nerve-injured rats and sham operated rats in either the open field test, motility box test or the elevated plus maze [9]. Nevertheless, other strategies might be available for indirectly measuring mood status in animal models of chronic pain. The desire to avoid pain is a clear indication of the affective response to injurious or intense stimulation of body tissue. Recently, a behavioural test paradigm that measures the aversiveness of nociceptive stimuli as an attempt to model the affective/motivational aspect of clinical pain states has been described [10], although it has only been partially chracterised pharmacologically.

Long-term stress contributes to the pathophysiology of depression, an effect that is mediated by inappropriate hypothalamo-pituitary adrenal (HPA) stress axis function. A number of chronic pain conditions in humans, such as rheumatoid arthritis and fibromyalgia are also associated with profound HPA axis dysfunction [3]. Animal models of chronic inflammatory disease such as adjuvant-induced arthritis (AA) are associated with overt behavioural hyperalgesia and allodynia in response to hindpaw sensory stimulation [5]. AA rats also display profound HPA axis dysfunction as typified by increased basal plasma levels of ACTH and corticosterone [5]. Furthermore, automated microsampling of plasma, made via a chronic indwelling jugular cannula in conscious, unrestrained rats has been used to show that pulse frequency of corticosterone is increased after AA. Although invasive this method might theoretically provide an indirect measure of non-evoked pain-like behaviour of the injured animal in its home environment. Surprisingly, little information exists pertaining to the effects of chronic pain (a known physical stressor) on HPA function in relevant animal neuropathic pain models.

Here, we will describe an automated method used for assessing spontaneous nociceptive behaviours and measuring analgesic efficacy of novel drug candidates in a rat model of persistent pain. Secondly, we will describe a behavioural paradigm that may be used for assessing the aversive nature of the pain response in rat models of chronic pain, and its sensitivity to pharmacologic manipulation. Finally, we will summarise the methods used and data obtained in describing HPA axis function in a well characterised rat model of peripheral neuropathic pain.

2 Methods

2.1 Animals and nerve injury procedures

Adult male Sprague-Dawley rats (Harlan Scandinavia) housed under standard conditions were used. They were allowed to habituate to the housing facilities for at least 1 week prior to surgery or behavioural testing. All experiments were performed according to the Ethical Guidelines of the International Association for the Study of Pain [15], and all procedures were conducted in accordance with the Danish National Guide for Care and Use of Laboratory animals.

Neuropathic pain was induced in rats by making a chronic constriction injury (CCI) of the sciatic nerve according to the method of Bennett and Xie [2]. Hindpaw mechanical allodynia (MA) was measured with a set of von Frey hairs. Mechanical hyperalgesia (MH) was assessed via the use of a safety pin according to the pin prick test. Only rats with robust MA (<2 g post-injury vs. 20 g pre-injury) and MH (8-15 s post-injury vs. <0.5 s pre-injury) were included in these studies.

2.2 Automated formalin testing

Assessment of formalin-induced flinching behaviour in normal, uninjured rats (body weight 150-180 g) was made via the use of an Automated Nociception Analyser (University of California, San Diego, USA; [14]) or manually by an observer. For automated counting a small C- shaped metal band (10 mm wide x 27 mm long) was placed around the hindpaw of the rat to be tested. Each rat (four rats were included in each testing session) was administered drug or vehicle according to the experimental paradigm being followed, and then placed in a cylindrical acrylic observation chamber (diameter 30.5 cm and height 15 cm). Individual rats were then gently restrained and formalin (5% in saline, 50 µl, s.c.) was injected into the dorsal surface of the hindpaw using a 27G needle. They were then returned to their separate observation chambers, each of which were in turn situated upon an enclosed detection device consisting of two electromagnetic coils designed to produce an electromagnetic field in which movement of the metal band could be detected. The analogue signal was then digitised and a software algorithm applied to enable discrimination of flinching behaviour from other paw movements. A sampling interval of 1 min was used and on the basis of the resulting response patterns three phases of nociceptive behaviour were identified and scored; first phase (0-5 min), interphase (6-15 min) and second phase (16-30 min), [4]. For manual testing, rats were administered drug or vehicle according to the experimental paradigm being followed, injected with and formalin and placed in open mesh steel cages enabling measurement of flinching behaviour too be initiated by a skilled observer blinded to drug treatment. On the basis of the response pattern, two distinct phases of nociceptive behaviour (0-5 and 15-40 min) characterised by flinching of the affected paw were identified and scored. Each flinch was registered on-line by the observer into a DOS-based PC programme. Each rat was observed for 15 s in sequence and the 15 s bins were then collated for each rat to obtain 5 min data bins for the 60 min duration of the experiment.

2.3 Place/escape avoidance paradigm

Place/escape avoidance testing [10] was performed on CCI rats 1-3 weeks after surgery. The rats were placed within a Plexiglas chamber (60 x 30 x 30 cm; one half of which was painted white and the other half black on the external surface) placed upon an elevated metal grid. Specifically, the rats were placed over the midline of the novel chamber, and stimulation of plantar surface of the hindpaw initiated with a 69 g von Frey hair. When residing within the dark side of the chamber the injured hindpaw was stimulated. Conversely, the non-injured hindpaw was stimulated when residing within the light side. The 69 g hair was chosen based on the observation that it initiates a reflex nociceptive response of both the injured and non-injured hindpaw. However, the nature of the stimulus is likely to be more aversive when presented to the injured hindpaw since the rats 'choose' to move from the dark to the light side of the chamber. All drugs were administered i.p. and the rats returned to their homecage (30, 120 and 60 min for morphine, gabapentin and duloxetine respectively) prior to initiation of testing in the light dark chamber. Throughout the 30 min test period rats were allowed unrestricted movement throughout the chamber.

2.4 HPA axis function

Blood sampling

CCI or sham surgery was performed on rats (day 0). Two days before the stress experiment (day 12 or day 19), rats were re-anaesthetised and a chronic indwelling polythene cannula (PE50) containing 0.9% saline was inserted in the carotid artery. Rats were administered post-operative analgesia and antibiotics and allowed to recover for 48 h. On day 14 or 21 after surgery for CCI, sham and age-matched control rats a baseline blood sample (500 µl) was obtained (0800 – 1300 h). Sixty min later rats were subjected to restraint stress using an acrylic container (20x5x7 cm) for 20 min. Serial blood samples were taken at 5, 30 and 60 min after termination of the stressor, with the volume of blood removed replaced immediately with warmed (37°C) saline. In a separate experiment, CCI and sham rats were decapitated 60 min after termination of restraint stress and trunk blood collected. All blood samples were transferred to plastic tubes containing EDTA and centrifuged for 20 min at 400 rpm (4ºC). Plasma aliquots were removed and stored at –80°C until either ACTH (serial blood samples) or corticosterone (trunk blood samples) levels were determined using commercially available [^{125}I] radioimmunoassay kits. Typical inter-assay variation for ACTH and corticosterone were 7.8-6.8% and 4.8-14.9% respectively. Typical intra-assay variation for ACTH and corticosterone were 3-3.2 and 4-12.2% respectively. Assay sensitivities for ACTH and corticosterone were 1 pg/ml and 0.6 ng/ml respectively.

Immunohistochemistry

CCI rats with marked pain-like behaviours together with sham rats were exposed to either 20 min restraint stress or left undisturbed in their homecage. Sixty min after termination of restraint stress rats were deeply anaesthetised , perfused transcardially with saline followed by 4% paraformaldehyde in 0.1 M PBS and the brain removed. After a period of post-fixation serial coronal sections (40 µm) containing the PVN were cut in series of six and stored in a solution containing ethylene glycol glycerol in 30% sucrose at -20ºC until ready for immunohistochemical processing on free-floating sections for c-Fos. The number of c-Fos immunostained neurones within sections of the medial hypothalamic paraventricular nucleus (PVN) were quantified bilaterally for each rat and for all treatment groups by an observer blinded to the treatment protocol.

In situ hybridisation

Rats (21 days post-injury) with marked pain-like behaviours together with sham and control rats were exposed to either 20 min restraint stress or left undisturbed in their homecage. After 4 h rats were decapitated and the brains rapidly removed, frozen on dry ice and stored at –80ºC. Brains were cut through the PVN using a Leica cryostat CM3050, and a series of twenty 12 μM thick coronal sections were obtained and subsequently mounted on Superfrost Plus slides Slides were allowed to dry and stored at –80ºC until in situ hybridisation (ISH) experiments were performed using one series of sections for hybridisation of each transcript.

CRF mRNA was detected using a RNA probe directed against rat CRF mRNA (bp 258-1018, GenBank accession number X03036). Sense RNA probes were used as negative controls. All probes were radioactively labelled using ^{33}P labelled UTP (Amersham Biosciences, Denmark). Following hybridisation, sections were exposed on a low energy screen (Molecular Dynamics; Amersham Biosciences, Denmark) and the screens were scanned using a Personal Molecular Imager FX phoshor imaging system (Bio-Rad, California, USA). They were then analysed using a Quantity One 4.30 phoshor imager program (Bio-Rad) and the signal quantified in terms of mean density per area of expression and expressed in arbitrary units. The optical density of the phoshor imager images was obtained from two consecutive PVN-containing sections per rat and an average value calculated.

Data are presented as mean ± S.E.M. For behavioural testing of CCI rats raw data are expressed as percent maximum possible effect (%MPE) according to the equation: %MPE=([Post-treatment value]-[Pre-treatment value])x100/ceiling value of assay-[Pre-treatment value].

3 Results and Discusssion

3.1 Automated formalin testing

Formalin injection into the rat hindpaw initiates biphasic spontaneous nociceptive behaviours consisting of either flinching, and licking and/or biting of the injected paw, which can be tentatively suggested to reflect the sensory and emotional aspects of pain processing respectively. The first phase in turn can be attributed to direct chemical stimulation of nociceptors, and the second phase to peripheral inflammatory processes and subsequent sensitisation of nociceptive spinal neurones [6]. Measurement of nociceptive behaviours via manual observation is a labour-intensive and tedious procedure. More crucially perhaps significant variation in the estimation of pain behaviours can exist between different observers. In contrast, the use of the automated system described here increases the screening capacity of the test enormously. In addition to there being a close correlation between the dose of test drug required to attenuate second phase nociceptive behaviours between the two modes of testing (Fig. 1), one of the key advantages of the automated set up is the reproducibility of results, both over time and between different observers.

3.2 Place/escape avoidance paradigm

Classical pain tests performed in animals require the experimenter to evoke a reflex nociceptive response. This measure may exclusively reflect sensory processing of nociceptive transmission. The place/escape avoidance paradigm may be used to assess drug effects on affective pain processing in rats with neuropathic pain [10]. The μ-opioid receptor agonist morphine, the antiepileptic

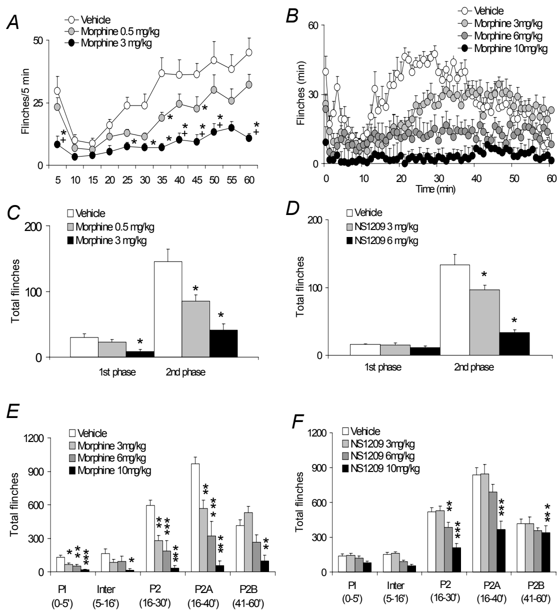


Figure 1. *Effects of the opioid agonist morphine and the AMPA/GluR5 receptor antagonist NS1209 in the rat formalin test. Both morphine (s.c.) and NS1209 (i.p.) were administered 30 min prior to formalin injection. Time course of antinociceptive effects of drugs measured (A) manually and (B) automatically. (C-F) total number of flinches recorded during each phase of the formalin test. (C and D) antinociceptive effects of morphine and NS1209 measured manually (E and F) antinociceptive effects of morphine and NS1209 measured automatically. *P<0.05, **P<0.01, ***P<0.001 (One Way ANOVA).*

gabapentin and the 5HT/NA reuptake inhibitor duloxetine all significantly reduced the %time spent in the light area of the chamber compared to control rats (Table 1). Whereas, morphine and gabapentin reduced mechanical hypersensitivity in response to hindpaw von Frey and pin prick stimulation (morphine only), duloxetine was ineffective against either measure under the conditions tested. Other drugs, such as the selective $GABA_A$ receptor agonist gaboxadol and the mixed CB1/CB2 receptor agonist WIN55212 were ineffective against all measures tested (data not shown). Thus, the place/escape avoidance paradigm may enable discrimination between selected drug classes on distinct components of sensory and affective nociceptive processing in rats with neuropathic pain.

3.3 HPA axis function

Disruption of normal HPA axis function has previously been demonstrated in rodent models of chronic inflammatory pain [5]. Whether the unremitting nature of pain or the action of pro-inflammatory mediators act as the primary stressor in chronic inflammatory disease has remained speculative. In the CCI rat model of neuropathic pain, both basal and restraint stress-induced plasma hormone responses, expression of the immediate early gene c-Fos, and PVN CRF mRNA levels were similar between nerve-injured and sham rats (Fig. 2). Thus, the current series of experiments do not readily support the hypothesis that HPA axis dysfunction arises as a consequence of chronic pain *per se*, and that HPA axis profiling might not easily lend itself to estimating the degree of distress and ongoing pain in animals after nerve injury. Finally, this suggests that in contrast to a number of chronic pain conditions in humans long-term stress might not necessarily impact negatively on mood status in animals.

Table 1. *Comparison of morphine, gabapentin and duloxetine actions on sensory (allodynia and hyperalgesia) and affective (aversion) components of nociceptive behaviour in rats with neuropathic pain. *P<0.05, **P<0.01, ***P<0.001 (One Way ANOVA).*

Drug/dose (mg/kg)	%MPE in light	%MPE MA	%MPE MH
Morphine			
0 (veh)	100±6.6	-1.7±1.4	1.6±6.7
3	**38.3±10*****	10.6±5	27.7±13.8
6	**35.2±10.7*****	**55.5±14*****	**62.8±17.9****
Gabapentin			
0 (veh)	100±3.1	-1.4±1.4	-1.2±2.1
50	**7.9±3*****	**25.0±7.8****	24.1±14.2
100	**22.6±14.7*****	**24.7±5.9***	2.5±17
Duloxetine			
0 (veh)	100±2.1	0.1 ± 0.6	4.1±6
10	**22.8±5.7*****	1.1 ± 1.4	16.1±12.6
30	**16.6±6.4*****	3.0 ± 1.8	25.5±14

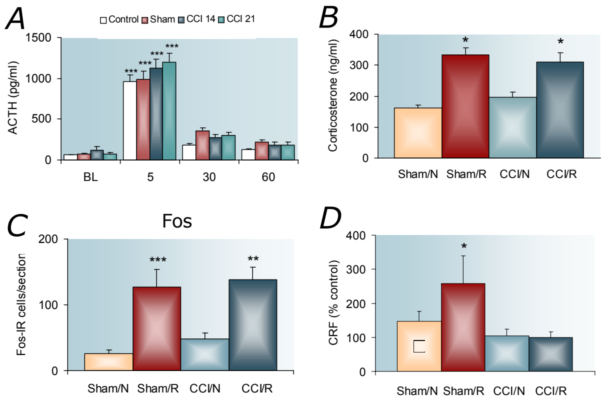


Figure 2. *HPA axis responsiveness in CCI rats. (A) and (B) restraint stress increased plasma ACTH and corticosterone similarly in 14 day and 21 day CCI, sham and control rats. (C) Restraint also increased expression of c-Fos within the PVN by a similar magnitude in CCI rats and sham controls (D) Restraint increased PVN CRF mRNA in sham rats only. Thus, the response was blunted in CCI rats which might indicate a resetting of the homeostatic set point for control of CRF expression. (A) ***P<0.001 vs baseline (Two way RM ANOVA), (B-D) *P<0.05; **P<0.01; ***P<0.001 vs control (One way ANOVA).*

4 Conclusions and future directions

Although non-evoked nociceptive behaviours are readily observed in animal models of persistent and chronic pain, they are all too infrequently reported. Various methods designed to aid quantification of spontaneous pain-like behaviours in animals are now available. Ideally, when developing novel animal models with improved face and construct validity to clinical neuropathic pain conditions, and when reporting the effects of novel drug candidates in animal pain models, spontaneous pain-like behaviours should be reported together with evoked hypersensitivity nociceptive behaviours. Alternatively, other measures of nociceptive processing which reflect the affective or emotional aspect of the pain response may represent valid parameters for describing the long-term consequences of ongoing spontaneous pain in animal models.

SFB and LHP were supported by the Danish Academy of Technical Sciences.

References:

1. Backonja, M.M.; Stacey, B. (2004). Neuropathic pain symptoms relative to overall pain rating. *Pain* **5,** 491-497.
2. Bennett, G.J.; Xie, Y.K. (1988). A peripheral mononeuropathy in rat that produces disorders of pain sensation like those seen in man. *Pain* **33,** 87-107.
3. Blackburn-Munro, G. (2004). Hypothalamo-pituitary-adrenal axis dysfunction as a contributory factor to comorbid pain and depression. *Curr. Pain Headache Rep.* **8,** 116-224.
4. Blackburn-Munro, G.; Bomholt, S.F.; Erichsen, H.K. (2004) Behavioural effects of the novel AMPA/GluR5 selective receptor antagonist NS1209 after systemic administration in animal models of experimental pain. *Neuropharmacology* **47,** 351-362.
5. Bomholt S.F.; Harbuz, M.S.; Blackburn-Munro, G. Blackburn-Munro, R.E. (2004). Involvement and role of the hypothalamo-pituitary-adrenal (HPA) stress axis in animal models of chronic pain and inflammation. *Stress* **7,** 1-14.
6. Coderre, T.J.; Fundytus, M.E.; McKenna, J.E.; Dalal, S.; Melzack, R. (1993) The formalin test: a validation of the weighted-scores method of behavioural pain rating. *Pain* **54,** 43-50.
7. Drevets, W.C. (2001). Neuroimaging and neuropathological studies of depression: implications for the cognitive-emotional features of mood disorders. *Curr. Opin. Neurobiol.* **11,** 240–249.
8. Hunt, S.P.; Mantyh, P.W (2001). The molecular dynamics of pain control. *Nat. Rev. Neurosci.* 2001, **2,** 83–91.
9. Kontinen, V.K.; Kauppila, T.; Paananen, S.; Pertovaara, A.; Kalso, E. (1999). Behavioural measures of depression and anxiety in rats with spinal nerve ligation-induced neuropathy. *Pain* **80,** 341-346.
10. LaBuda, C.J.; Fuchs, P.N. (2000). A behavioral test paradigm to measure the aversive quality of inflammatory and neuropathic pain in rats. *Exp. Neurol.* **163,** 490-494.
11. Mogil, J.S.; Crager, S.E. (2004). What should we be measuring in behavioral studies of chronic pain in animals? *Pain* **112,** 12-15.
12. Monassi, C.R.; Bandler, R.; Keay, K.A. (2003). A subpopulation of rats show social and sleep-waking changes typical of chronic neuropathic pain following peripheral nerve injury. *Eur. J. Neurosci.* **17,** 1907-1920.
13. Rainville, P. (2002). Brain mechanisms of pain affect and pain modulation. *Curr. Opin. Neurobiol.* **12,** 195–204.
14. Yaksh, T.L.; Ozaki, G.; McCumber, D.; Rathbun, M.; Svensson, C.; Malkmus, S.; Yaksh, M.C. (2001). An automated flinch detecting system for use in the formalin nociceptive bioassay. *J. Appl. Physiol.* **90,** 2386-2402.
15. Zimmermann, M. (1983). Ethical guidelines for investigations of experimental pain in conscious animals. *Pain* **16,** 109-110.

Preclinical assessment of analgesic drugs in assays of acute pain

S. Stevens Negus
Alcohol and Drug Abuse Research Center, McLean Hospital-Harvard Medical School, Belmont, MA, USA

Abstract
Meaningful research on pain and analgesia depends on the development of validated procedures for identifying the presence of pain and quantifying its magnitude. The most widely used procedures focus almost exclusively on measurement of withdrawal responses or other behaviors that increase in rate, frequency or intensity following the acute presentation of a noxious stimulus. In the radiant-heat tailflick procedure in rats, for example, a rat's tail is placed beneath a radiant heat source, the heat source is activated, and the latency to tail withdrawal is measured. Drugs can then be evaluated for their ability to attenuate the noxious stimulus-enhanced behaviors. Advantages of this type of procedure include technical simplicity, sensitivity to morphine-like analgesics, and some face validity. Disadvantages include subjectivity in measuring behavioral endpoints, unreliable relationships between stimulusenhanced behaviors and pain, difficulties in dissociating sensory and motor effects of drug treatments, and inconsistent predictive validity with non-opioid drugs.

Keywords
Pain, analgesia, nociception, antinociception

1 Introduction
Pain is a pervasive public health problem, and analgesic drugs play a central role in its treatment. Currently, the most widely used analgesics include mu opioid agonists such as morphine and non-steroidal anti-inflammatory drugs (NSAIDS) such as aspirin. Although these drugs are useful across a wide range of conditions, they are not uniformly effective, and undesirable side-effects often limit their use. Consequently, one long-standing focus of drug discovery has been the search for novel analgesics

Meaningful research on pain and analgesia depends on the development of validated procedures for identifying the presence of pain and quantifying its magnitude. Pain has been defined by the International Association for the Study of Pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" [18]. Analgesia is a selective reduction in pain without altering other sensitivity in other sensory modalities. As such, pain and analgesia are essentially subjective experiences, and their existence in humans are typically assessed using verbal reports. However, verbal reports are obviously not suitable for measuring pain and analgesia in animals, and this presents a challenge to preclinical research. How does one tell if an animal is in pain, and how does one tell if a putative analgesic is effective in reducing that pain?

In the absence of verbal reports, the existence of pain in animals has typically been inferred by measuring changes in behavior induced by acute exposure to noxious stimuli, which are defined as stimuli that produce or are capable of producing tissue damage. For nearly a century, preclinical researchers interested in such issues as the genetics, neurobiology and pharmacology of pain and analgesia have focused almost exclusively on behaviors that increase in rate, frequency or intensity following the presentation of a noxious stimulus [2, 3, 8, 12-14]. The radiant-heat tail-flick test as developed by D'Amour and Smith [7] exemplifies this type of procedure. In this procedure, a rat's tail is placed beneath a radiant heat source (i.e. a light bulb), the heat source is activated, and the primary dependent variable is the latency to tail withdrawal. Rapid tail withdrawal is considered to be evidence for "nociception" (i.e. the ability to detect a noxious stimulus), and nociception is thought to be functionally related to pain. Slowing of tail withdrawal latencies by drugs is considered as preliminary evidence of "antinociception," which is thought to be related to analgesia. D'Amour and Smith examined the antinociceptive effects of morphine and four other opioids in their original paper, and since then, versions of the tail flick test (including tests that deliver noxious thermal stimuli via other means, such as hot plates or exposure to warm water) have been used to study thousands of candidate analgesic drugs as well as the underlying mechanisms of nociception and antinociception.

The early evolution of this type of procedure was quickly influenced by the importance of two key independent variables: stimulus intensity and stimulus modality. As stimulus intensity increases (i.e. as temperature increases in assays of thermal nociception), the rate, frequency or intensity of the evoked response also typically increases, and the evoked response also becomes more resistant to attenuation by drugs. The value of this ability to modulate both evoked behaviors and drug effects my manipulating stimulus intensity rapidly became apparent, and preclinical studies of pain and analgesia often incorporate measures of behaviors evoked by a range of noxious stimulus intensities [4, 15, 19, 26, 28]. Moreover, by testing a range of stimulus intensities, the threshold intensity for evoking the target behavior can be determined, and this threshold can then serve as a dependent measure for evaluating drug effects [16].

Stimulus modality has also been extensively manipulated in assays of nociception. Thermal stimuli are still widely used, and as noted above, various methods have been developed for delivering both noxious heat and noxious cold stimuli. Other commonly used modalities include mechanical and chemical noxious stimuli [2, 3, 8, 12-14]. Mechanical noxious stimuli physically deform tissue (e.g. by pinching or applying pressure via probes), and as with thermal tests, the dependent measure is usually the latency to a withdrawal response or the threshold stimulus intensity for evoking a withdrawal response. Commonly used chemical noxious stimuli include intraperitoneal injections of dilute acetic acid or intraplantar injections of formalin. When chemical noxious stimuli are used, withdrawal from the stimulus is rarely possible, and the primary dependent variables are typically behaviors such as writhing or flinching. The number of these behaviors is then counted during an observation period of some set duration

In summary, preclinical studies of pain and analgesia have been conducted largely using approaches that involve the delivery of an acute noxious stimulus and the measurement of a resultant change in behavior. This brief overview should make it clear that, despite extensive and productive manipulation of variables related to the stimulus intensity and modality in these procedures, there has been little evolution in the types of behaviors that are measured. In virtually all cases, the target behavior has been either a withdrawal response (if the stimulus can be escaped), or some other behavior that increases in rate, frequency or intensity following presentation of the noxious stimulus (e.g. writhing or flinching when the noxious stimulus cannot be escaped). For the purposes of discussion in the remainder of this manuscript, behaviors that increase in rate, frequency or intensity after delivery of a noxious stimulus will be referred to as "noxious stimulus-enhanced behaviors."

2 Introduction

The popularity of procedures that measure noxious stimulus-enhanced behaviors testifies to their utility in preclinical pharmacology, and there are both methodological and heuristic advantages to these procedures as experimental tools. At least four methodological advantages can be enumerated. First, although expensive and sophisticated devices have been developed for conducting studies of nociception, such devices are certainly not necessary. A warm-water tail-withdrawal study in rats requires only a source of hot water (as the stimulus) and a stop watch (to measure the response latency). Studies of acetic acid-induced writhing require little more than a hypodermic syringe to deliver the acid and the patience to observe the resulting number of writhes. A second methodological advantage is that experimental subjects require little or no training in the procedure. A third advantage is that baseline expression of the target behaviors is typically very low and very stable, and as a result, it is relatively easy to detect changes in these behaviors induced by the noxious stimulus. A final advantage is that noxious stimulus-enhanced behaviors are reliably attenuated by some classes of drugs, such as morphine-like opioid agonists. Consequently, these procedures have been invaluable for in vivo evaluation of the potency, time course, efficacy and receptor mechanisms of actions of some classes of drugs such as opioids (e.g. [6]. Overall, studies of noxious stimulus-enhanced behaviors are relatively easy to perform, and they can be used to rapidly and efficiently collect data on some clinically relevant classes of drugs.

Studies of noxious stimulus-enhanced behaviors in animals also benefit from a measure of face validity with regard to acute pain in humans. For example, the heat threshold to elicit withdrawal responses in animals is very similar to the threshold to produce verbal reports of pain in humans. Indeed the tail flick test of D'Amour and Smith was based on a procedure that had been developed to measure thermal pain thresholds in humans [21]. Consequently, it is easy to imagine that delivery of noxious stimuli to animals might produce a subjective experience similar to what humans mean by pain and that behaviors induced by noxious stimuli could serve as measures of that pain state. The impression of face validity is further nourished by the sensitivity of noxious stimulus-enhanced behaviors to suppression by morphine-like analgesics. As a result, it is perhaps an easy step to derive the more general notion that drug-induced reductions in noxious stimulus-enhanced behaviors are suggestive of reductions in pain—i.e. analgesia. The pitfalls of this line of reasoning will be addressed below, but it must be acknowledged that compelling parallels exist between acute pain/analgesia in humans and acute nociception/antinociception in animals, and these parallels have made the behavioral science of nociception nociception appear intellectually accessible and clinically relevant.

3 Disadvantages of procedures that focus on noxious stimulus-enhanced behaviors

Perhaps the technical simplicity, pharmacological utility and face validity of procedures that assess noxious stimulusenhanced behaviors contribute to the apparent high reinforcing efficacy of these procedures among scientists (including the author of this manuscript). However, it is important to recognize the limitations as well as the strengths of these procedures, and at least four such limitations can be described. First, the procedures often require relatively intrusive involvement of the experimenter both in the handling of the subject and in the measurement of behavioral endpoints. For example, one common type of procedure measures tail-withdrawal latencies by mice from water heated to noxious temperatures (usually >50°C). In a typical experimental set-up, the water will be heated to a constant temperature in a water bath, and the mouse will be picked up by the experimenter and held over the water bath so that the tip of the mouse's tail is immersed in the water. Clearly, from the mouse's perspective, the warm-water stimulus is just one of several major stimuli associated with the total experience, and uncontrolled stimuli associated with the experimenter (smell, tightness of grip etc) can profoundly influence tail withdrawal responses. In addition, the dependent measure (tail withdrawal latency) implicitly assumes that tail withdrawal is a quantal response that can be scored as present or absent, but in practice, the topology of the withdrawal responses can vary considerably, and one challenge to the experimenter is to decide when tail withdrawal has occurred and to apply those criteria consistently across time. The profound impact of uncontrolled stimuli associated with the experimenter on noxious stimulus-enhanced behaviors was recently highlighted by a review of baseline tail-withdrawal data in mice [5]. This study found that experimenter identity had the greatest impact on tail-withdrawal latencies of all variables measured, exceeding even the impact of genotype.

A second disadvantage concerns the reliability of noxious stimulus-enhanced behaviors as surrogate measures for the subjective experience of pain. Although there is generally good concordance between acute stimuli that produce noxious stimulus-enhanced behaviors in animals and pain in humans, it appears that this relationship does not always hold. For example, intraperitoneal injection of dilute acetic acid elicits a writhing response in mice, and this response is routinely used as a measure of nociception. However, the curve that relates concentration of acid to number of writhes has an inverted-U shape, such that maximal numbers of writhes are produced by a concentration of approximately 0.5-

0.6% acetic acid (E. Bilskey, personal communication). At higher acid concentrations, the number of writhes declines again, but it is unlikely that these higher acid concentrations would be associated with decreased "pain." Thus, the number of writhes should probably not be interpreted as a straightforward measure of pain. One troubling implication of this inverted-U shaped concentration effect curve for acetic-acid induced writhing is that drug treatments could reduce acid-induced writhing either by reducing sensitivity to the stimulus or by *increasing sensitivity to the stimulus—i.e. by making the acid seem more painful.*

The relationship between noxious stimulus-enhanced behaviors and pain is even more problematic in cases of chronic pain associated with inflammation or neuropathy. For example, patients with chronic pain often report spontaneous pain, which by definition does not involve a response to external stimulation. Thus, spontaneous pain cannot be assessed using standard stimulus-response approaches to pain measurement. Moreover, even when pain-evoked withdrawal responses are studied in chronic pain patients, it is not clear that these responses correlate reliably with subjective pain states. For example, the nociceptive flexion reflex (NFR), which assesses withdrawal responses following delivery of a painful electrical stimulus, has been used to study pain and analgesia in humans. The NFR response appears to correlate reliably with subjective reports of ACUTE pain; however, there is often a dissociation between NFR responses and subjective pain experience in patients with various CHRONIC pain conditions [23]. Similarly, a recent functional magnetic resonance imaging (fMRI) study in fibromyalgia patients and matched controls found that delivery of a painful pressure stimulus activated comparable regions involved in stimulus detection (e.g. somatosensory cortex), but fibromyalgia patients displayed significantly lower levels of activation in motor areas involved in the coordination of withdrawal responses, and patients also displayed fewer overt motor responses to painful stimuli [9]. The authors concluded that the results were consistent with "…a state of reduced affective appraisal and responsiveness in fibromyalgia," and they go on to suggest the results may "…reflect an adaptive mechanism in which patients with chronic pain have become so accustomed to persistent pain that the brief, moderate-to-strong pain evoked in the experimental paradigm does not produce the emotional responses observed in those unaccustomed to pain." [9]. This description is reminiscent of "learned helplessness" theories of depression, and evidence for depressed motor activity in chronic pain may also be related to the high degree of comorbidity between chronic pain and clinical depression [1]. For example, one recent review of methods for pain assessment in animals argues that "In humans, chronic pain often causes depression, and behaviors seen in animals suffering from chronic pain are sometimes taken as indicators of 'learned helplessness,' where animals become unresponsive and apathetic following unsuccessful responses to pain" [20]. Overall, these results suggest that assessment of noxious stimulus-enhanced behaviors is not sufficient to fully characterize subjective experiences of pain in chronic pain patient, and by extension, measurement of these behaviors is unlikely to be sufficient to model chronic pain in animals.

A third disadvantage associated with the use of procedures that measure noxious stimulus-enhanced behaviors is that it is difficult to dissociate the sensory from the motor effects of drugs. Ideally, a drug might decrease expression of a noxious stimulus-enhanced behavior by reducing sensory detection of the stimulus (true analgesia). However, drugs may also reduce expression of these behaviors by reducing the ability of the subject to respond to the stimulus. This difficulty in dissociating sensory motor effects of drugs is generally well-appreciated, and to address this issue, many investigators compare drug potency to attenuate noxious stimulus-enhanced behaviors with drug potency to alter motor behavior in separate procedures, such as assays of locomotor activity or rotorod performance (e.g. [22]. Of course, the reliance on this comparison introduces a whole new set of problems, such as what tests of motor behavior are most appropriate, and what potency ratios are deemed sufficient to justify further testing.

The final disadvantage of procedures that measure noxious stimulus-enhanced behaviors is inconsistent predictive validity for non-opioids, and it has been common to observe both false negatives (i.e. drugs ineffective in preclinical procedures but effective clinically as analgesics) and false positives (i.e. drugs deemed effective preclinically but that are not effective clinically as analgesics). As one prominent example of a false negative, non-steroidal antiinflammatory drugs (NSAIDS) are widely and effectively used as analgesics, but these drugs typically do not produce antinociceptive effects in preclinical procedures of acute thermal or mechanical nociception (although they may be effective in assays of acute chemical nociception). In part due to the insensitivity of traditional acute nociception procedures to NSAIDS, one step in the evolution of preclinical assays of pain and analgesia has been the development of procedures in which basal nociceptive sensitivity to thermal or mechanical stimuli is increased. This is usually accomplished in one of two ways intended to model mechanistically distinct pain states in humans [3, 13, 14, 27]. First, nociceptive sensitivity may be enhanced by experimental induction of inflammation or an inflammatorylike state. For example, in one approach developed in the monkey, subcutaneous injection of the inflammatory mediators bradykinin or prostaglandin E2 in the tail of the monkey reduces thermal thresholds for warm water-induced tail withdrawal [16, 17]. Depending on the means used to produce the thermal hypersensitivity, NSAIDS may be effective (i.e. the NSAID ketorolac was effective against thermal hypersensitivity induced by bradykinin, but not by prostaglandin). The second approach seeks to model neuropathic pain by imposing some type of nerve injury. Hypersensitivity associated with nerve damage is typically not sensitive to NSAIDS.

Drugs that disrupt motor behavior often emerge as false positives in assays of noxious stimulus-enhanced behaviors, and as noted above, comparison of drug effects on nociception and on tests of motor function is often used to screen out this type of false positive. However, the vulnerability of these procedures to false positives goes well beyond drugs that produce obvious motor impairment. The ongoing experience with neurokinin receptor antagonists as candidate analgesics illustrates the general problem [10, 11, 24, 25]. The tachykinin substance P has long been implicated as a key neurotransmitter in the mediation of nociceptive signals, and at least some neurokinin receptor antagonists produced promising effects in preclinical assays of noxious stimulus-enhanced behaviors (e.g. [22]. To date, however, these drugs have not shown reliable analgesic efficacy in humans.

4 Summary and future directions

Assays of noxious stimulus-enhanced behaviors emerged early in the 20th century as a first step in the development of procedures designed to assess pain and analgesia in animals. These procedures involve the delivery of a noxious stimulus to an organism and the measurement of a response, and procedures such as the tail-flick test in rats are still widely used today. Advantages of these procedures include their technical simplicity, utility for pharmacological characterization of some classes of drugs, and face validity in relation to acute pain in humans. Disadvantages include subjectivity in the conduct of experiments and measurement of dependent variables, unreliable relationship between noxious stimulus-enhanced behaviors and pain, difficulty in dissociating sensory from motor effects of drugs, and inconsistent predictive validity for analgesic drug development.

Early refinements of these procedures involved the manipulation of independent variables related to stimulus intensity and stimulus modality. More recently, the development of models of inflammatory and neuropathic pain have manipulated independent variables related to the basal pain state of the organism, and introduction of these models marked a significant advance in preclinical procedures designed to explore mechanisms and treatment of pain. However, it is critical to recognize that even with these newer procedures, the primary dependent variable usually remains a stimulus-enhanced behavior, such as a withdrawal response from a mechanical or thermal stimulus. As such, these procedures remain vulnerable to all the limitations described above. The next stage in the evolution of preclinical procedures related to pain and analgesia will likely involve development of novel behavior endpoints that will address the limitations of existing procedures. Some of those novel endpoints will be the topics of other talks in this symposium.

Supported by RO1-DA11460 from the National Institutes of Health, National Institute on Drug Abuse

References:

1. Bair, M.J.; Robinson, R.L.; Katon, W.; Kroenke, K. (2003). Depression and pain comorbidity: a literature review. *Arch. Intern. Med.*, **163**, 2433-2445.
2. Barbour, H.; Maurer, L. (1920). Tyramine as a morphine antagonist. *J. Pharamcol. Exp. Ther.*, **15**, 305-330.
3. Bennett, G.J. (2001). Animal Models of Pain. In: Kruger, L. (ed): *Methods in Pain Research*. Boca Raton, FL, CRC Press, 67-91.
4. Carstens, E.; Wilson, C. (1993). Rat tail flick reflex: magnitude measurement of stimulus-response function, suppression by morphine and habituation. *J. Neurophysiol*., **70**, 630-639.
5. Chesler, E.J.; Wilson, S.G.; Lariviere, W.R.; Rodriguez-Zas, S.L.; Mogil, J.S. (2002). Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neurosci. Biobehav. Rev*., **26**, 907-923.
6. Coop, A. (2005). Biologcal evaluation of opioids, stimulants, and depressants. I. An overview of the studies performed by the Drug Evaluation Committee of the College on Problems of Drug Dependence. *NIDA Res. Monogr.*, **185**, 105-130.
7. D'Amour, F.; Smith, D. (1940). A method for determining loss of pain sensation. *J. Pharmacol. Exp. Ther*., **72**, 74-79.
8. Dykstra, L.A. (1985). Behavioral and pharmacological factors in opioid analgesia. In: Seiden, L.S.,Balster, R.L. (ed): *Behavioral Pharmacology*: The Current Status. New York, Alan r. Liss, Inc., 111-129.
9. Gracely, R.H.; Petzke, F.; Wolf, J.M.; Clauw, D.J. (2002). Functional magnetic resonance imaging evidence of augmented pain processing in fibromyalgia. *Arthritis Rheum*., **46**, 1333-1343.
10. Hill, R. (2000). NK1 (substance P) receptor antagonists -why are they not analgesic in humans? *Trends Pharmacol. Sci.*, **21**, 244-246.
11. Hill, R. (2000). Reply: will changing the testing paradigms show that NK1 receptor antagonists are analgesic in humans? *Trends Pharmacol. Sci*., **21**, 465.
12. Krueger, H.; Eddy, N.; Sumwalt, M. (1942). The Pharmacology of the Opium Alkaloids.
13. Le Bars, D.; Gozariu, M.; Cadden, S.W. (2001). Animal models of nociception. *Pharmacol. Rev*., **53**, 597-652.
14. Mogil, J.S.; Wilson, S.G.; Wan, Y. (2001). Assessing nociception in murine subjects. In: Kruger, L. (ed): *Methods in Pain Research*. Boca Raton, FL, CRC Press, 11-39.
15. Morgan, D.; Cook, C.D.; Smith, M.A.; Picker, M.J. (1999). An examination of the interactions between the antinociceptive effects of morphine and various muopioids: the role of intrinsic efficacy and stimulus intensity. *Anesth Analg*, **88**, 407-413.
16. Negus, S.S.; Butelman, E.R.; Al, Y.; Woods, J.H. (1993). Prostaglandin E2-induced thermal hyperalgesia and its reversal by morphine in the warm-water tail-withdrawal procedure in rhesus monkeys. *J. Pharmacol Exp. Ther*., **266**, 1355-1363.
17. Negus, S.S.; Butelman, E.R.; Gatch, M.B.; Woods, J.H. (1995). Effects of morphine and ketorolac on thermal allodynia induced by prostaglandin E2 and bradykinin in rhesus monkeys. *J Pharmacol Exp Ther.*, **274**, 805-814.
18. Pain, I.A.f.t.S.o. (1979). Pain terms" a list with definitions and notes on usage. *Pain*, **6**, 249-252.
19. Qi, J.N.; Mosberg, H.I.; Porreca, F. (1990). Modulation of the potency and efficacy of mu-mediated antinociception by delta agonists in the mouse. *J. Pharmacol. Exp. Ther*., **254**, 683-689.
20. Rutherford, K.M.D. (2002). Assessing pain in animals. *Animal Welfare*, **11**, 31-53.
21. Schumacher, G.; Goodell, H.; Hardy, J.; Woolff, H. (1940). Uniformity of pain threshold in man. *Science*, **92**, 110-112.
22. Seguin, L.; Le Marouille-Girardon, S.; Millan, M.J. (1995). Antinociceptive profiles of non-peptidergic neurokinin1 and neurokinin2 receptor antagonists: a comparison to other classes of antinociceptive agent. *Pain*, **61**, 325-343.
23. Skljarevski, V.; Ramadan, N.M. (2002). The nociceptive flexion reflex in humans -- review article. *Pain*, **96**, 3-8.
24. Urban, L.A.; Fox, A.J. (2000). NK1 receptor antagonists- are they really without effect in the pain clinic? *Trends Pharmacol. Sci*., **21**, 462-464.
25. Villanueva, L. (2000). Is there a gap between preclinical and clinical studies of analgesia? *Trends Pharmacol. Sci*., **21**, 461-462; author reply 465.
26. Walker, E.A.; Butelman, E.R.; Decosta, B.R.; Woods, J.H. (1993). Opioid thermal antinociception in rhesus monkeys: Receptor mechanisms and temperature dependency. *J. Pharmacol. Exp. Ther*., **267**, 280-286.
27. Wang, L.X.; Wang, Z.J. (2003). Animal and cellular models of chronic pain. *Adv. Drug Deliv. Rev*., **55**, 949-965.
28. Woolfe, G.; MacDonald, A. (1944). The evaluation of the analgesic action of pethidine hydrochloride (Demerol). *J. Pharmacol. Exp. Ther*., **80**, 300-307.

Analysis of Pain Related to Movement in Monoarthritis

K. Hygge Blakeman[1], S. Eriksson[2], K. Ängeby Möller[1], O.G. Berge[1]
[1] *Bioscience, AstraZeneca, Södertälje, Sweden*
[2] *Trial Form Support AB, Stockholm, Sweden*

Abstract

Behavioral response to pain in monoarthritic rats can be measured by stance and gait analysis. We have investigated the effect of movement on pain behavior and pharmacological treatment by comparing two analysis paradigms; the box recording and the Paw Print walkway. We have found that there is a difference between the two analysis setups in both the behavioral endpoint and pharmacological sensitivity.

Keywords

Gait analysis, Monoarthritis, Walking pain, Rat, Standing pain.

1 Introduction

There are several ways to study spontaneous pain behaviors in rats with monoarthritis. In our lab we have used the box recording (adapted from [1]). However, the manual evaluation of pain behavior is labor intensive with this setup and this has cultivated a desire to increase throughput. The Paw Print walkway is a modified version of the CatWalk introduced by Hamers et al. in 2001 [2]. It has been developed with hardware that allows for faster manual scoring and has a computerized gait detection algorithm extracting several parameters pertaining to the rat's gait pattern and weight-bearing. The Paw Print walkway consists of a glass-floored corridor where rats are trained to make a swift continuous passage. Internally reflected light in the glass floor is refracted at contact points, illuminating the rat paws and thereby enhancing the visibility of the prints. In the box recording, rats are placed on a glass floor and recorded from below in a chamber allowing for minimal movement. Our aim was to investigate the potential behavioral and pharmacological differences between the Paw Print setup and the box recording.

To begin with, we wanted to see if the same animals run on both setups would generate the same pain scores (Study I). Also, we investigated whether known analgesics would have comparable efficiency in the two setups. In Study II the analgesic efficacy of morphine and dexamethasone were investigated for both Paw Print and box scoring. Finally, we wanted to see how changing the concentration of the monoarthritis induction agent would influence the outcome (Study III).

2 Materials and Methods

2.1 Animals

Male Sprague-Dawley rats (M&B, Denmark or Scanbur BK, Sweden) were used in all studies. Housing environment was maintained at 12:12 hour light/dark cycle, 21±2°C, and 55-60% humidity. Food and water were available *ad libitum*. After arrival, the animals were acclimatized to their new surroundings for at least a week before they were included in a study. At the end of a study, all animals were euthanized by an overdose of pentobarbital. The procedures followed the ethical guidelines of the IASP [3] and were approved by the appropriate ethics committee (Stockholms Södra försöksdjursetiska nämnd).

2.2 Monoarthritis induction

Under isoflurane anaesthesia, 40 μL Freund's Complete Adjuvans (FCA, Sigma Chemical Co., Sweden) was injected intraarticularly in the tibio-tarsal (ankle) joint using a 26G needle (method adapted from [4]). The FCA concentration was 1 mg/ml in Study I and Study II. In Study III, the animals run on the Paw Print setup were induced with 0.4 mg/ml FCA.

2.3 Paw Print Walkway

The Paw Print walkway consists of a black acrylic box, 21 cm tall, 10 cm wide and 100 cm long. Rats are trained to enter the dark walkway by a door through one short end of the box, make a swift, continuous passage and exit by a door at the other short end. The floor of the box, i.e. the walkway itself, is made of glass with light projected into the long edge via fiber optics. Projecting the light in this fashion allows for virtually complete internal reflection of the light. Only where an object, such as a rat paw, is allowed to touch the glass, light is scattered at the point of contact and can exit the glass, hence producing an illuminated pawprint [5]. The intensity of the illumination depends on the degree of contact against the floor and increases with the applied pressure [6]. A wide-angle video camera placed under the walkway to record the rats walking over the walkway and their illuminated pawprints. The recorded sequence can be manually scored and given a pain score between one and three (similarly to the box setup described below). Also, the Paw Print setup is equipped with gait detection algorithms for quantification of gait pattern and weight-bearing. However, only manual scores are presented here to allow for comparison with the box setup.

2.4 Box Recording

During the box recordings, three rats are placed in a 30×45 cm black acrylic box with glass flooring. The individual rats are separated from each other by black acrylic walls in 14×30 cm compartments, allowing only limited movement (adapted from [7]). Under the glass flooring a video camera is mounted and animals are recorded in 4-minute sessions. From the recordings, manual behavioral scoring is performed using a pain score scale from 0 to 3 (scoring scale from [1]), where 0 represents no behavioral response to pain, and score 3 means the animals do not use the injured paw at all.

2.5 Drug treatment

In Study II, morphine and dexamethasone was administered s.c. to all animals run in both the box recording and the Paw Print setup. Dexamethasone was administered 24 hours after FCA-monoarthritis induction. Rats run on the Paw Print walkway were given 0.78

µmol/kg and recorded 5 hours after drug injection. For the animals recorded in the box setup, the dexamethasone dose was 1 µmol/kg and they were recorded 6 hours after drug administration. Morphine was administered 48 hours after FCA-induced monoarthritis. The rats were given 12.5 µmol/kg morphine in the box recordings and 10 µmol/kg in the Paw Print setup. For morphine-administered rats, recordings were performed one hour after drug administration, in both setups.
Naproxen 100 µmol/kg was used in Study III. Both groups were administered naproxen perorally by gavage. The group in the box recording were filmed 6 hours after, and the Paw Print group 5 hours after drug administration. Naproxen was administered 24 hours after FCA-induction.

2.6 Statistics

Values are displayed as mean values ± SEM (standard error of the means) and the number of animals in each study is presented in parenthesis. Unless otherwise specified, Astute 1.9.1 in Microsoft Excel 9 for Windows 2000 was used for data analyses and the alpha level set to 0.05. In Study I, the non-parametric Kendall rank correlation coefficient test was used to determine the strength of the correlation between the scores given in the two setups. In order to produce a line fit for the curve, linear regression in Microsoft Excel 9, was used. In Study II and III, the non-parametric Wilcoxon-Mann-Whitney test was used to compare treated and untreated animals. In the tables an asterisk indicates significance. The tables show average percent analgesia produced by a drug compared to the vehicle control at the same time point . Percent analgesia was calculated with the formula:
100 - (score drug-treated group * 100 / average score vehicle treated group).

3 Results

3.1 Study I

The manual pain scores generated by the same set of animals on the two setups show a strong positive correlation (Kendall's τ =0.6 8) when plotting individual animals. Linear regression produced a good line-fit (R^2=0.984, y = 1.6945Ln(x) - 0.0638) revealing a logarithmic relationship between the two scores (Figure 1). Animals on the Paw Print walkway display higher pain scores.

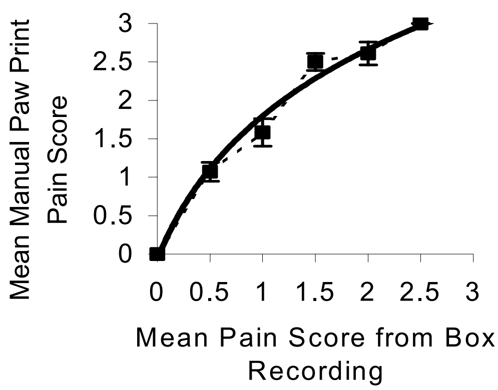


Figure 1. *The figure above shows the relationship between pain-scores derived from manually scoring the same rats on the Paw Print walkway versus in the box recording. The raw data, plotted with a dashed line, has been fitted with a trend line in solid black.*

3.2 Study II

In the box recordings, both morphine and dexamethasone produce significant analgesia at of a biologically relevant magnitude. In the Paw Print setup however, both drugs displayed low analgesic efficacy at non-significant levels (Table 1).

Table 1. *The table below illustrates the efficacy of morphine and dexamethasone on the Paw Print walkway and the box recording. The numbers in the cells indicate percent pain reduction compared to vehicle treated animals ± SEM. The numbers in parenthesis indicate the number of animals included in each study.*

	Paw Print	Box Recording
Morphine	15 ± 6.6 (9)	74 ± 11.9* (6)
Dexamethasone	5 ± 3.3 (10)	48 ± 0.0* (5)

3.3 Study III

Decreasing the FCA concentration by half produced comparable pharmacological sensitivity to naproxen in the two setups (Table 2).

Table 2. *The table below illustrates the efficacy of naproxen on the Paw Print walkway and the box recording. The numbers in the cells indicate percent pain reduction compared to vehicle treated animals. The numbers in parenthesis indicate the number of animals included in each study.*

	Paw Print (0.4 mg/ml FCA)	Box Recording (1 mg/ml FCA)
Naproxen	61 ± 15.0* (9)	59 ± 6.9* (5)

4 Discussion

The manual scores from the same animals in the Paw Print walkway and box recording were logarithmically related. The Paw Print scores were slightly higher. One might speculate that the pain experienced by the animals in the Paw Print setup is slightly different than in the box recordings and that this difference might not only be quantitative but also qualitative. Study II supports this hypothesis. Here, morphine and dexamethasone both show a diminished efficacy in the Paw Print setup compared to the box recording, indicating a difference in pharmacological sensitivity. Based on previous studies we do not believe this difference is simply an effect of slight differences in dose. In the Paw Print setup we have tried dexamethasone doses up to 2.55 µmol/kg and for morphine as high as 30 µmol/kg without reaching analgesia comparable to the levels seen in the box recordings here (unpublished findings). By decreasing the concentration of the induction agent by more than half in Study III, the sensitivity to naproxen is comparable between the two scoring methods.

The studies described herein show a difference in the two behavioral evaluation methods, both in behavioral endpoints and pharmacological sensitivity. One can hypothesize that the difference in the evaluated pain might be qualitative as well as quantitative. It is recognized that patients can differentiate between standing and walking pain, and a difference between the two can also be seen in rats [1]. It remains to be determined if the combination of these two models also can make a distinction between standing and walking pain.

References:

1. Coderre, T.J.; Wall, P.D. (1987). Ankle joint urate arthritis (AJUA) in rats: an alternative model of arthritis to that produced by Freund 's adjuvant. *Pain*, **28**, 379-393.
2. Hamers, F.P.; Lankhorst, A.J.; van Laar, T.J., Veldhuis, W.B.; Gispen, W.H. (2001). Automated quantitative gait analysis during overground locomotion in the rat: its application to spinal cord contusion and transection injuries. *J Neurotrauma*, **18**, 187-201.
3. Zimmermann, M. (1983). Ethical guidelines for investigations of experimental pain in conscious animals. *Pain*, **16**, 109-110.
4. Butler, S.H.; Godefroy, F.; Besson, J.M.; Weil-Fugazza, J. (1992). A limited arthritic model for chronic pain studies in the rat. *Pain*, **48**, 73-81.
5. Clarke, K.A. (1992). Disturbance of spatiotemporal footfall contact patterns in the rat by TRH analogue CG3703. *Neuropeptides*, **23**, 33-38.
6. Betts, R.P.; Duckworth, T. (1978). A device for measuring plantar pressure under the sole of the foot. *Eng Med*, **7**, 223-228.
7. Schött, E.; Berge, O.-G.; Ängeby Möller, K.; Hammarström, G.; Dalsgaard, C.-J.; Brodin, E. (1994). Weight bearing as an objective measure of arthritic pain in the rat. *J. Pharmacol. Toxicol. Methods*, **31**, 79-83.

The automation of observing and analyzing rodent behavior: possibilities and limitations

B.M. Spruijt
Utrecht University, Utrecht, The Netherlands

The increasing number of genetically modified animals to be characterized demands for the development of a reliable tool for observing and analyzing challenge-induced and baseline behavior. Since a functional interpretation of the gene effect on behavior is the ultimate goal of many phenotyping studies, an ethological rationale underlying testing is a logical consequence. So far batteries of already existing tests that have been applied involve two caveats:

- Behavior is induced in a novel environment and, thus, the limited observation time excludes assessment of habituation, baseline levels, rhythmicity, etc.
- Most tests have a specific focus on one motivational system, e.g. exploration, anxiety, etc., and a limited number of measurement parameters, whereas complex functions are the result of interacting motivational systems.

Recently, several systems have been developed equipped with hardware and software allowing observations over days. Since the animals are very precisely monitored by computers, behaviors have to be explicitly defined in terms of movement, position and form. Implementation, thus, requires redefining ethograms, which is an exciting process. Explicitly defining behavioral elements makes one aware of subjective influences in previously defined behaviors. The inevitable validation of data collected with new systems against data obtained with previously used well-known tasks can never be full proof as the latter tests have never been standardized themselves. Video imaging techniques also has disadvantages in that they do not cover neurological disturbances.

Apart from standardization and validation another issue emerges: the challenge resides not only in what can be measured, but how can large sets of data be reduced to a functional description of behavior and to what extent can relevant changes in complex behaviors of vast numbers of genetically modified animals be detected. For instance, the at first sight simple behavior "activity" consists of different components (speed, bouts, angularity, etc.) in combination with different spatial zones; modeling methods are required for an appropriate description and deduction of critical parameters, which then have to be subjected to statistical procedures such as principal component analysis and t-pattern analysis. In fact, a complex test requires complex statistics to have the full benefit of this approach. The fact that data can now be collected over days contributes significantly to the power of statistics.

Presently different systems are being developed these days and continuously extended in the near future with devices and sensors allowing the measurement of anxiety and conditioned place preference, conditioned place aversion, object recognition, etc. It is self-evident that the automated collection of data obtained in standardized and validated cages contributes to the reliability and reproducibility of such behavioral changes displayed by mutant mice. Different data sets illustrating and validating the applied techniques will be presented with a perspective on future: possibilities and limitations. In this symposium different systems which are being used are highlighted: IntelliCage and PhenoLab. Secondly, attention is paid to video imaging and the data streams generated by such an automated system and possibilities for analysis.

The 'mouse fitness centre': A standardized test system to assess motor function

S. S. Arndt

The non-automated multi-task testing apparatus mouse fitness centre (FC), allows testing a large number of animals within a short period of time. This apparatus is simple-to-do and enables the quick assessment of an important aspect of the general health of mice, namely motor (dys-)functions.

The system integrates eight tests that enable to detect motor dysfunctions:

1 The vertical pole.
2 Placing responseFootprint gangway.
3 Horizontal wire and –gimp.
4 Grip strength.
5 The inclined- and upside-down grid.

All motor tests should be performed in close succession in order to minimize testing-related stress in the mice. This is realized by installing all components of the FC on a so-called 'basis', which guides all necessary adaptations for performing the next test in a series. These adaptations are de.ned in a standard protocol that must be applied strictly. Working with 'standardized' animals (i.e. animals with similar motor abilities) increases the likelihood that results from different studies and laboratories are similar and that findings can be reproduced. To our knowledge no automated system exists covering the functionality of the FC. However, specific features of the FC probably can be automated.

The FC, as part of the assessment of the general health status, should proceed any further behavioral testing of mice (and other rodents). Pre-screening and detection of motor dysfunctions allows, for example, to select cognition tests that do not require the compromised motor function. Moreover, these data help to interpret results obtained by use of other tests and to put them into perspective.

Automated behavioral analysis of mice using IntelliCage: Inter-laboratory comparisons and validation with exploratory behavior and spatial learning

H.-P. Lipp, O. Litvin, M. Galsworthy, D.L. Vyssotski, A.L. Vyssotski, A.E. Rau, F. Neuhäusser-Wespy, H. Würbel, R. Nitsch and D.P. Wolfer

IntelliCage™ is a large home cage containing four complete operant conditioning units placed in the corners. A central computer continuously controls and monitors

activity and learning of up to 16 transponder-tagged mice per cage without human interference.
In a recent standardized multi-lab study, we noted identical strain rank order in the open-field, elevated Null-maze, water maze and object exploration, but comparison of absolute values revealed significant differences between laboratories, or strain-by-laboratory interactions. A subset of 46 female mice was also tested in 4 IntelliCages, located in different animal facilities of the University of Zurich (2 cages per site), each unit housing 11- 12 transponder-tagged mice from different strains. The system measured continually visits of drinking sides, development of place preference and activity parameters. IntelliCage revealed significant strain differences in initial exploratory activity and baseline activity during the first day. Most importantly, the behavioral scores of the strains as observed in both laboratories were statistically indistinguishable.
We then compared the first 10 min after introduction to IntelliCage with standard tests of exploration lasting about the same time span. The measure in IntelliCage was the number of visits to the yet unfamiliar test corners. It was positively correlated with visits to the center in the open-field. For the elevated Null-maze, the number of head dips correlated strongly with the number of corner visits in IntelliCage, while measures of anxiety such as time in the protected area did not correlate significantly. In the water maze (WM), we found a significant positive correlation between average escape latencies and place preference learning in IntelliCage. No correlations were found with WM probe trial scores. These findings indicate that either test measures the acquisition of spatial learning, but the WM testing took four man-weeks, while IntelliCage showed the same findings after 24 hours without presence of an experimenter.
We conclude that optimal standardization and comparability is best achieved by the use of automated procedures, and that such automation measures the same behavioral dimensions as standard manual tests, yet with much less stress for animals and experimenters.

Monitoring animal behavior in the Smart Vivarium

Serge J. Belongie

In the course of modern medical research, it is common for a research facility to house thousands of caged mice, rats, rabbits, and other mammals in rooms known as vivaria. In any experiment involving a group of animals it is necessary to perform environmental and physiological monitoring to determine the effects of the procedure and the health of the animals involved. Such monitoring is currently performed by human observers, and for practical reasons, only a small subset of cages can be inspected for limited amounts of time. In this talk, I will outline the computer vision and machine learning technology behind the Smart Vivarium, a system for automated, continuous animal behavior monitoring. The Smart Vivarium will serve as an invaluable tool for medical researchers as it will make better use of fewer animals. Early discovery of sick animals will prevent diseases from spreading, and in general will lead to more efficient caretaking of animals. Additionally, the proposed technology can serve as a powerful tool for monitoring sentinel cages in potential bioterrorism targets and chemical agent research facilities.
The Smart Vivarium project is a California Institute for Telecommunications and Information Technology (Calit2) collaboration between the Jacobs School 's Computer Science & Engineering and Bioengineering Departments and the UCSD Animal Care Program.

Multi-level analysis of mouse behaviour in a home cage environment using PhenoLab

L. de Visser, R. van den Bos and B.M. Spruijt

To contribute to the refinement of behavioral phenotyping methods for inbred and mutant mice, we developed a reliable tool for observing and analyzing behavior in a home cage-like environment (PhenoLab®, Noldus Information Technology, Wageningen, The Netherlands). Testing animals in their home cage environment holds several advantages; it allows for continuous observations over consecutive days and the evaluation of both challenge-induced and baseline behaviors. Home cage testing also minimizes human intervention (such as handling) and reduces interactions with other environmental factors not related to the behavioral test (such as animal transport). By carefully designing the home cage environment, different behavioral domains can be studied simultaneously, for example, by providing the animals with different stimuli and tasks (light, sound, novel objects, cognition tasks).
Here, data is presented of studies in inbred strains of mice (C57BL/6, DBA/2, C3H and 129S2/Sv) on locomotor activity and anxiety-related behavior in the PhenoLab system. First, strain differences in locomotor activity were dependent on the time of testing (novel vs. baseline conditions) due to differences in rate of adaptation to the environment. Second, Principal component analysis (PCA) revealed two major components within the domain of locomotor activity, which could be interpreted as ‘general activity’ (or: how active is the animal?) and ‘way of moving’ (or: when active, how is the animal moving through the cage?). Anxiety-related behavior was further studied by introducing an aversive light stimulus in the cage after an adaptation period of 4 days. The light stimulus illuminated the area around the feeding station, creating an approach-avoidance conflict for the animals.
The results of these experiments are discussed in the light of the possibilities and limitations of PhenoLab as a tool for the behavioral phenotyping of genetically modified mice.

Behavioral phenotyping: Getting more information from video tracking data

W.W. Kuurman

Mice are used worldwide as a research animal or model for a large variety of biological processes. The search for mice with interesting or exceptional phenotypes is important, so that this research can be done effectively. To facilitate this search PhenoLab®, (Noldus Information Technology bv, The Netherlands) was developed. In this PhenoLab several mice are individually housed for a week in a home cage and tracked using and overhead camera and EthoVision® (Noldus Information Technology bv, The Netherlands). Based on contrast differences, X- and Y-coordinates of the center of gravity for the mouse are recorded a number of times each second. During the week the mouse shows various spontaneous and induced behaviors. The behavioral phenotype of the mouse needs to be deduced from the X- and Y-coordinates (integration of location and movement).
Video tracking data is often summarized in time bins and then mean of each time bin is used for statistical analysis.

For data which does not have a known unimodal distribution, however, mean and standard deviation ignore variation in the data. This variation can be utilized using all data available and mathematical modeling of each variable. Parameter estimates from these models can then be used as phenotypic information for each mouse tested. This approach yields high-resolution phenotypic information, so that detection of mice with interesting or exceptional phenotypes is more successful.

Some examples of this approach are velocity and distance to home cage wall data. For velocity a mathematical model was developed to distinguish low, intermediate, and high velocities. This model can be used to analyze locomotive activity in greater detail. For distance to home cage wall a mathematical model was developed to determine how often the mouse stays close to the wall, where the border is between being close and being away from the wall, and how these measures change over time. This model can be used to analyze anxiety and habituation in greater detail.

Additional models should be developed to utilize information collected by video tracking software. The future challenge is to integrate information obtained with this approach, so as to facilitate interpretation of the data.

The Erasmus Ladder: a new tool for the automated measurement of motor performance and motor learning in mice

A. Cupido, S.P. Krygsman, C.M. Snoeck, J.A.Th. Bos, C.I. De Zeeuw and S.K.E. Koekkoek

Recent developments in the creation of mouse models for human diseases have called for tools that are able to quickly screen mutant mice for their deficits. In our department, we have developed two successful tools for detecting mouse mutants with cerebellar deficits: eye blink conditioning and VOR adaptation. However, for screening large amounts of mice these systems are too time consuming, as they require specialized and invasive surgery. Therefore we developed a new tool: The Erasmus Ladder.

The Erasmus Ladder is a horizontal ladder with two cages. The horizontal ladder has 25 rungs. The rungs are divided in a left side and a right side. All subdivisions of the rungs are equipped with pressure sensors. The rungs can be automatically protracted and retracted.

The mice are trained to walk with a constant speed from one cage to the other. We have developed software to analyze the walking pattern of the mice.

In this presentation we show quantitatively the level of the motor performance and motor learning of mice with cerebellar deficits.

The advantages of the Erasmus Ladder are fourfold. First, the experiments do not require surgery; second, the Erasmus Ladder will be completely computer controlled, thus the experiments are not time consuming; third, the Erasmus Ladder has a feedback control, which allows us to modify the position of rungs during ongoing locomotion; fourth, with a few adjustments it is also possible to screen for hippocampal and amygdaloidal deficits.

The "mouse fitness centre": an efficient, standardized scan of the physical condition and sensorimotor integrity of mice

S.S. Arndt
Anatomy II, Cognitive Neurobiology, University of Cologne, Cologne, Germany
Present address: Laboratory Animal Science, Utrecht University, Utrecht, The Netherlands

Abstract
The "mouse fitness centre" (FC) [1] is a multi-task testing system that allows for efficient physical testing of large numbers of mice. These tests are simple-to-do and enable the assessment of motor (dys-) functions, an important aspect of the general health of mice.
The system integrates eight tests: the vertical pole, placing response, footprint gangway, horizontal wire and –gimp, grip strength, and the inclined- and upside-down grid. All tests are performed in one integrated procedure in order to minimize testing-related stress in the mice. The FC, as part of the assessment of the general health status, should proceed any further behavioral testing of the animals. Pre-screening and detection of motor dysfunctions is inevitable to select appropriate follow-up tests that do not require the compromised motor function. Moreover, FC data help to interpret results obtained by use of other tests and to put them into perspective.
To our knowledge no automated system exists covering the functionality of the FC. However, specific features of the FC eventually can be automated.

Keywords
Mice, health assessment, motor functions, behavior, phenotyping.

1 Introduction
"Mens sana in corpore sanum" [based on Juvenalis 98 – 128 A.D.] points out the idea behind assessing the health status and sensory and motor abilities. Only healthy mice will be able to meet the physical requirements of many standard behavioral tests. If they prove to be unable to meet these requirements, special tasks have to be selected that are not depending upon the compromised sensory and/or motor functions. One of the first general health assessment routines was published by Irwin in 1968 [8] to evaluate the response of mice to psychoactive drugs. It can also be used to assess the general heath of, for example, genetically engineered mice, or of mice with experimentally induced brain lesions. The Irwin protocol includes, among others, assessment of the body position, bizarre behaviors, transfer arousal, locomotor activity and respiration rate. Another test battery, the SHIRPA protocol, is a multi-tired approach to the general examination of mice [17]. This protocol consists of preliminary observations such as sensory and muscle functions, autonomic functions, lower motor neuron functions, exploratory behavior, histology, biochemistry, more complex behaviors and physiological measurements. Taken together, these protocols enable a very detailed analysis of the animals` general heath status.

1.1 The mouse fitness centre
The need to detect efficiently motor dysfunctions in laboratory mice prior to any further behavioral testing resulted in the development of the FC.
To allow for the performance of all motor tests within one procedure, all test components of the FC are installed on the so-called "base" unit. All necessary adaptations of the apparatus for the next test to be performed are guided by the basis and are defined in a standard protocol that must strictly be adhered to. Intertest intervals because of modifications of the test apparatus are reduced to a minimum. This system has been developed to realize the idea of a standardized testing system [18] that makes testing less time-consuming for the researcher and less stressful for the animals.

2 Materials and methods

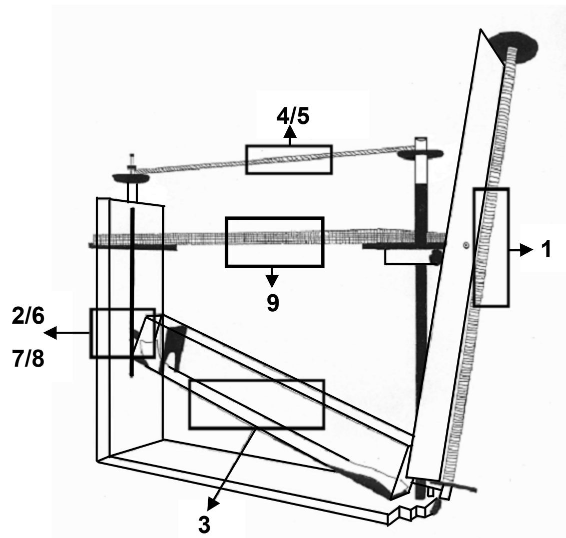


***Figure 1**. Fitness centre basis equipped with horizontal gimp (4), safety net (9) vertical pole (1) and footprint gangway (3) which can be replaced by the grip strength meter (6; not shown), the placing response grid (2; not shown), the upside-down (8; not shown) or inclined -grid (7; not shown).*

All tests are run in close succession i.e. one integrated procedure. The animals might be given a resting period of e.g. 3 minutes in-between tests, depending on their health constitution or other strain peculiarities that otherwise may interfere with the testing procedure. The tests are run in a fixed order as described below.

2.1 Horizontal wire manoeuvre
To measure motor strength, the horizontal wire manoeuvre is performed [5,8,10]. To hold its body suspended, a mouse requires grip strength and balance. A horizontally positioned wire (Ø: 7 mm, length: 81 cm), attached to the FC 27 cm above the table on which the apparatus stands is used (see Figure 1, frame 4). To prevent harm in the event of a fall, a safety net is placed 15 cm beneath the wire. The

animal is held above the wire by tail suspension and is lowered to allow the forelimbs to grip the wire. The mouse is held in extension and rotated around the horizontal and released. If the mouse does not manage to grasp the wire with its hind paws within 2 minutes it is put back in its homecage and given a 3 minutes rest period before the test is repeated. If the animal does not manage to grasp the wire with its hind paws within 2 minutes during the second attempt it is put back in its homecage and given a 3 minutes rest period before the test is repeated for the third and last attempt.

2.2 Horizontal gimp manoeuvre

To measure motor strength, the horizontal gimp manoeuvre is performed [5,8,10]. To hold its body suspended, a mouse requires grip strength and balance.
A horizontally positioned gimp (Ø: 0.05 mm, length: 81 cm), attached to the FC 27 cm above the table on which the apparatus stands is used (see Figure 1, frame 4). The animal is held above the gimp by tail suspension and is lowered to allow the forelimbs to grip the wire. The mouse is held in extension and rotated around the horizontal and released. If the mouse does not manage to grasp the gimp with its hind paws within 2 minutes it is put back in its homecage and given a 3 minutes rest period before the test is repeated. If the animal does not manage to grasp the gimp with its hind paws within 2 minutes during the second attempt it is put back in its homecage and given a 3 minutes rest period before the test is repeated for the third and last attempt.

2.3 Vertical pole

To measure motor coordination and balance, the vertical pole is used [5,10] that measures negative geotaxis, a reflex behavior. It requires balance and grip strength [5,10]. A cord-wrapped pole (Ø: 16 mm; length: 81 cm) with a platform (Ø: 9 cm) catered on top is used (see Figure 1, frame 1). The animal is removed from the home cage by holding the tail between thumb and index-finger. The pole is held in horizontal position, the mouse is placed with its forelimbs adjacent to the platform in the centre of the pole. Then, the pole is tilted into a vertical position with the mouse facing the floor. Mice with intact motor functions will stay on the pole and may walk up or down the length of the pole. The cut-off time in this test is 2 minutes. The procedure is repeated three times successively.

2.4 Grip strength

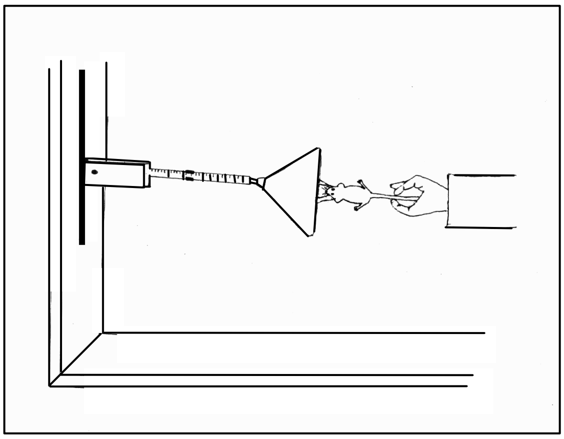


Figure 2. Mouse gripping the trapeze and gently horizontally pulled backwards.

This is a test of neuromuscular function, which measures the force required to make a mouse release its grip [8,17].
A spring scale equipped with a wire (Ø: 0.2 mm) triangle (trapeze) is used (see Figure 2). The animal is gripped by its tail as described before. Then, the animal is lowered and allowed to grip the trapeze with its forepaws. A gentle horizontal backwards pull is applied. The animals try to stop this involuntary backward movement, until the pulling force overcomes its grip strength. After the animal loses its grip on the grasping trapeze, the grip strength can be read off the spring balance.

2.5 Upside down grid

Neuromuscular abnormalities can be detected by measuring motor strength. Balance and grip strength are required for a mouse to hold its body suspended and are measurable by performing the upside-down grid test [5,172]. The grid (width: 23 cm, height: 35 cm) is made of stainless steel wire (Ø: 0.2 with grid holes of 0.5 x 0.5 cm). A frame (width: 6 cm) made of grey PVC and placed around the perimeter of the grid prevents the mouse from walking off the edge (see Figure 3). The grid is held in a horizontal position. A mouse is placed on it approximately in the centre. Then the grid is turned upside down (180°). The upside down grid is held at a height approximately 28 cm above the table, high enough to prevent the mouse from easily climbing down. For this test the cut-off time is 30 seconds.

2.6 Inclined grid

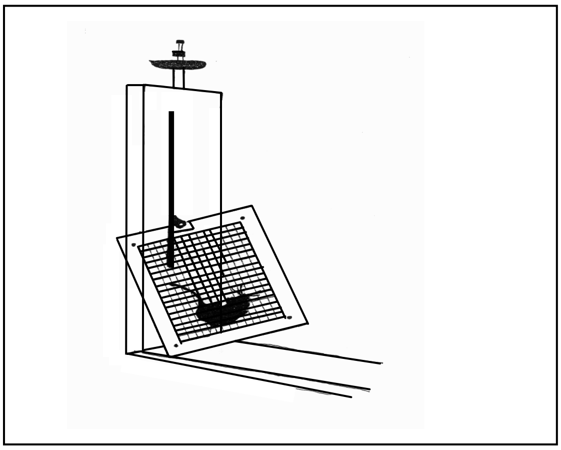


Figure 3. *Mouse up to turn on the inclined grid.*

Negative geotaxis, a reflexive behavior that requires balance and grip strength, can be observed using the inclined grid. The same apparatus (grid) as for the upside down grid (see 2.5) is used (see Figure 3). Turning on the inclined grid is assessed by a modification of the procedure described by Marshall [13]. The grid is held in a horizontal position. A mouse is placed on it approximately in the centre. The nose of the animal points to the edge of the grid that is to be lowered. Then, the grid is moved until it attains a negative inclination of 30° with respect to the horizontal plane. For this test the cut-off time is 60 seconds. The procedure is repeated three times successively.

2.7 Placing response

This response represents a reflex behavior and is sensitive to damage to corticospinal systems [8,17,19]. The same apparatus (grid) as for the upside down grid test (see 2.5) and for the inclined grid test (see 2.6) is used (see Figure

3). The experimenter can manipulate the inclination of the grid by moving the opposite side up or down. The grid is attached to the FC 28 cm above the table on which the apparatus stands (see Figure 3). The grid is tilted horizontally. The animal is held by its tail as described for the vertical pole test. Then the animal is lowered slowly from a height of approximately 15 cm towards the grid.

2.8 Footprint gangway

Neurological dysfunction can be assessed by analyzing footprint patterns and walking tracks. A sloping gangway (length 70 cm, width 5 cm, with 5 cm high side walls) is used. The floor of the gangway is lined with ordinary white paper. The gangway forms an angle of 30° with respect to the horizontal plane (see Figure 1, frame 3). To obtain the footprints, the hind feet are dipped into non-toxic food dye. The animals walk up the gangway into a dark compartment. After a run, an animal is placed in a cage with 1 cm warm water for approximately 1 min in order to wash off the dye.

The footprints are scanned and evaluated using the FOOTPRINTS [11] program. The measures of interest are obtained by clicking with the computer mouse on particular points of the digitized print. The program calculates the measures and stores them in a table, which can be imported into commonly used calculation and statistics programs. The output parameters by the program are: number of steps, additional steps and slidings, toe spred 1 to 5 and 2 to 4, length of the foot step, stride length and width, the foot angle, and the areas touched by single steps, additional steps, and by slidings, respectively.

3 Scoring

Test	Scoring
Vertical pole	Latency [sec] to turn 180º Descend: 0 = no; 1 = yes Dropping off 0 = no; 1 = yes Sliding down 0 = no; 1 = yes Reclimbing: 0 = no; 1 = yes
Placing response	0 = none 1 = upon nose contact 2 = upon vibrissae contact 3 = before vibrissae contact (distance to lid approximately 18 mm) 4 = early vigorous extension (distance to lid approximately 25 mm)
Wire-/gimp manoeuvre	0 = active grip with hind paws 1 = difficulty to grasp with hind paws 2 = unable to grasp with hind paws 3 = unable to lift hind legs, falls off after 3 seconds have elapsed 4 = falls off within 3 seconds
Grip strength	Grip strength [N]
Inclined grid	Latency [sec] to turn 180º: three times. If animal does not turn within 60 seconds. The performance is awarded a score of 0
Upside down grid	0 = falls off immediately (< 2 sec) 1 = grip with forepaws only 2 = grip with all 4 paws

***Table 1.** Scoring scheme.*

Data are input during testing into a Microsoft-Access-Database. If necessary, peculiarities are recorded.

4 Statistical analysis

For tests with non-parametrical data that are performed three times, such as vertical pole -sliding, -descend and –drop-off, a sum score (rating) is calculated. The data are analyzed non-parametrically with Kruskal-Willis Tests (Chi-Square Approximation), Wilcoxon Two-Sample Test and Kolmogorov-Smirnov Two-Sample Test.

For tests with parametrical data that are performed three times, such as latency to turn 180º on the vertical pole and latency to turn 180º on the inclined grid, the mean of the three recordings is taken for analysis. The data are analyzed by an analysis of variance (ANOVA). For all analyses a difference between test animals and their controls is considered to be absent when its P-value exceeds 0.05.

The data derived from the footprint testing are also analyzed by ANOVA. In addition, post-hoc Fisher's Least Significant Difference (LSD) comparisons are performed to analyze e.g. the differences between groups in more detail. A difference between groups is considered significant if the P-value is below 0.05.

5 Discussion

The scientific literature provides several examples especially of mutant [6,7,12,14,15] and inbred [2,16] mouse strains that suffer from sensory and/or motor deficits that in turn affect their behavior. A striking example is given by Contet et al. [3]. They tested 129S2/SV and C57BL/6 mice in the Morris water maze (MWM). The 129S2/SV animals showed a longer escape latency than the C57BL mice. However, no differences in the distance moved were detected between the two strains. Most likely this was due to a slower swim speed of the 129S2/SV mice. A poor swimming performance may be indicative for motor impairments. However, as shown e.g. by Klapdor and van der Staay [9] some mouse strains show floating behavior in the MWM. The BALB-mouse, for example, does not suffer from motor impairments but develops floating as a swim strategy in contrast to other mouse strains tested in the same study. These findings illustrate that a detailed investigation of differences in swim speed and their underlying motor deficits must be part of the behavioral analysis to allow for a proper interpretation of the results.

Such a detailed investigation is given e.g. by Costa et al. [4] who, for this purpose, carried out analyses of grip strength, RotaRod performance and footprint patterns.

Tests of learning and memory usually require intact motor functions and rely on sensory (e.g. visual or olfactory) cues. A blind mouse is useless for any kind of visual discrimination task, a deaf mouse cannot be cued with an auditory tone and an anosmic mouse shows no response to olfactory cues [5]. The general health assessment gives a detailed picture of the animals' physical status. If existent impairments remain undetected, the results of subsequent tests might be interpreted incorrectly.

It may be an advancement to equip the different tests of the FC with sensors that allow an automated collection of data on motor (dys-) functions.

I am indebted to Frauke Ohl and Franz Josef van der Staay for a critical revision of the manuscript. Jens Alber is acknowledged for technical assistance. David Surjo is acknowledged for his *technical expertise. Hansjörg Schröder is acknowledged for his support.*

References

1. Arndt, S. (2004). Phenotypic characterisation with special emphasis on behaviour, circadian rhythmicity and metabolism considering an Insulin receptor knockout mouse model as example. *PhD-thesis.*
2. Brown, R.E.; Stanford, L.; Schellinck, H.M. (2000). Developing standardized behavioral tests for knockout and mutant mice. *ILAR. J.*, **41**, 163-174.
3. Contet, C.; Rawlins, J.N.; Bannerman, D.M. (2001). Faster is not surer - a comparison of C57BL/6J and 129S2/Sv mouse strains in the water maze. *Behav. Brain Res.*, **125**, 261-267.
4. Costa, A.C.; Walsh, K.; Davisson, M.T. (1999). Motor dysfunction in a mouse model for Down syndrome. *Physiol. Behav*., **68**, 211-220.
5. Crawley, J.N. (2000). *What's wrong with my mouse?* Wiley-Liss.
6. Del Punta, K.; Leinders-Zufall, T.; Rodriguez, I.; Jukam, D.; Wysocki, C.J.; Ogawa, S.; Zufall, F.; Mombaerts, P. (2002). Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature*, **419**, 70-74.
7. Ferguson, G.D.; Anagnostaras, S.G.; Silva, A.J.; Herschman, H.R. (2000). Deficits in memory and motor performance in synaptotagmin IV mutant mice. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 5598-5603.
8. Irwin, S. (1968). Comprehensive observational assessment: Ia. A systematic, quantitative procedure for assessing the behavioral and physiologic state of the mouse. *Psychopharmacologia*, **13**, 222-257.
9. Klapdor, K. ; Van der Staay, F.J. (1996). The Morris water-escape task in mice: strain differences and effects of intra-maze contrast and brightness. *Physiol Behav.*, **60**, 1247-1254.
10. Klapdor-Dulfer, K. (1996). On the assessment of functional deficits in rodent models of cerebrovascular disorders. *PhD-Thesis.*
11. Klapdor, K.; Dulfer, B.G.; Hammann, A; Van der Staay, F.J. (1997). A low-cost method to analyse footprint patterns. *J. Neurosci. Methods*, **75**(1), 49-54.
12. Lalonde, R.; Strazielle, C. (2003). Motor coordination, exploration, and spatial learning in a natural mouse mutation (nervous) with Purkinje cell degeneration. *Behav. Genet.*, **33**, 59-66.
13. Marshall, J.F. (1982). Sensorimotor disturbances in the aging rodent. *J. Gerontol.*, **37**, 548-554.
14. Montag-Sallaz, M.; Montag, D. (2003). Severe cognitive and motor coordination deficits in tenascin-R-deficient mice. *Genes Brain Behav.*, **2**, 20-31.
15. Paylor, R.; Hirotsune, S.; Gambello, M.J.; Yuva-Paylor, L.; Crawley, J.N.; Wynshaw-Boris, A. (1999). Impaired learning and motor behavior in heterozygous Pafah1b1 (Lis1) mutant mice. *Learn. Mem.*, **6**, 521-537.
16. Picciotto, M.R.; Wickman, K. (1998). Using knockout and transgenic mice to study neurophysiology and behavior. *Physiol. Rev.*, **78**, 1131-1163.
17. Rogers, D.C.; Fisher, E.M.; Brown, S.D.; Peters, J.; Hunter, A.J.; Martin, J.E. (1997). Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. *Mamm. Genome*, **8**, 711-713.
18. Van der Staay, F.J.; Steckler, T. (2001). Behavioural phenotyping of mouse mutants. *Behav. Brain Res.*, **125**, 3-12.
19. Whishaw, I.Q.; Haun, F.; Kolb, B. (1999). Analysis of behavior in laboratory rodents. *Modern Techniques in Neuroscience*, 1243-1275.

Automated behavioral analysis of mice using IntelliCage: inter-laboratory comparisons and validation with exploratory behavior and spatial learning

H.-P. Lipp [1], O. Litvin[2], M. Galsworthy [1], D.L. Vyssotski [2], A.L. Vyssotski [1], P. Zinn[1] , A.E. Rau [1], F. Neuhäusser-Wespy [1], H. Würbel[3], R. Nitsch [2] and D.P. Wolfer [1].

[1] *Institute of Anatomy, Division of Neuroanatomy & Behavior, University of Zürich, Switzerland*

[2] *Division of Psychiatry Research, University of Zürich, Switzerland*

[3] *Animal Welfare & Ethology, Institute of Veterinary Physiology, University of Giessen, Germany*

Abstract

IntelliCage^{TM} is a large home cage containing four complete operant conditioning units placed in the corners. A central computer continuously controls and monitors activity and learning of up to 16 transponder-tagged mice per cage without human interference over hours, days or weeks. A common problem in behavioral phenotyping of mice is that genetic differences are inconsistent across laboratories. This is partially due to uncontrollable interactions between experimenter, variations of experimental set-up and variable stress reactions of mice being exposed singly to a variety of stimuli and situations. Here we compare mice that have been raised under identical conditions but then tested in two different laboratories for exploratory behavior and spatial learning, first by manual testing using identical test set-ups, and then in the automated IntelliCage. The data presented here illustrate that automated testing provides much more consistent data across laboratories, and they show that IntelliCage measures at least partially the same behavioral traits as observed in standard tests of mouse exploratory behavior and spatial learning.

Keywords

Mouse – Cognition – Behavioral phenotyping – Automation – Transponder tagging

1 What is IntelliCage

Figure 1 shows a single IntelliCage housing 8-16 mice carrying commercially available microchips for identification (TROVAN and DATAMARS). One controller computer can handle up to 8 cages, monitoring or testing a maximum of 128 mice, each one individually, multitasking for different tests possible. The flow of information between computers and test cages is shown in Figure 2.

Essentially, an IntelliCage is a behavioral micro-laboratory permitting automated individual assessment of spontaneous and cognitive behavior typical for mice (Table 1). Each cage contains four identical test corners accommodating one mouse at a time only. A mouse entering a test corner is identified by passive transponders implanted subcutaneously. Depending on schedule, the computer can reward correct responses by opening access to the liquid bottles, or it may refuse access or even punish incorrect choices by delivering air-puffs. Spontaneous behavior such as basic activity levels, circadian variations, exploratory behavior, neophobia and behavioral stereotypies are concomitantly assessed in most testing modules.

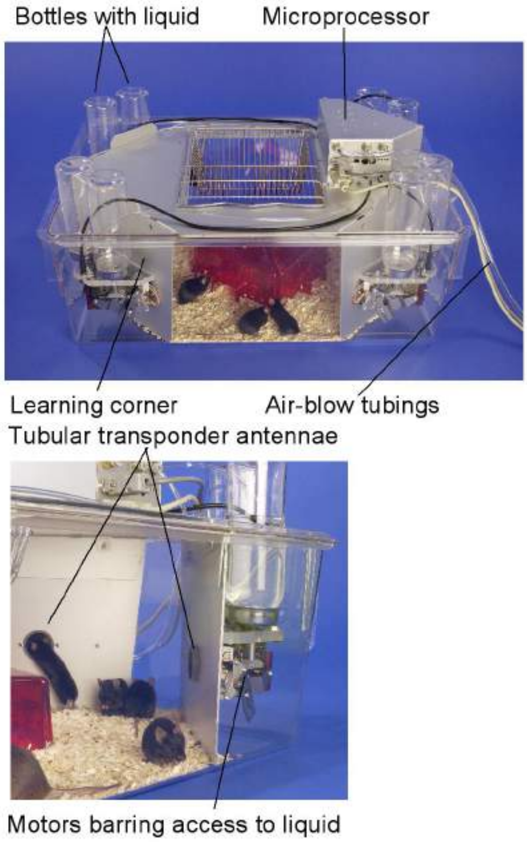


Figure 1. *IntelliCage housing 8-16 mice over days and weeks*

Social hierarchies and changes thereof can be assessed by competition for reward after programmable deprivation periods. Taken together, the IntelliCage system represents an almost universal tool for fully automated behavioral screening yet allows, for expert users, sophisticated behavioral analysis for mechanistic purposes.

IntelliCage operates most efficiently over long periods with female mice while males will sooner or later require the insertion of compartment barriers housing 2-3 males. However, this limits merely the analysis of spatial behavior. Contrary to a widely held belief, behavioral analysis is not confounded by social interactions, and we have shown recently that standardization of behavioral testing in mice is not prevented by enriched environment.

It should be emphasized that mice need not to be kept permanently in IntelliCages. Test cohorts can be formed and kept in ordinary cages. Typical test programs last one week. Normally, the animals adapt very rapidly when being re-introduced to the system. This allows for efficient long-term monitoring.

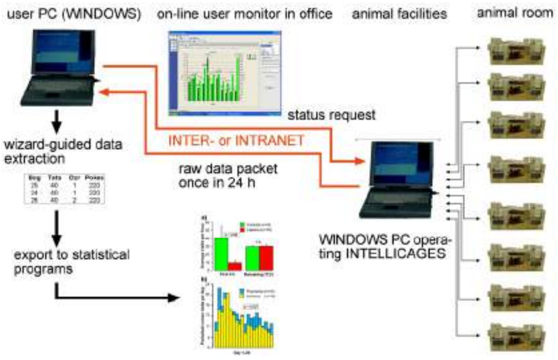


Figure 2. Organization of an IntelliCage test battery

On the other hand, IntelliCage has not been designed to replace conventional behavioral laboratories. It simply relegates the chore of monotonous manual testing to screening robots, freeing expertise for detailed analysis of behavioral problems found in IntelliCage.

Table 2: Examples of programmmable behavioral tests

Spontaneous behavior
Basic activity levels, circadian activity
Anxiety, neophobia and exploration
Spatial and temporal
Stereotypical place preferences
Spatial preference and avoidance learning
Spatial reversal learning
Spontaneous alternation
Temporal conditioning
Temporo-spatial conditioning
Systematic patrolling schedules
Radial maze-like patrolling
Discrimination learning & preferences
Visual discrimination
Gustatory discrimination learning
Spontaneous drug preference or avoidance
Memory
Procedural memory
Habituation
Spatial short-term (working memory)
Visceral/gustatory memory
Social & others
Competition rank orders
Approach-avoidance conflicts
Operant conditioning
Procedural learning
Fixed ratio conditioning (motivation)
DRL (Differential reinforcement of low responding, response inhibition, timing)

Table 1. Examples of programmable behavioral tests

2 Inter-laboratory comparisons

A main problem in analyzing mouse behavior is that treatment effects and genetic differences can be replicated in one laboratory but not in another. This is most obvious in strain comparisons [1,3].

In a recent standardized multi-lab study, we noted identical strain rank order in the open-field, elevated Null-maze, water maze and object exploration [2] but comparison of absolute values revealed significant differences between laboratories, or strain-by-laboratory interactions. A subset of 46 female mice was also tested in 4 Intellicages, located in different animal facilities of the University of Zurich (2 cages per site), each unit housing 11-12 transponder-tagged mice from different strains.

The system measured continually visits of drinking sides, development of place preference and activity parameters. IntelliCage revealed significant strain differences in initial exploratory activity and baseline activity during the first day. Most importantly, the behavioral scores of the strains as observed in both laboratories were statistically indistinguishable.

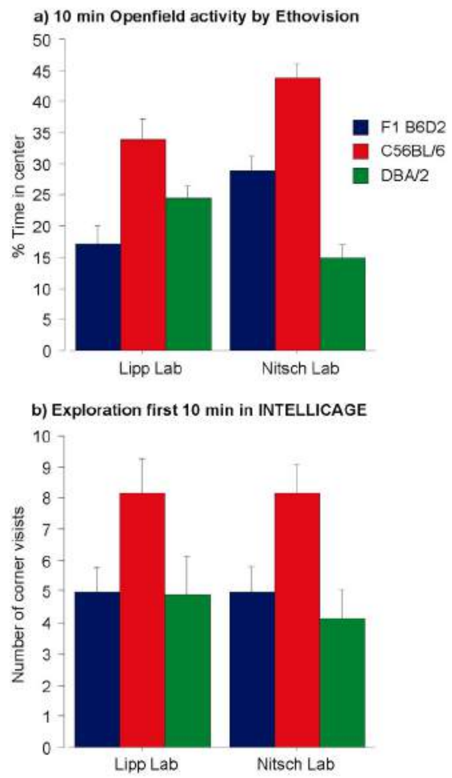


Figure 3. Inter-laboratory comparisons of exploratory behavior. a) Standard open field as assessed by video-track analysis (Noldus Ethovision 3.0). Note the typically high activity of strain C57BL/6, but also significant differences between laboratories and strain-by-laboratory interactions. b) Same mice tested later in IntelliCage systems. Note the uniform scores between the laboratories.

3 Water maze and IntelliCage

The perhaps most widely used test in assessing cognition of rodents is the water maze test, in which the animals must learn the position of a platform hidden under the water surface. Although popular in mice, one should recognize that this test does not simply measure cognition. In brief, it measures both the capacity of procedural learning (how to get to the platform) and, concomitantly, the development of spatial memory. The presence of the latter is usually assessed by means of a probe trial during which the animal searches for the removed platform.

The 46 mice of the multi-lab study had undergone an abbreviated version of water maze learning, performing 4 consecutive trials per day, during 4 days. This was followed by a probe trial on the fifth day. For procedural details, see [2]. As in the exploration studies, identical test set-ups were used in the two laboratories.

As in the exploratory tests, water maze learning revealed significant strain differences, the B6D2 F1 hybrid mice learning best, and no statistical difference being observed between the two inbred strains C57BL/6 and DBA/2.

The same animals were tested in IntelliCage after having adapted to the system during 4 days during which they learned to operate, by nose-pokes, the gates barring access to water.

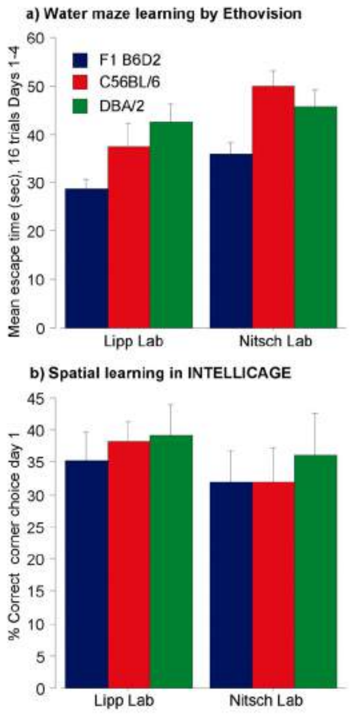


Figure 4. Inter-laboratory comparison of spatial learning. a) Watermaze learning as assessed by video-track analysis (Noldus Ethovision 3.0). Note the typically much shorter escape latencies of the hybrid mice, and significant differences between laboratories for all strains. b) Same mice tested later in IntelliCage for development of spatial preference to the only corner allowing drinking. These corners were different for different animals. Chance level for corner visit is 25%. Bars and whiskers indicate Mean and S.E.M.

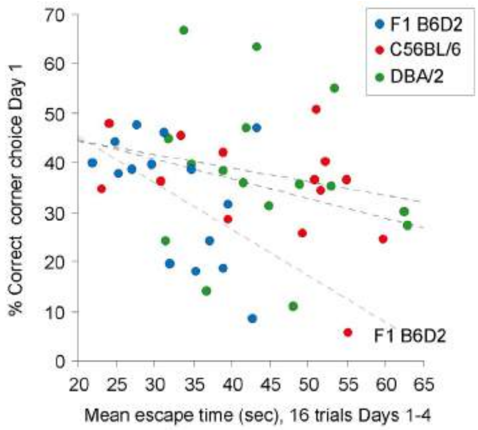


Figure 5. Correlation between average escape latency in the watermaze and the percentage of correct corner choices during the first day of spatial learning in the IntelliCage, data from the two laboratories pooled. This pooling increased variability and reduced correlations. The predicted variable (percent correct choices in IntelliCage) is corrected for overall activity. With the exception of the DBA/2 mice, there is a clear trend in that mice with short escape latencies in the water maze show higher corner preference, best visible in the B6D2 hybrid mice.

Afterwards, mice had to learn to obtain water from one corner only. The mice soon showed increasing preference for visiting the corner were they still had access to water. However, they still visited other corners. Comparison of strains and laboratories revealed neither strain differences, nor differences between laboratories, and no strain-by-laboratory interactions (Figure 4).

Correlation analysis showed that averaged measures from the water maze were partially predicting later spatial preference in IntelliCage, chiefly for the first day of spatial preference learning. Significant correlations were found between the average escape latency and the number of correct corner visits ($r = -0.32$, $p < 0.05$), mice escaping faster showing higher numbers of correct visits. Another water maze variable predicting the degree of spatial preference was swim speed ($r = 0.43$, $p = 0.003$). On the other hand, the probe trial scores did (expectedly) not correlate with the degree of spatial preference on the first day of learning. As illustrated in Figure 5, the correlations were strain-dependent, best visible in the F1 B6D2 hybrid mice, and lacking in DBA/2 mice.

One should note that this analysis is preliminary and was conducted using very global scores only. A refined analysis will need better temporal matching, e.g., using scores from the first day of water maze learning. At present, it allows to conclude that several behavioral scores of water maze learning characterizing escape performance (but not spatial memory) are predicting the efficiency of mice learning place preferences in the IntelliCage. As the motivations underlying spatial learning in the water maze and IntelliCage are different (aversion in the water maze, reward in IntelliCage), the observed correlations imply a common process underlying at least partially general spatial learning.

4 Discussion

4.1 Comparability

In our mice subjected to exploratory tests, automated home-cage testing in IntelliCage provided much more reliable and comparable results than conventional testing of individual mice, even when the latter tests were standardized stringently. By and large, both tests recognized the highly active C57BL/6 strain, but manual testing showed strain-by-laboratory interactions that are usually non-interpretable. For example, two strains showed higher open-field activity in one laboratory (Nitsch), but one strain (DBA/2) showed significantly less activity than the DBA/2 mice in the Lipp laboratory. In contrast, IntelliCage showed identical scores for all three strains in both laboratories. The variability of scores in the home cage system appears somewhat higher, but one should note that the stress level in a social group of mice living together for long time is less than in the typical open-field situation. Stress, however, decreases activity scores, particularly in early phases of exposures to novel environments, and may thus have exerted a flooring effect, that is, generally lower scores entailing reduced variability.

4.2 Social housing

A common but unsubstantiated belief in experimental psychology is that animals must be tested individually, separated from others, in order to obtain comparable results. Testing in groups is believed to suffer from interactions between animals, resulting in biased test scores. The data here show that this is clearly not the case. The IntelliCage system recognized precisely the same strain differences and activity levels in both locations. One should note that every cage contained individuals from the three strains. Social interactions between animals familiar with each other would be expected to produce unpredictable variation obscuring standardization and comparability between laboratories. The opposite was found here. These observations are in agreement with learning studies of socially living transponder-tagged mice in semi-naturalistic environments where clear genetic differences emerged and persisted (4).

Likewise, unpublished data from our laboratory indicate that social housing does not interfere with the assessment of treatment effects and genotypes.

4.3 Common behavioral mechanisms

The data presented here show that IntelliCage measures, at least partially, the same behavioral dimensions as in the classic behavioral tests. For example, frequent crossing of the central zone in an open-field is commonly considered as an indicator of exploratory behavior and reduced agoraphobia. The significant correlation of this score with the number of corner visits in IntelliCage fits this notion well, since entering the tubes leading to the test corner is probably a more direct measure of explorative tendencies than overcoming the typical initial wall hugging of mice placed in an open-field arena. Likewise, the observed correlations between spatial learning scores in the water maze and place (corner) preference in IntelliCage indicate some common underlying mechanisms, although these observations warrant further detailed analysis.

4.4 Economy of testing

Behavioral testing of individual animals is laborious. Even using highly standardized procedures and computer-assisted video-tracking of 4 mice at a time during 30 min, the manually conducted open-field analysis of 46 mice required almost two workdays (9 hours for testing, 6 hours of off-line data analysis). In contrast, the same task, using 4 IntelliCages, took one hour. Thus, the workload for the experimenter is reduced 15 times. The savings in terms of time and efforts are even more pronounced in the case of water maze learning. The employed schedule was optimized for fast assessment of learning. Yet it took 2 experimenters for four days each, plus one day for data analysis, a total of 10 man-days or 80 h. On the other hand, testing the mice for development of spatial preferences required to set-up the IntelliCage by loading a program module, then leaving the mice for 24 hours under the supervision of a computer. Downloading the data and performing a routine analysis can be done in less than an hour. The effort of the experimenter is thus reduced at least 40 times.

5 Conclusions

We conclude that optimal standardization and comparability of behavioral testing in mice is best achieved by the use of automated procedures, and that such automation measures, at least partially, the same behavioral dimensions as standard manual tests for exploration and spatial learning, yet with much less stress for animals. In terms of animal welfare, the presented data show that behavioral assessment in social groups does not interfere with recognition of even subtle genetic differences. Economically, the use of computer-controlled social home-cage test settings such as IntelliCage provides more than substantial savings in terms of human resources.

Supported by Swiss National Science Foundation, the NCCR "Neural Plasticity and Repair" and the Foundation 3R. We thank Inger Drescher-Lindh and Rosmarie Lang for expert support in testing animals.

References

1. Crabbe, J.C.; Wahlsten, D. ; Dudek, B.C. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science*, **284**, 1670-1672.
2. Vyssotski, A.L.; Dell'Omo, G.; Poletaeva, I.I.; Vyssotski, D.L., Minichiello, L.; Klein, R.; Wolfer, D.P.; Lipp, H.-P. (2002) Long-term monitoring of hippocampus-dependent behavior in naturalistic settings: mutant mice lacking the neurotrophin receptor TrkB in the forebrain show spatial learning but impaired behavioral flexibility. *Hippocampus*, **12**, 27-38.
3. Wahlsten, D.; Metten, P.; Phillips, T.J.; Boehm, S.L.; 2nd, Burkhart-Kasch, S.; Dorow, J.; Doerksen, S.; Downing, C.; Fogarty, J.; Rodd-Henricks, K.; Hen, R.; McKinnon, C.S.; Merrill, C.M.; Nolte, C.; Schalomon, M.; Schlumbohm, J.P.; Sibert, J.R.; Wenger, C.D.; Dudek, B.C.; Crabbe, J.C. (2003) Different data from different labs: lessons from studies of gene-environment interaction. *J. Neurobiology* **54**, 283-311.
4. Wolfer, D.P.; Litvin, O.; Morf, S.; Nitsch, R.M.; Lipp, H.-P.; Würbel, H. (2004) Cage enrichment and mouse behaviour. *Nature*, **432**, 821-822.

Monitoring Animal Behavior in the Smart Vivarium

Serge Belongie, Kristin Branson, Piotr Dollár and Vincent Rabaud
Department of Computer Science & Engineering, University of California, San Diego, USA

Abstract

In the course of modern medical research, it is common for a research facility to house thousands of caged mice, rats, rabbits, and other mammals in rooms known as *vivaria*. In any experiment involving a group of animals it is necessary to perform environmental and physiological monitoring to determine the effects of the procedure and the health of the animals involved. Such monitoring is currently performed by human observers, and for practical reasons, only a small subset of cages can be inspected for limited amounts of time. This short paper outlines the computer vision and machine learning technology behind the *Smart Vivarium*, a system for automated, continuous animal behavior monitoring. The Smart Vivarium will serve as an invaluable tool for medical researchers as it will make better use of fewer animals. Early discovery of sick animals will prevent diseases from spreading, and in general will lead to more efficient caretaking of animals. Additionally, the proposed technology can serve as a powerful tool for monitoring sentinel cages in potential bioterrorism targets and chemical agent research facilities.

Keywords

mice, vivarium, tracking, behavior recognition

1 Introduction

A single vivarium can contain thousands of cages of mice, making close monitoring of individual mice impossible. Automated behavior analysis of individual mice will allow for earlier detection of abnormal behavior, and thus an improved level of animal care, as well as more detailed and exact data collection which will improve the efficiency of medical experiments.

Video surveillance of mice has the important characteristic of being non-intrusive; no modification to the environment is necessary. It is now feasible because of the recent availability of low-cost video cameras. Because of the huge number of medical experiments conducted on caged mice, this feasibility has led to a large amount of research on this problem; see for example the proceedings of *Measuring Behavior,* 1996 – present. To our knowledge, all current approaches (for example, [1]) require overhead mounted cameras. This simplifies the problem because the amount of occlusion is reduced. However, this kind of surveillance requires a specially designed cage, since in a standard mouse cage the overhead view is obstructed by the feeder and cage top (see Figure 1). In this short paper, we describe recent advances from our research group at the University of California, San Diego, on the problems of non-intrusive mouse tracking and behavior recognition from a side cage view.

Figure 1. *Still frame captured from a video sequence of three mice (240 by 360 pixels). The metal container at the top of the cage holds food pellets and a water bottle. It also prevents the use of an overhead mounted camera. The bedding on the floor of the cage is the only dynamic part of the background, other than reflections.*

2 Tracking Multiple Mouse Contours

Our tracking research focuses on the problem of tracking the contours of multiple identical mice from video of the side of their cage; see Figure 1 for an example frame. Although existing tracking algorithms may work well from an overhead view of the cage, the majority of vivaria are set up in a way that prohibits this view.

This problem is uniquely difficult from a computer vision standpoint. Because mice are highly deformable 3D objects with unconstrained motion, an accurate contour model is necessarily complex. Because mouse motion is erratic, the distribution of the current mouse positions given their past trajectories has high variance. The biggest challenge to tracking mice from a side view is that the mice occlude one another severely and often. Tracking the mice independently would inevitably result in two trackers following the same mouse. Instead, we need a multitarget algorithm that tracks the mice in concert. As the number of parameters that must be simultaneously estimated increases linearly with K, the number of mice, the search space size increases exponentially with K [4]. Thus, using existing approaches to directly search the contour space for all mice at once is prohibitively expensive.

In addition, tracking individual mouse identities is difficult because the mice are indistinguishable. We cannot rely on object-specific identity models (e.g., [3]) and must instead accurately track the mice *during* occlusions. This is challenging because mice have few if any trackable features, their behavior is erratic, and edges (particularly between two mice) are hard to detect. Other features of the mouse tracking problem that make it difficult are clutter (the cage bedding, scratches on the cage, and the mice's tails), inconsistent lighting throughout the cage, and moving reflections and shadows cast by the mice.

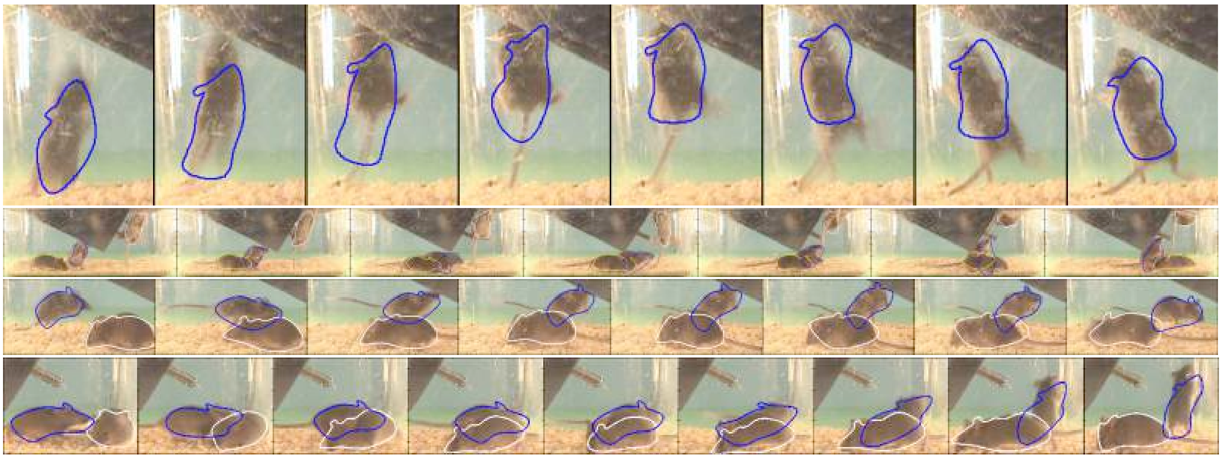


Figure 2. Still frame summary of the successes of our algorithm. We plot the average affine transformation applied to the contour with the most total weight.

We propose a solution that combines existing blob and contour tracking algorithms. However, just combining these algorithms in the obvious way does not effectively solve the difficulties discussed above. We propose a novel combination of these algorithms that accentuates the strengths of each individual algorithm. In addition, we capitalize on the independence assumptions of our model to perform most of the search independently for each mouse. This reduces the size and complexity of the search space exponentially, and allows our Monte Carlo sampling algorithm to search the complex state parameter space with a reasonable number of samples. Our algorithm works with a detailed representation of a mouse contour to achieve encouraging results.

We evaluated our blob and contour tracking algorithm on a video sequence of three identical mice exploring a cage, available at http://smartvivarium.calit2.net. This sequence contained 11 occlusions of varying difficulty. Summary still frames are shown in Figure 2. These results demonstrate the following strengths of our algorithm:

- Our contour tracking algorithm is robust to erratic mouse behavior – we never lose a mouse. For instance, we follow mice that jump, drop from the ceiling, and make quick turns and accelerations that are not fit by our simple dynamics model.
- Two contours never fit the same mouse.
- Our algorithm is rarely distracted by background clutter. This implies that our feature extraction methods and the blob and contour combination provide robust observation likelihoods. The only exceptions are when *both* algorithms make mistakes: when the blob tracker mistakes shaded bedding for foreground and the contour tracker fits to the edge of a tail.
- Perhaps the most impressive result is that our algorithm accurately tracks the mice through 7 out of 11 occlusions and partway through the other 4. This is because of the detailed fit provided by the contour tracking algorithm and its ability to use features available during occlusions.
- In general, our algorithm usually found very good contour fits outside of occlusions, much better than those obtained using contour tracking alone.

More information about our algorithm can be found in [2].

3 Behavior Recognition via Sparse Spatio-Temporal Features

After tracking and thus determining the positions and identities of each mouse, we focus our attention on recognizing their behavior. The method we have developed so far is successful in determining five basic behaviors: sleeping, drinking, exploring, grooming, and eating. Its design is very general and easily portable to other activities (like scratching or nesting that were omitted simply for lack of training footage).

Most behavior recognition methods developed in the computer vision literature are focused on human activity recognition (for a survey see [6]), but these methods are generally inapplicable to rodent behavior for the following reasons:

- With the exception of the eyes and ears, there are very few distinguishable features on the body of a rodent
- Rodent limbs are almost imperceptible in a single frame (except perhaps for the tail)
- Relevant activities can happen in a burst (e.g. less than a second)

Furthermore, many of the traditional approaches for human activity recognition assume simplifications that we cannot make, including:

- Simple backgrounds and little or no occlusion
- Small or no variation in the behaviors and posture of the subjects
- High resolution, clean data

Many of the problems described above have counterparts in object recognition. The inspiration for our approach comes from approaches to object recognition that rely on sparsely detected features in a particular arrangement to characterize an object, e.g. [8] and [7]. Such approaches tend to be robust to image clutter, occlusion, object variation, and the imprecise nature of the feature detectors. In short they can provide a robust descriptor for objects without relying on too many assumptions.

Our approach is based on describing a behavior in terms of local regions of interesting motion. Figure 3 shows example frames of two clips of a mouse grooming where the global appearance and motion are quite different but local regions of motion are quite similar.

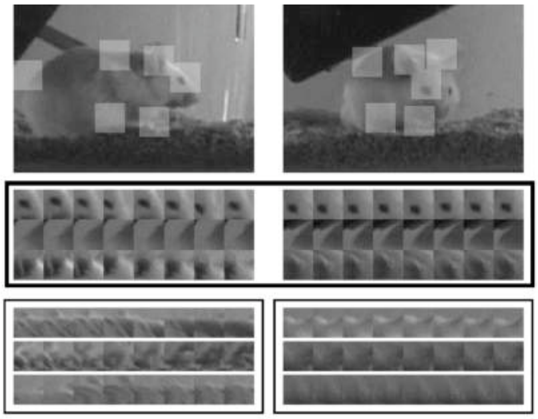


Figure 3. Highlighted regions of local motion in example frames of two from videos of mice grooming. Six prominent areas of motion were extracted from each behavior. Note that although the posture of the mouse is quite different in the two cases, three of the six regions (shown in the top three rows) for each mouse are quite similar. The other three have no obvious correspondence although it is very hard to perceive what these are without motion.

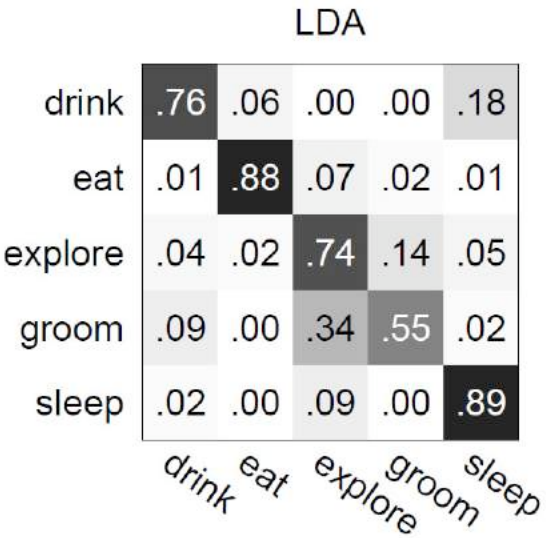


***Figure 4.** The confusion matrix on test data shows the efficiency of our approach. Each row represents how our algorithm classified a given activity (using Linear Discriminant Analysis). The most confusion occurs when a mouse is grooming and the algorithm incorrectly classifies it as exploring.*

Although our method is still in development, it is already very reliable (cf. figure 4). These results were obtained with only a small amount of training data. As we increase the amount of training data, and incorporate a more robust model for activity, the accuracy and number of activities we can detect shall continue to increase.

4 Conclusion

We have described our recent progress in two areas of the *Smart Vivarium* project at UC San Diego: multiple mouse tracking and behavior recognition in the home cage environment. In our ongoing work we are collaborating with medical researchers to apply this technology for purposes of behavioral phenotyping in inbred mouse strains.

This work is funded by NASA GSRP grant NGT5-50479, the UCSD Division of Calit2 – the California Institute for Telecommunications and Information Technology – under the Smart Vivarium project, NSF Career grant #0448615, subcontracts B542001 and B547328 under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48, and the Alfred P. Sloan Fellowship.

References:

1. Twining, C; Taylor, C.; Courtney, P. (2001). Robust tracking and posture description for laboratory rodents using active shape models. In *Behavior Research Methods, Instruments and Computers, Measuring Behavior Special Issue*.
2. Branson, K.; Belongie, S. (2005). Tracking multiple mouse contours (without too many samples). *To appear, IEEE International Conference on Computer Vision and Pattern Recognition*, San Diego, CA.
3. Comaniciu, D.; Ramesh, V.; Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564-575.
4. MacCormick, J.; Blake, A (2002). A probabilistic exclusion principle for tracking multiple objects. *International Journal of Computer Vision*, 39(1):57-71.
5. Dollár, P.; Rabaud, V.; Cottrell, G.; S. Belongie. (2005, *submitted*). Behavior recognition via sparse spatio-temporal features. *IEEE International Conference on Computer Vision*, Beijing, China.
6. Gavrila, D. M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98.
7. Agarwal, S.; Awan, A.; Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1475– 1490.
8. Fergus, R.; Perona, P.; Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. *IEEE International Conference on Computer Vision and Pattern Recognition*, Madison, WI.

Multi-level analysis of mouse behavior in a home cage environment using PhenoLab®

L. de Visser[1,2], R. van den Bos[1,2], W.W. Kuurman[1], B.M. Spruijt[1,2]
[1] *Animals, Science and Society, Utrecht University, Utrecht, The Netherlands*
[2] *Rudolf Magnus Institute of Neurosciences, UMC, Utrecht, The Netherlands*

Abstract

To contribute to the refinement of behavioral phenotyping methods for inbred and mutant mice, we developed a automated tool for observing and analyzing behavior in a home cage-like environment (PhenoLab®, Noldus Information Technology, Wageningen, The Netherlands). Testing animals in their home cage environment holds several advantages; it allows continuous observations over consecutive days and the evaluation of both challenge-induced and baseline behaviors. Home cage testing also minimizes human intervention (such as handling) and reduces interactions with other environmental factors not related to the behavioral test (such as animal transport). Here, data will be presented of studies in inbred strains of mice (C57BL/6 and DBA/2) on locomotor activity and anxiety-related behavior in the PhenoLab® system. We found that different aspects of activity behave differently over time. Anxiety-related behavior was studied by introducing an aversive light stimulus in the cage after an adaptation period of 4 days. The light stimulus illuminated the area around the feeding station, inducing pronounced avoidance behavior in the animals. Possibilities and limitations of PhenoLab® as a tool for the behavioral phenotyping of mice are discussed.

Keywords

Automated home cage observations, mice, locomotor activity, anxiety.

1 Introduction

The continuous increase in mouse models for central nervous system functioning demand extensive behavioral assays to elucidate the impact of genetic alterations [2,4,7]. Current methods for behavioral phenotyping of mice often involve batteries of individual tests, each addressing different motivational systems, such as activity in novel environment, anxiety, spatial cognition, etc [1,3]. Although most tests are pharmacologically validated and relatively easy to perform, some limitations are evident. First, long term development of behavior as a response to stimuli (e.g. novelty or drug effects) and circadian processes are ignored due to the limited testing period. Second, information on the complex interactions between motivational systems is often difficult to acquire.

We developed a new tool to study mouse behavior in a home cage like environment [5]. This approach has several advantages; it allows continuous observations over days and the evaluation of both challenge-induced and baseline behavior. Furthermore, the confounding influence of stress caused by handling and transport is minimized.

In this paper we present data from two separate experiments to demonstrate some of the possibilities of automated home cage observations. The first experiment addresses time-courses of specific aspects of locomotor activity as mice gradually become habituated to the novel home cage system. In the second experiment, an aversive stimulus is used to create an approach-avoidance conflict, which allows for the detection of parameters indicative of anxiety.

2 Methods

2.1 Animals

Female mice of the C57BL/6OlaHsd (n=6) and DBA/2OlaHsd (n=10) strain were used in the first experiment. Male mice of the C57BL/6OlaHsd strain (n=10) were used in the second experiment. All mice were purchased from Harlan (Horst, The Netherlands). Upon arrival at the animal facility, mice were either housed in pairs (females) or single (males) and allowed to acclimatize for two weeks under a reversed light/dark cycle (lights on: 19.00 hrs). All mice were provided with a shelter, tissue and paper shreds as enrichment. Humidity was kept at a constant level and room temperature was maintained at 21.0 ± 2.0 °C. The Animal Ethical Committee of Utrecht University approved the experiments.

2.2 Automated home cage observations

Locomotor activity was automatically recorded with video tracking in specially designed home cages (PhenoLab®, Noldus Information Technology, Wageningen, The Netherlands, see Figure 1). Each PhenoLab® system consists of four cages, connected to a PC that runs EthoVision 3.0 (Noldus Information Technology) for videotracking. On top of each cage a unit is placed containing a digital infrared sensitive camera and infrared lights. This allows continuous recordings during both light and dark period of the day. The top unit further contains a bright white light stimulus that could be switched on automatically by programming EthoVision 3.0. This lightspot illuminates approximately one quarter of the cage with a light intensity of 1000 lux.

Figure 1. *Single cage with top unit of the PhenoLab® system for automated home cage observations. Each cage contains a fixed water bottle, feeding station and shelter.*

\

2.3 Experiment 1

For the first experiment, mice were introduced to the PhenoLab® cages and locomotor acitivity was recorded for six consecutive days. Cages were equipped with a fixed water bottle and feeding station, and shelter, tissue and paper shreds for enrichment. For details on parameter settings in EthoVision 3.0, see [6]). Parameters presented in the present paper are 'duration of movement' as a percentage of total observation time and 'velocity' in cm/s. Values for each parameters are calculated for each individual in 1-hour bins and subsequently averaged over 12 hours to differentiate between dark and light period of the day.

2.4 Experiment 2

In the second experiment, mice were introduced to the PhenoLab® cages following the procedure of experiment 1, but now the bright white light stimulus was programmed to switch on, on day 5 immediately at the onset of the dark period. The light spot was directed on the feeding station for three hours continuously. This created an approach-avoidance conflict for the animals. Again, locomotor activity was recorded for six consecutive days in total. Parameters analyzed for this paper were 'duration of movement' in seconds per hour and the 'time spent in feeding zone'. The feeding zone was defined as one body length around the feeding station. During the time the light spot was on, this zone was illuminated.

2.5 Statistical analysis

Statistical analysis was conducted using SPSS 10.0 for Windows. Repeated measures ANOVA were performed to test for overall effects of within-subjects factor 'day' and between-subjects factor 'strain' on duration of movement and velocity in experiment 1. *Posthoc* independent samples t-tests were used to compare strains on each experimental day. Levels of significance were assigned at p=0.05.

3 Results

3.1 Experiment 1

Results from experiment 1 are modified from the study described in [5]. Figure 2 presents the 'duration of movement' during the dark period of each experimental day. Overall, C57BL/6 mice showed higher levels of movement (repeated measures ANOVA between-subjects factor 'strain' $F_{1,14}$=6,663; p=0.022) compared to DBA/2 mice. Posthoc comparisons per day revealed significant higher duration of movement in C57BL/6 on day 1, 5 and 6. Duration of movement changed over time (repeated measures ANOVA within-subjects factor 'day' $F_{5,70}$=41,559; p<0.001) but this was independent of strain (repeated measures ANOVA 'day' x 'strain': $F_{5,70}$=1,227; p=0.310). In Figure 3 strain differences on the parameter 'velocity' are shown. Overall, DBA/2 mice moved with higher velocity compared to C57BL/6 (repeated measures ANOVA between-subjects factor 'strain': $F_{1,14}$=7,060; p=0.019). However, when tested per experimental day differences were only significant on day 5 and 6. Velocity changed significantly over time (repeated measures ANOVA within-subjects factor 'day': $F_{5,70}$=5,041; p=0.008) independent of strain (repeated measures ANOVA 'day' x 'strain': $F_{5,70}$=2,339; p=0.100).

3.2 Experiment 2

The mice showed a strong cyclic activity pattern with high levels of movement duration during the dark phase as compared to the light phase.

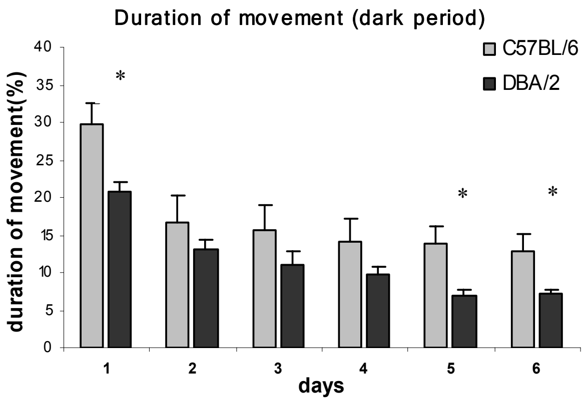


*Figure 2. Duration of movement (as a percentage of time) during the dark period for each experimental day (1-6). Means + SEM per12- hour period are used. * p<0.05 for differences between strains; independent sample t-test.*

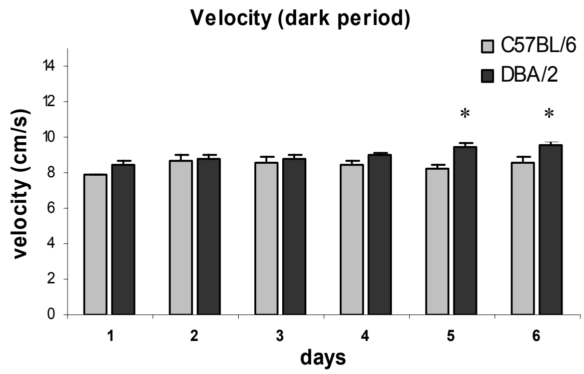


*Figure 3. Velocity (in cm/s) during the dark period for each experimental day (1-6). Means + SEM per12- hour period are used. * p<0.05 for differences between strains; independent sample t-test.*

The light spot induced a small decrease in overall activity as reflected by lower duration of movement (Figure 5). Furthermore, there was a shift in activity towards the end of the dark phase on day 5, compared to day 4. 24 hours after the light spot (day 6), duration of movement was still lower than what would have been expected from day 4.
The light spot induced pronounced avoidance behavior in male C57BL/6 mice (see Figure 4). Time spent in the feeding zone decreased markedly during the time the light spot was switched on (day 5) compared to the day before (day 4). Some degree of habituation to the stimulus occurred, which was reflected by the increase in time spent in the feeding zone within the three hours the light spot was on. Notably, after the spot was switched off, there was a marked increase in the time spent in the feeding zone during the fourth hour of the dark phase. Values exceeded the duration in feeding zone during the fourth hour of the dark phase of day 4 and day 6.
24 hours after the light spot (day 6), time spent in feeding zone was still decreased compared to day 4. However, during the first three hours of day 6, time in the feeding zone increased faster compared to this time period on day 5.

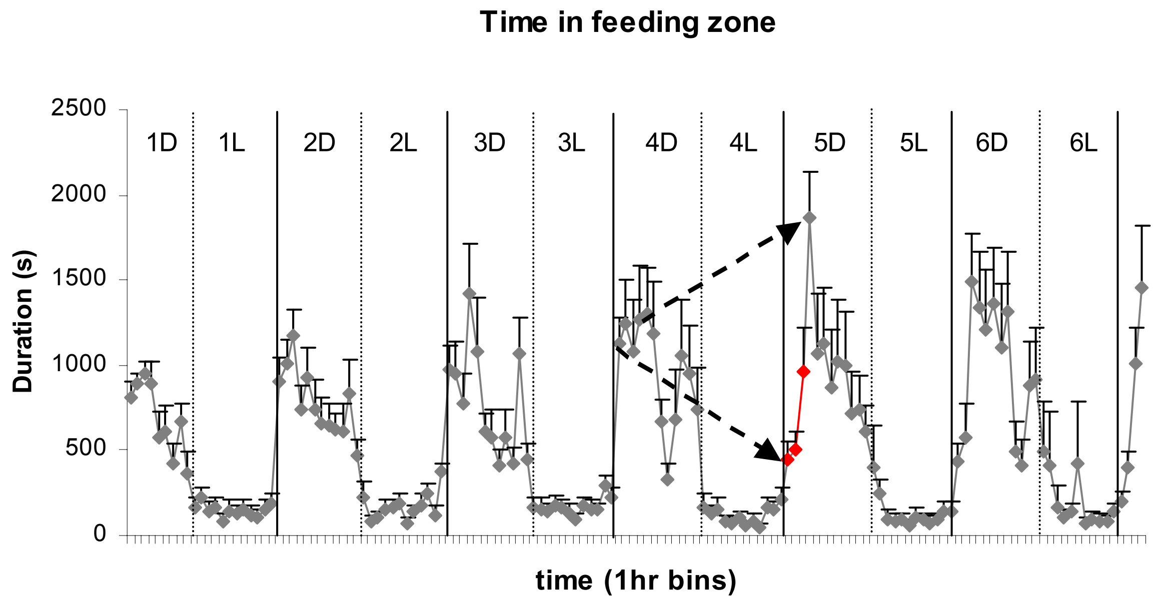


Figure 4. Time spent in feeding zone during six consecutive days. Means and SEM of 1hr bins are used. Days are indicated by numbers (1-6) and dark and light periods are distinguished by either D (dark period) or L (light period). Black data points on day 5 represent time window when light spot was switched on. Arrows indicate the main findings, see text for further explanation.

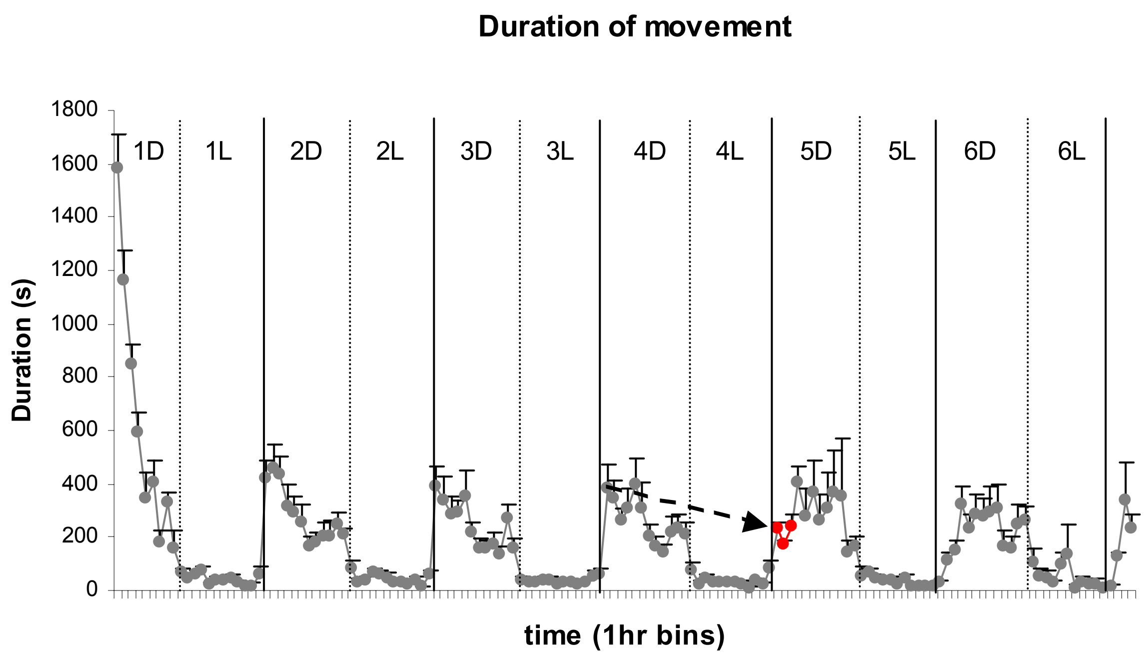


Figure 5. Duration of movement during six consecutive days. Means and SEM of 1hr bins are used. Days are indicated by numbers (1-6) and dark and light periods are distinguished by either D (dark period) or L (light period). Black data points on day 5 represent time window when light spot was switched on. Arrow indicates the main finding, see text for further explanation.

4 Discussion

Automated home cage observations offer several possibilities for the behavioral characterization of mice. Novelty induced exploration can be distinguished from activity in a familiar environment using a single test. It appeared that some aspects of locomotor activity, such as duration of movement, are more dependent on familiarity with the environment than others, such as velocity. Furthermore, by presenting an aversive stimulus when the animals were habituated to the environment, we were able to create an approach-avoidance conflict. This allowed evaluation of possible anxiety-related behavior, while avoiding the confounding influences of novelty, handling and transport. Without these external stress factors it is possible to distinguish "state" from "trait" characteristics. Currently, experiments are being performed with the anxiolytic drug diazepam to find further evidence for the use of the aversive light spot as an anxiety test.

However, limitations of the presented method are in the time- consuming process of data handling and analysis. The vast amounts of data generated by the system demand sophisticated data handling software and statistical tests. Moreover, a functional interpretation of this computer output in terms of ethologically relevant profiles is a challenge in itself. To facilitate this, statistical tools for data reduction and clustering, such as Principal

Components Analysis, are adopted [6]. With these tools, the interrelation of parameters can be studied to reveal underlying motivational systems. Pharmacological experiments can further prove the predictive validity of parameters that showed differential potential between e.g. inbred mice.

References:

1. Crawley, J.N.; Belknap, J.K.; Collins, A.; Crabbe, J.C.; Frankel, W.; Henderson, N.; Hitzemann, R.J.; Maxson, S.C.; Miner, L.L.; Silva, A.J.; Wehner, J.M.; Wynshaw-Boris, A. & Paylor, R. (1997) Behavioral phenotypes of inbred mouse strains: Implications and recommendations for molecular studies. *Psychopharmacology*, **132,** 107-124.
2. Gerlai, R. (2002) Phenomics: fiction or the future? *Trends Neurosci*, **25**, 506-509.
3. Rogers, D.C.; Jones, D.N.C.; Nelson, P.R.; Jones, C.M.; Quilter, C.A.; Robinson, T.L.; Hagan, J.J. (1999) Use of shirpa and discriminant analysis to characterize marked differences in the behavioural phenotype of six inbred mouse strains. *Behav Brain Res,* **105**, 207-217.
4. Tecott, L.H.; Nestler, E.J. (2004) Neurobehavioral assessment in the information age. *Nat Neurosci*, **7**, 462-466.
5. de Visser, L.; van den Bos, R.; Spruijt, B.M. (2005) Automated home cage observations as a tool to measure the effects of wheel running on cage floor locomotion. *Behav Brain Res*, **160**, 382-388.
6. de Visser, L.; van den Bos, R.; Kuurman, W.W.; Kas, M.J.H.; Spruijt, B.M. (in press) Novel approach to the behavioural characterization of inbred mice: automated home cage observations. *Genes Brain Behav.*
7. Wahlsten, D.; Rustay, N.R.; Metten, P.; Crabbe, J.C. (2003) In search of a better mouse test. *Trends Neurosci,* **26**, 132-136.

Behavioral phenotyping: getting more information from video tracking data

W. W. Kuurman and L. de Visser
Department of Animals, Science & Society, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

Abstract
Discovery of mice with a specific behavioral phenotype is of considerable interest for neuroscience research and drug development. A method for high throughput phenotyping (HTP), therefore, is being developed. In HTP single mice are subjected to behavioral tests for one week in a home cage-like environment and tracked using an overhead camera and video tracking software. One of the main challenges of this approach is handling and analysis of the large amount of data generated in the HTP. The goal of this paper, therefore, is to present an outline for handling and analysis of the data to eventually define ethologically relevant characteristics of the mice. After data acquisition, noise in the raw data is reduced by several procedures. After this, data is analyzed by computing parameters traditionally used in behavioral research and from mathematical models that are currently developed. The total set of parameters computed for each mouse forms a high resolution behavioral phenotypic map. For evaluation purposes this set of parameters is reduced to a small set of factors using principal components analysis. Standardized deviations for each factor and each mouse are summarized in a phenogram. This phenogram is used by an expert committee to select mice with an interesting behavioral phenotype. After selection, mice are subjected to low throughput phenotyping for detailed neurobiological research.

Keywords
Behavior, Phenotyping, Mathematical Modeling.

1 Introduction
Mice are used as models in a large variety of biological studies. Discovery of mice with a specific behavioral phenotype is of considerable interest for neuroscience research and drug development. To find interesting phenotypes Neuro-Bsik Mouse Phenomics Consortium was assembled, consisting of ten university neuroscience research groups and two commercial companies.
For this consortium the behavioral phenotype for large numbers of mutant and recombinant mice will be determined in a standardized manner using high throughput phenotyping (HTP). In the HTP single mice are tested for several behavioral domains; locomotion, activity, habituation, perception, cognition and anxiety. An expert committee evaluates the test results so as to select mice with an interesting behavioral phenotype. These selected mice are subjected to low throughput phenotyping (LTP). In the LTP mice can be studied in greater detail at various levels.
For the HTP a PhenoLab®, consisting of several PhenoTyper® home cages (Figure 1), was developed (Noldus Information Technology, Wageningen, The Netherlands). In these cages single mice are tested, without human intervention, for a week. Location of the mouse is determined using an overhead camera and EthoVision video tracking software (Noldus information technology).
Various test paradigms are being developed for these PhenoTyper® cages and validated against traditional behavioral tests [3] such as the open field and elevated plus maze [2]. Due to the large amount of data obtained in the HTP (6.5 million lines of data for each mouse), development of methods, algorithms, mathematical models and statistics to evaluate these tests and present results such that they can be interpreted by the expert committee is an important part of HTP design. The goal of this paper, therefore, is to present an outline for handling and analysis of the data to eventually define ethologically relevant characteristics of the mice, including some examples for clarification. This outline consists of four stages; data acquisition, noise reduction, analysis of data, and presentation of results.

Figure 1. *The PhenoTyper® home cage is used for high throughput phenotyping. In a battery of these cages single mice are housed for a week, subjected to a large number of behavioral tests and data is acquired using an overhead camera and video tracking software.*

2 Data Acquisition
For the HTP single mice are housed and tested, without human intervention, in the PhenoTyper® for one week. In the cage the mouse has free access to feed and water, wood chips are used for bedding and a shelter and nesting material are used as enrichment. During testing, location of the mouse is determined using an overhead camera and video tracking software.

Based on contrast differences, this software records size of the detected area (the mouse) and X- and Y-coordinates of the center of gravity for the mouse at 0.08 second intervals. Additionally, the zone; in the shelter, on the shelter, in open space, vicinity of feeder etc., in which the mouse is present is recorded. After one week of data acquisition 6.5 million samples containing the location of the mouse are recorded.

During the HTP-week both spontaneous and stimuli-evoked behaviors are observed. Examples are response to novelty, circadian rhythmicity, emotional reactivity to both appetitive and aversive stimuli, adaptation to the environment over days and cognitive behavior. The behavioral phenotype of the mouse needs to be deduced from the recorded information, i.e., integration of location and movement.

3 Noise Reduction

Raw data obtained during the HTP-week has been subject to various types of measurement error. A correction of some of these errors can help to improve data quality for further analysis. Several algorithms are being developed and used for this purpose. Some of the decision rules in the algorithms might be considered an educated guess still subject to change, but are given for illustrative purposes.

3.1 Missing Samples

During tracking samples are occasionally recorded as missing. This might be due to the small area size of the detected mouse because it burrowed under wood chips and tissue or temporary high processor loads of the computer. If data is missing and, based on known X- and Y-coordinates before and after the missing samples, the mouse has been missed for less than 2 sec. or has moved less than 2 cm in both the X- or Y-direction than data is interpolated.

3.2 Zone Transistions

It occurs often that the mouse (its center of gravity) enters a zone during a few samples before returning to the previous zone. This adds to the number of zone transitions while the mouse might not actually be in that zone. If the mouse enters a zone less than 1/3 sec. before returning to the previous zone, than this is not considered a zone transition.

3.3 Sleeping Outside Shelter

Although most mice use the shelter to sleep or rest, some mice do not. If the mouse has moved less than 2 cm in both the X- or Y-direction for more than 5 min. in the open space than the mouse is considered to have slept or rested outside the shelter.

3.4 Path Shape Smoothing

Recording of the X- and Y-coordinates is subject to variations in area size, body wobble, errors in detecting the mouse. The X- and Y-coordinates form the bases for many locomotive parameters, so that and accurate estimation of these points is important. Assuming the mouse follows a smooth curved path without large sudden deviations, path shape smoothing repositions the X- and Y-coordinates using the LOWESS algorithm [6]. After path shape smoothing locomotive parameters such as velocity, distance moved, turn angles, etc. can be computed.

4 Analysis of Data

After noise reduction the data can be used for analysis to determine the behavioral phenotype of each mouse. The goal of the analysis is to summarize the data into a set of parameters. This set should such that each behavioral test result is expressed in one or more parameters and that as much of the variation present in all mice tested is captured. This approach results in a high-resolution phenotypic map, increasing the chance of detecting mice with an interesting behavioral phenotype.

Video tracking data is at present used to calculate frequencies, latencies, durations, totals, and means, which are summarized in time bins. These parameters are then used for further statistical analysis.

In addition to these traditional measures mathematical models are now being developed from which the parameter estimates can be added to the total set of parameters. Mathematical models can capture variation in the data not adequately expressed by traditional parameters, or can capture changes over time during the HTP-week in a single parameter, i.e. slope or rate. For data that does not have a known unimodal distribution, for example, mean and standard deviation ignore variation in the data, which can be captured using a mathematical model. To illustrate this, two examples of mathematical models are given; 1) distance to arena wall, and 2) gearbox. For both mixture models are used [7]. Mixture modeling is suited for analysis of variables with a multimodal distribution. In such a model each mode is modeled by a unimodal distribution (often the normal distribution) and the proportionality between these modes is estimated.

4.1 Distance to Arena Wall

Mice tend to move close to the wall, and occasionally move into the open space in the middle of their cage. Distance to arena wall (DAW) might be indicative of fear/anxiety. The distribution for DAW is bimodal, and is modeled using two unimodal distributions and a proportionality parameter

$$DAW \sim p\, D_1(\mu_1, \sigma_1) + (1-p)\, D_2(\mu_2, \sigma_2)$$

where ~ denotes “distributed as,” p is the proportionality parameter, D_i is distribution i (i = 1 or 2) with mean μ_i and standard deviation σ_i. Distribution 1 models the proportion of the data when the mouse is close to the arena wall, and Distribution 2 models this when the mouse is not close to the wall. Assuming this distribution can change over time each of the five parameters might change over time. Each parameter, therefore, can be substituted for a linear model. For the proportionality parameter this linear model would be

$$p = p_0 + p_1 \times \text{time}$$

where p_0 is the intercept (p at time 0) and p_1 is the slope or the amount of change for p per unit of time. Similarly this can be done for the remaining four parameters in the model, resulting in a mixture model containing ten parameters.

Results are depicted for two mice in an open field (80 cm diameter) at three different times in Figure 2. Figure 2 depicts differences between the two mice, but these differences are also expressed as model parameter estimates. The transition for being close to the arena wall and away from the wall is about 7 cm for Mouse 1 and about 8 cm for Mouse 2. As time increases Mouse 1 spent more time away from the wall, while Mouse 2 spent more time close to the wall. When each mouse was away from the wall, Mouse 1 staid closer to the wall than Mouse 2. These differences between the two mice would have been less predominant using a traditional analysis; calculating the proportion of time each mouse has spent closer than 10 cm (or any other distance) from the wall.

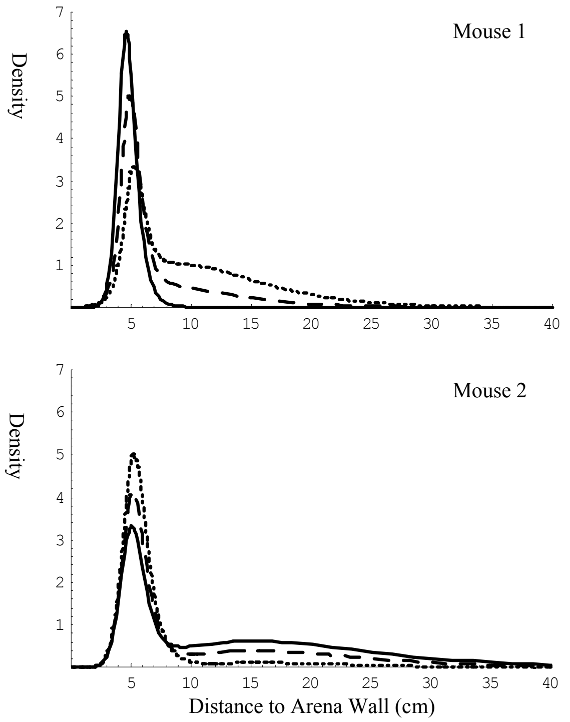


Figure 2. *Predicted probability density for distance to arena wall in an open field of two mice at time t = 0 (——), t = 7.5(– – –) and t = 15 (- - -) minutes.*

4.2 Gearbox

Mice move at various velocities and these velocities might be grouped into various modes of movement [4,5] or "gears." Similar to a model developed by Drai *et al.* (2000) an alternative model was developed to distinguish various gears for movement. In contrast to the model by Drai *et al.* (2000), which models maximum velocities during bouts of movement using Gaussian distributions, in the alternative model all velocities at which the mouse moves and different distributions are used to estimate the different gears for each mouse. For this alternative model a mixture of three distributions is used for velocities at which the mouse was moving, i.e. velocities when the mouse was lingering or stopped were excluded.

$$Vel \sim p_1\, G_1(\mu_1, \sigma_1) + p_2\, G_2(\mu_2, \sigma_2) + (1- p_1 - p_2)\, G_3(\mu_3, \sigma_3)$$

where p_i is the proportion of time the mouse moves using Gear i (i = 1, 2, 3) [Note $p_3 = 1- p_1 - p_2$, so that $1 = p_1 + p_2 + p_3$], G_i is the distribution with mean μ_i and standard deviation σ_i for velocities in Gear i. The three gears describe low, intermediate and high velocities.

A frequency table and predicted probability density functions for each of the three gears for one mouse are depicted in Figure 3.

The gearbox model provides for each mouse mean and standard deviation for each of three gears and the proportional use for each of those gears. In addition model parameters can be used to calculate the proportional use for each of the three gears for any time interval, moving bout, or behavioral test of interest.

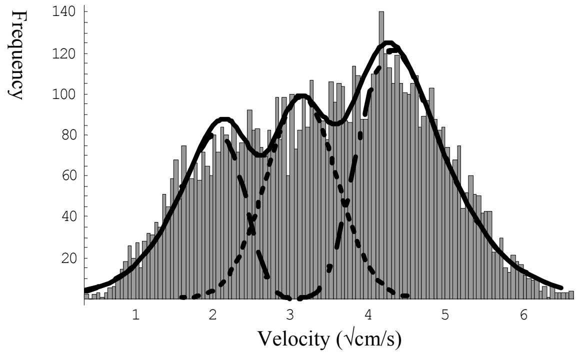


Figure 3*. Frequency table and imposed probability density for velocity of a mouse moving at low(—— ——) , intermediate (– – –) and high (—– – —) velocity, in addition to imposed density of the mixture distribution of the three densities (——).*

4.3 Further Analysis

More mathematical models are being developed, e.g. for habituation [3], in addition to models for pathway analysis and models for specific behavioral tests. After one week of testing in the HTP-week each mouse could have a set of 150 to 250 parameter estimates.

5 Presentation of Results

The set parameters derived for each mouse from the HTP-week contains too many parameter estimates for an expert committee to effectively select mice with an interesting behavioral phenotype. This set, therefore, needs to be summarized into a short list of endpoints. This list of endpoints should contain as much information as possible. Principal Component Analysis (PCA) is an algorithm that reduces a set of parameters into a list of consecutive factors. Each factor is a linear combination of the parameters such that the largest amount of variation in the data is explained by the first factor, and the second largest amount of variation by the second factor, etcetera. The factors themselves are uncorrelated. The first few factors usually explain most of the variation in the data [1]. For each factor mouse scores can be calculated, standardized (mean is zero and standard deviation is one), and represented in a phenogram. Such a phenogram can be used to detect mice visually that deviate for one or more factors compared to other mice tested in the HTP-week (Figure 4).

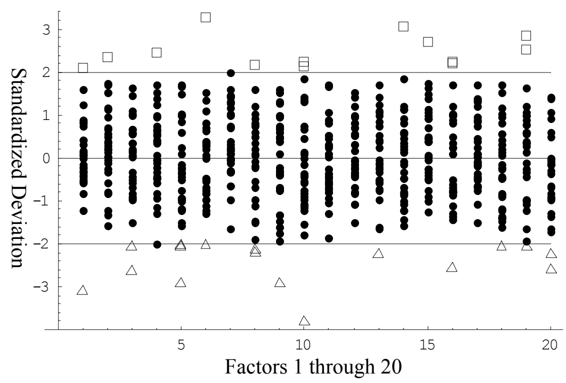


Figure 4*. Phenogram of standardized deviation scores for 30 mice for each of 20 factors. Mice can either deviate positively (□) or negatively (Δ) compared to mice that do not deviate (●).*

For each mouse all parameter estimates are retained so that results for mice that appear interesting based on the phenogram can be reviewed in detail.
Further research is needed to examine if specific factors in the phenogram can be associated with behavioral domains. Mice can so be selected for these domains instead of the abstract concept of factors.

6 Discussion and Conclusion

High throughput phenotyping of mice using video tracking data requires the analysis of large amounts of data. The challenge is to reduce the data to a limited set of parameters that contains most of the variation contained in the data. In addition to the traditional parameters obtained from video tracking data, mathematical models are being developed to achieve this goal. The advantage of mathematical models is that these can model variables distributed in a great variety of shapes. Furthermore, effects of time and can be accounted for as well as responses to stimuli.
Two examples of mathematical models are given. First, a mixture model for distance to arena wall consisting of a mixture of two distributions. Parameter estimates from this model can be used to detect how often the mouse is close to the wall, where the transition is between being close to the wall and being away from the wall, and how these measures change over time. Second, a mixture model for velocity containing a mixture of three distributions, representing a gearbox with three gears. Parameter estimates from this model can be used to detect the characteristics of each gear for each mouse and to determine which gears are used at specific moments or moving intervals.
With more mathematical models being developed, the amount of information obtained from the data increases. Each traditional and mathematical model parameter contains specific information for each HTP tested mouse. The entire set of parameters forms a high-resolution map of the behavioral phenotype for each mouse.
To facilitate interpretation of the data the set of parameters is reduced, using PCA, to a small number of factors containing most information represented in the set of parameters. For each mouse tested and for each factor a standardized deviation score can be computed. These scores are visualized in a phenogram, which can be used to make a first selection of mice showing an interesting behavioral profile by an expert committee.
Mice selected from the HTP are subjected to LTP. Data obtained from the LTP can in addition to being used for LTP purposes also be used to evaluate, validate, improve, and interpret the mathematical models, parameters, and PCA analysis for the HTP data.
Although methodology and models are being developed based on data from the HTP, they might be applied to other behavioral tests using video tracking data. This can result in a more efficient use of data, hence reducing resources such as animals, labor and facilities, needed to obtain these data. In addition these models might provide insight into behavioral phenotype not disclosed using traditional parameters.

References:

1. Armitage, P.; Berry, G.; and Matthews, J. N. S. (2002). *Statistical Methods in Medical Research*. Blackwell Publishing Company, Oxford, UK.
2. Crawley, J. N (2000). *What's Wrong With My Mouse? Behavioral Phenotyping of Transgenic and Knockout Mice*. John Wiley & Sons, New York, USA.
3. De Visser, L.; Van den Bos, R.; Kuurman, W. W.; Kas M. J. H.; and Spruijt, B. M. (2005). Novel approach to the behavioral characterization of inbred mice: automated home cage observations. Submitted.
4. Drai, D.; Benjamini, Y.; and Golani, I. (2000). Statistical discrimination of natural modes of motion in rat exploratory behavior. *Journal of Neuroscience Methods*, **96,** 119-131.
5. Drai, D.; Kafkafi, N.; Benjamini, Y.; Elmer, G. I.; and Golani, I. (2001). Rats and mice share common ethologically relevant parameters of exploratory behavior. *Behavioral Brain Research*, **125,** 133-140.
6. Hen, I.; Kafkafi, N.; Golani, I.; and Benjamini, Y. (2004). The dynamics of spatial behavior: how can robust smoothing techniques help? *Journal of Neuroscience Methods*, **133,** 161-172.
7. Mclachlan, G.; and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons, New York, USA.

The Erasmus Ladder: a new automated screening tool for motor performance and motor learning deficits in mice

Cupido, A.[1], Krygsman, S.P.[1], Snoeck, C.M.[1], Bos, J.A. Th.[2], De Zeeuw, C.I.[1] and Koekkoek, S.K.E.[1]

[1]*Neuroscience, Erasmus MC, Rotterdam, The Netherlands*
[2]*Experimental Medical Instrumentation, Erasmus MC, Rotterdam, The Netherlands*

Abstract
To quickly screen a large amount of mutant mice for deficits in motor learning and motor performance, we created a new tool, the Erasmus Ladder. The Erasmus Ladder analyses the walking pattern of mice, and screens for cerebellar deficits. With a few adjustments it is also possible to screen for hippocampal deficits.

Keywords
Mouse models, motor performance, motor learning, screening tool, cerebellum.

Introduction
Recent developments in the creation of mouse models for human diseases have called for tools that are able to quickly screen mutant mice for their deficits. In our department, we have created two successful tools for detecting mouse mutants with cerebellar deficits: eye blink conditioning [3] and vestibulo-ocular reflex adaptation [2]. However, for screening large amounts of mice these systems are too time consuming, as they require specialized and invasive surgery. Therefore we developed a new tool: the Erasmus Ladder.

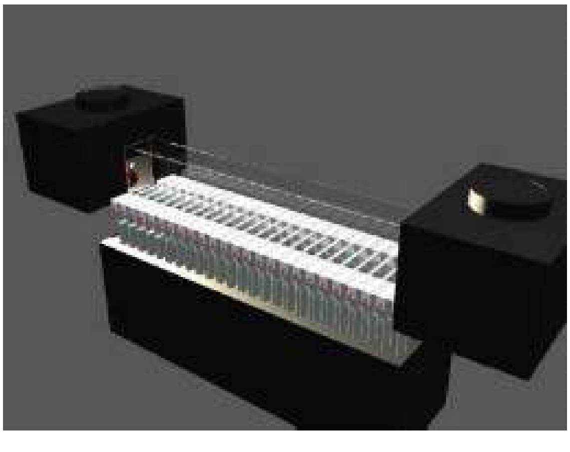


Figure 1. The Erasmus Ladder is a horizontal ladder with two boxes. The horizontal ladder has 25 rungs. The rungs are divided in a left side and a right side. All subdivisions of the rungs are equipped with pressure sensors. The rungs can be automatically protracted and retracted. The Erasmus Ladder is designed in such a way that in every part the airflow can be regulated.

Walking Pattern
The Erasmus Ladder (figure 1) is designed to analyze the walking pattern of mice. Mice normally explore their surroundings, which complicates their walking pattern. Therefore we designed several protocols to force the mouse to walk with a constant speed. All these protocols are based on regulating the airflow, which will stimulate the mouse to walk in a certain direction. Since increasing the airflow stresses the mouse, it is possible to condition the mouse after a couple of training sessions to walk with a constant speed.

Screening for Cerebellar Deficits
To screen for cerebellar deficits, mice will be conditioned in the following way: with our software it is possible to retract a rung on which the mouse is about to place its paw. Just before this rung will be retracted a tone will be given. If the mouse does not learn to adjust its walking pattern when the tone is given, there is high probability of a cerebellar deficit [1].

Advantages
The advantages of the Erasmus Ladder are fourfold. First, the experiments do not require surgery; second, the Erasmus Ladder will be completely computer controlled, thus the experiments are not time consuming; third, because the software of the Erasmus Ladder is a custom written modular LabView 7.1 real-time application [4], the configuration of the Erasmus Ladder can immediately adjust to the behavior of the mouse; fourth, with a few adjustments it is also possible to screen for hippocampal deficits, since six Erasmus ladders can form a maze-like configuration (figure 2).

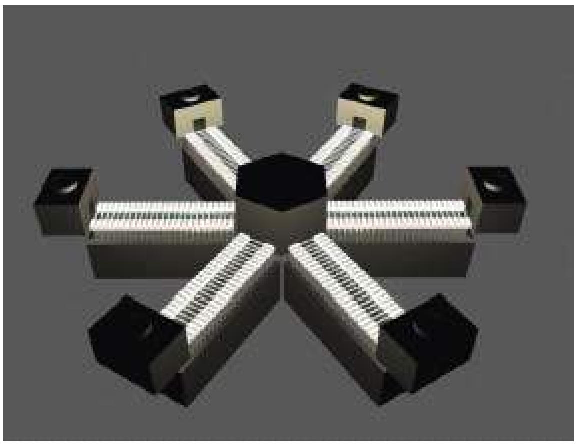


Figure 2. When six Erasmus Ladders are connected, it is possible to condition a mouse to prefer a specific box.

References

1. Apps, R.; Garwicz, M. (2005). Anatomical and physiological foundations of cerebellar information processing. *Nat Rev Neurosci*, **6(4)**, 297-311.
2. De Zeeuw, C.I.; Hansel, C.; Bian, F.; Koekkoek, S.K.; van Alphen, A.M.; Linden, D.J.; Oberdick, J. (1998). Expression of a protein kinase C inhibitor in

Purkinje cells blocks cerebellar LTD and adaptation of the vestibulo-ocular reflex. *Neuron*, **20(3)**, 495-508.

3. Koekkoek, S.K.; Hulscher, H.C.; Dortland, B.R.; Hensbroek, R.A.; Elgersma, Y.; Ruigrok, T.J; De Zeeuw, CI. (2003). Cerebellar LTD and learning-dependent timing of conditioned eyelid responses. *Science*, **301(5640)**, 1736-9.
4. www.ni.com.

Measuring and analyzing dispersal

R.P.J. Potting[1], F.L. Wäckers[2]
[1]*Wageningen University, The Netherlands*
[2]*Netherlands Institute of Ecology (NIOO)*

Abstract

The dispersal and movement behavior of animals has a central role in modern ecology and is fundamental to many issues in applied ecology. The challenge of quantifying and understanding the mechanisms of dispersal is widely acknowledged [1, 2]. The measurement of dispersal and movement of free ranging animals has been a very difficult task. However, the rapid advancement of techniques to track animals, especially with regard to miniturasation and increasing capability at reduced cost, continually extends the limits of what can be achieved with established methods [8]. These novel tracking techniques has enabled to measure dispersal of animals that were previously impossible to measure. The dispersal symposium gave a state of the art of techniques used to monitor and analyze animal movement and dispersal in the field at different spatial scales.

Keywords

Dispersal, mark-recapture, tracking individuals, modeling

1 The definition of dispersal and movement

Dispersal refers to intergenerational movement and incorporates movement of individuals from the location where they were born, to another location where they will reproduce. Dispersal has important implications at multiple scales: for the survival of individuals, for the dynamics of populations and communities and for the persistence and geographical distribution of species [7]. Intragenerational movement mediates how species respond to environmental conditions and mainly includes the searching behavior for resources within the home range or territory of the animal. This type of movement behavior is significant for behavioral ecologists who are interested in the proximate (how) and ultimate (why) causes of dispersal. At this scale the encounters with habitats, food, mates, competitors and predators within the home range of the animal are studied. Hereafter, we do not make a specific distinction between movement and dispersal.

2 The scale and measurement of dispersal

Dispersal determines the distribution of animals at a variety of spatial scales and is central to the population dynamics of species. Depending on the spatial scale, dispersal can be measured directly by following individuals, by marking and subsequently recapturing them, or tagging them electronically and detecting them automatically. At the largest scale, such as annual bird or insect migrations to new breeding or overwintering areas, dispersal is studied using rings or isotopes [6]. At a smaller scale of hundreds of meters, direct observation methods are used, in which the animal carries an active or passive tag or marking [5]. For example, to track the spatial dynamics of small insects at a field or landscape scale, mark-release recapture is a standard method. With this technique a sample of insects is captured, color marked, released at a particular point and recaptured with a grid of traps in the landscape. With this technique only limited information is gathered on the dispersal behavior, because these marking techniques are limited in their ability to identify individuals with unique codes. The direct tracking of individual movements depends on appropriate methods being used relative to the body size of the animal. Larger animals are able to carry relatively heavy transmitters, which enable tracking of individuals.

3 The significance of dispersal

Dispersal is a crucial process that affects the spatial dynamics of animal species in fragmented landscapes. Direct measurements of the movement of individuals, such as mark-release recapture and the tracking of individuals, are informative for questions of habitat use, local population structure, and spatial dynamics. Knowledge of the decision mechanisms, such as when, how and where to travel and settle, is essential for understanding dispersal, for modeling it and for putting theories to practical use. In applied ecology there is a desire to predict how animals use the habitats in their environment in order to determine how the species may be best managed. Agro-ecologists may want to slow the spread of pest species or enhance the efficacy of their enemies. Conservation biologists try to design landscapes that enhance the survival probability of vulnerable species. Despite the fact that dispersal has been identified as a key process it is the least understood factor in conservation biology and agro-ecology. One of the reasons for this gap in knowledge is that it is a difficult task to track the whereabouts of individual animals in the field.

4 The Dispersal Symposium

The dispersal symposium gave an overview of existing and novel techniques for the objective and automated tracking and analysis of the movement behavior of animals in the field. Mammals and birds are able to carry battery operated active tags that are used in radio tracking. Radio tagging is a powerful study technique, but requires careful planning of equipment and attachment procedures [3]. The paper by van den Brink gave an overview of the possibilities and constraints of the use of telemetry for assessing habitat use and home ranges of vertebrates in the field. Tracking larger animals is generally easier than tracking small animals, such as insects. But as three symposium speakers demonstrated, technology is advancing rapidly. The paper of Osborne reviewed the unique work on a scanning harmonic radar system which enables accurate tracking of the complete flight paths of individual insects, such as honeybees, over hundreds of meters. Griffiths and coworkers presented another novel technique to track small insects in the field utilizing a laser marker to etch a unique codes on the body of beetles. An alternative to tagging is the use of physiological natural markers to identify the whereabouts of the animals. An example of this strategy is the work by Wäckers who presented a novel technique to study the feeding behavior of small wasps in the field by measuring the sugar profile of individual insects with the HPLC technique.

Dispersal data from empirical studies are essential to formulate dispersal theories and test these in predictive dispersal models. A fundamental problem with modeling dispersal is the lack of understanding of the behavioral mechanisms that cause individuals to disperse particular distances and directions and to what extent this is a result of responses to the landscape or response to conspecifics. Spatially explicit population modeling incorporating individual movement behavior offers a powerful technique to study the mechanisms involved in dispersal behavior and spatial dynamics of populations in heterogeneous environments [4]. Spatially explicit population models incorporate assumptions about the mechanisms that initiate movement and settling. These models can be applied to determine optimal management strategies to guide pest management or conservation. The paper by Potting gave an overview how a behavior based simulation technique can be used to predict the dynamics of interacting individuals in complex landscapes.

References:

1. Bullock, J.M.; Kenward, K.E.; Hails, R.S. (2002). *Dispersal Ecology*. Blackwell, Oxford.
2. Clobert, J.; Danchin, E.; Dhondt, A.A.; Nichols, J.D. (2001). *Dispersal*. Oxford University Press.
3. Harris, S.; Cresswell, W.J.; Forde, P.G., Trewella, W.J.; Woollard, T.; Wray, S. (1990). Home-range analysis using radio-tracking data – a review of problems and techniques particularly as applied to the study of mammals. *Mammal Review* **20**, 97-123.
4. Judson, O.P. (1994). The rise of the individual-based model in ecology. *Trends in Ecology and Evolution* **9**, 9-14.
5. Kenward, R.E. (1987). *Wildlife radio tagging – equipment, field techniques and data analysis.* Academic Press, London.
6. Kruuk, H.; Gorman, M.; Parish, T. (1979). The use of 65Zn for estimating populations of carnivores. *Oikos* **34**, 206-208.
7. Nathan, R. (2001). The challenges of studying dispersal. *Trends in Ecology and Evolution* **16**, 481-483.
8. Riley, J.R.; Drake, V.A. (2002). Technologies for insect movement and migration research. *Computers and Electronics in Agriculture*, **35**, 57-62.

Tracking individual insect flights with harmonic radar

J.L. Osborne, A.D. Smith
Plant and Invertebrate Ecology Division, Rothamsted Research, Harpenden, UK

Abstract
A harmonic radar system for tracking individual flying insects at low altitude has now been used successfully on bumblebees, honeybees, butterflies and moths. A small transponder is attached to the dorsal thorax of an insect, and this is detected by the radar over a range of about 900m (horizontally). The radar has provided novel insights into how individuals search for food resources in complex landscapes. It has also enabled us to answer questions about the navigational abilities of these insects whilst in flight.

To illustrate the power of this technique, examples are given of how honeybees, bumblebees and butterflies explore the landscape during their first flights, and how they exploit the landscape for food resources.

Keywords
Harmonic radar, bee behavior, tracking individuals, insect flight.

1 Introduction
Tracking insect movements over space and time presents a great challenge, because insects are generally small, often very numerous, frequently winged and can travel extremely long distances. Direct methods of studying insect movement involve either characterizing the actual path of the insect's movement, or performing mark-recapture experiments. Movement paths have usually been characterized either by eye or by video. The insect's position at different intervals can be recorded digitally, on audio tape or with numbered flag markers. Although this is the most accurate and precise way of recording insect movement, it is time-consuming and can only be done on a relatively small spatial scale. Mascanzoni & Wallin[5] were the first to use a detection system, based on the harmonic radar principle (originally developed for finding avalanche victims) to locate ground-dwelling beetles. It has subsequently been used on other carabids[3,4,16,17], and on butterflies[14] and caddis flies (Briers *et al.* unpublished data). The insect is tagged with a diode and aerial of 3-5cm length which, when illuminated with microwaves, radiates a signal at the second harmonic frequency of the transmitted one. The microwaves are emitted by portable detection equipment, used to locate the tagged insect, and a marker is placed in the ground where the insect is detected. After a time gap (e.g. 15 mins[16]), another search is made and another positional fix is taken. Although described as a "radar", this system gives no range information and is, rather, a portable *direction-finding* device: alone it cannot produce geometrically accurate maps. It is useful for slow-moving, or stationery objects, but is not suited to tracking "real time" trajectories, or use with fast-moving insects.

The scanning harmonic radar system described in this paper, developed by Professor Joe Riley and Alan Smith, is currently the only technique which enables accurate tracking of the complete flight paths of individual, large, low-flying insects over hundreds of metres[7,11].

2 Methods

2.1 Harmonic radar system
The harmonic radar is a 3.2 cm wavelength, 25 kW peak power, azimuthally scanning, dual frequency system. A transponder, consisting of a vertical dipole aerial (16mm long) with a small Schottky diode and inductive loop in the centre, is carefully attached to the dorsal surface of the insect (Figure 1.). The transponder weighs approximately 12mg. It captures some of the energy in the radar transmissions and re-radiates part of it at double the transmitted frequency. This returned signal is easily distinguished from even strong echoes reflected by ground features. Since the illuminating radar delivers the energy to operate the transponder, no "on-board" battery is required and extreme miniaturization is therefore possible.

Figure 1. *Bombus terrestris worker with harmonic radar transponder, feeding on oilseed rape (Brassica napus) flowers*

The transponder can be detected within a circle of radius 900m centered on the radar, in unobstructed flat terrain and can be detected from ground level to about 6m altitude. The two radar antennae (Figure 2.) rotate in azimuth at 20 revolutions per minute, so position fixes (precision ±3 m) from the transponder are received approximately every 3 seconds. These are digitized to provide a temporally and spatially explicit, or geometrically accurate, track of the insect's flight path.

Figure 2. *Scanning harmonic radar equipment*

For individual insects, the direction, distance and straightness of flight can be measured, and even destination if it is within range. The insect's course control and navigational performance can be investigated by examining flight speed and direction, and the relationship with wind speed and direction.

2.2 Flight performance of insects with transponders

The flight performance of honeybees and bumblebees is not affected to a significant degree by the transponders. For example, bumblebees were found to collect similar quantities of nectar and pollen with and without the transponder in place[6], although foraging trip duration can be longer for bees with transponders. This is likely to be due to the transponder affecting how the bees manipulate certain flowers. Calculations indicated that the increased drag imposed by the transponder was negligible compared to the drag from the bee's own body[7]. Also, honeybees equipped with transponders have been tracked doing path integration on the way to a feeder and to the hive. The distances and directions of the vectors that displaced bees fly is exactly what other studies predict[12,13].

Experiments in a flight room showed that small tortoiseshell (*Aglais urticae*) and peacock (*Inachis io*) butterflies are not hindered by the presence of the transponder and continue to fly and feed as frequently as they do without the transponder in place[1].

2.3 Constraints of the system

The constraints of the system include the size of the transponder. Lighter versions (~3mg), have been used on moths[8] but they have not proved to be robust for use in the field on strong-flying insects such as bees. Individual insects with transponders cannot be distinguished from each other, so only one or two can be tracked at a time. A relatively flat landscape, devoid of tall vegetation, is required to utilize the maximum spatial range of the radar. The system can not be used in forests, tree-filled landscapes or hilly areas, and altitudinal range of the present system limits the variety of insects that can be studied. The radar equipment is large and, so far, has not been mobile during studies, although that would increase its horizontal range.

2.4 Studies of insects

The technique has been used to investigate flight behavior of bumblebees[6,9,10], honeybees[2,12,13], moths[8] and butterflies[1]. Details of the individual experiments are available in those papers. Some of the main findings are summarized here.

3 Results

3.1 Bee behaviour

In the first study of bumblebee foraging flight, 65 bees were tracked flying to and from their colony over a 700 m range (Fig. 2 in Osborne *et al.* 1999[6]). The results showed that bumblebees do not necessarily forage close to their nest. They fly along fast, straight flight paths and show route constancy between trips. They also actively compensated for wind speed and direction (Fig. 3 in Riley *et al.* 1999[9]).

The most recent harmonic radar research has provided detailed insights into honeybee navigation mechanisms (Figure 3.). Honeybees were tracked to provide a convincing proof of Von Frisch's hypothesis of the role of the waggle dance in bee communication[13,15]. It was also found that trained honeybees use an "automatic pilot" mode, making vector flights and tending to disregard landscape cues[12] when they were displaced from a feeder with which they were familiar. Naïve honeybees have also been tracked on their first flights away from their colony. They were observed to make characteristic orientation flights that were very different from those of experienced foragers[2]. A subsequent study of bumblebee orientation flights using the radar showed that the orientation flights of bumblebees, before they start to forage, are more wide-ranging than those of honeybees and involve sampling of different forage sources. Bumblebee search strategies may differ from those of honeybees because they cannot obtain information on forage location from nestmates within the colony, as honeybees can through the waggle dance communication.

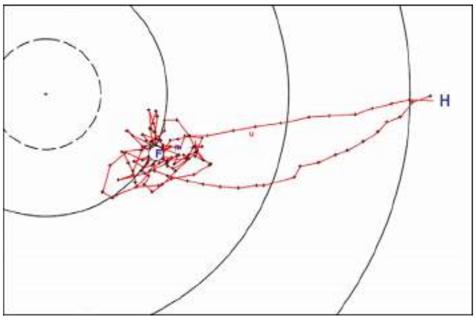


Figure 3. *Harmonic radar track (red) of honeybee leaving hive H, searching for feeder at F (with which they had become familiar, but which had been removed for test), and returning to hive. Radar rings are 100m apart.*

3.2 Butterfly behaviour

For the first time, the flight paths of five butterfly species were successfully tracked using harmonic radar within an agricultural landscape[1]. Until now, butterfly mobility has been predominantly studied using visual observations and mark-release-recapture experiments. For small tortoiseshell (*Aglais urticae*) and peacock (*Inachis io*) butterflies, two main styles of track were identified; (A) fast linear flight and (B) slower non-linear flights involving a period of foraging and/or looped sections of flight. The results provide tentative support for non-random dispersal and a perceptual range of 100 – 200m for these species and demonstrated that the harmonic radar methodology will be of significant value for future investigation of butterfly mobility and dispersal.

Figure 4. *Small tortoiseshell (Aglais urticae) butterfly with radar transponder*

4 Conclusions

This harmonic radar technique has thus proved invaluable for investigating the navigation and foraging behavior of bees and butterflies. It is most suitable for studying the flight patterns of large insects, flying at low altitudes over a range of hundreds of meters in relatively open landscapes. To date, honeybees and bumblebees have been most frequently studied. Butterflies or moths, particularly those grassland species thought to be localized in their movement, are also an ideal choice. Other hymenopterans, such as large solitary bees or wasps, would also be suitable, and it would be interesting to try the technique on Coleoptera, Orthoptera or Odonata; orders containing large, sturdy insects capable of carrying the transponder.

We thank Prof J Riley, D. Reynolds, E. Cant, A. Martin and N. Carreck for their contribution to the work in this paper. Rothamsted Research receives grant-aided support from the Biotechnology and Biological sciences Research Council of the United Kingdom.

References:

1. Cant, E. T.; Smith, A. D.; Reynolds, D. R. & Osborne, J. L. (2005). Tracking butterfly flights across the landscape with harmonic radar. *Proceedings of the Royal Society: Series B* **272**, 785-790.
2. Capaldi, E. A. et al. (2000). Ontogeny of orientation flight in the honeybee revealed by harmonic radar. *Nature* **403**, 537- 540.
3. Kennedy, P. J. & Young, M. R. (1993). Radar tracking the movements of ground beetles in a farmland mosaic. *Aberdeen Lectures in Ecology* **6**, 18-19.
4. Kennedy, P. J. (1994). in *Carabid beetles: ecology and evolution* (eds. Desender, K., Dufrêne, M., Loreau, M., Luff, M. L. & Maelfait, J.-P.) pp. 439-444 (Kluwer Academic Publishers, Dordrecht, The Netherlands).
5. Mascanzoni, D. & Wallin, H. (1986). The harmonic radar: a new method of tracing insects in the field. *Ecological Entomology* **11**, 387-390.
6. Osborne, J. L. et al. (1999). A landscape scale study of bumble bee foraging range and constancy, using harmonic radar. *Journal of Applied Ecology* **36**, 519-533.
7. Riley, J. R. et al. (1996). Tracking bees with harmonic radar. *Nature* **379**, 29-30.
8. Riley, J. R. et al. (1998). Harmonic radar as a means of tracking the pheromone-finding and pheromone-following flight in male moths. *Journal of Insect Behavior* **11**, 287-296.
9. Riley, J. R. et al. (1999). Compensation for the wind by foraging bumble bees. *Nature* **400**, 126.
10. Riley, J. R. & Osborne, J. L. (2001). in *Insect movement: mechanisms and consequences* (eds. Woiwod, I. H.; Reynolds, D. R. & Thomas, C. D.) pp. 129-157 (CABI, Publishing, Wallingford).
11. Riley, J. R. & Smith, A. D. (2002). Design considerations for an harmonic radar to investigate the flight of insects at low altitude. *Computers and Electronics in Agriculture* **35**, 151-169.
12. Riley, J. R. et al. (2003). The automatic pilot of honeybees. *Proceedings of the Royal Society B* **270**, 2421-2424.
13. Riley, J. R.; Greggers, U.; Smith, A. D.; Reynolds, D. R. & Menzel, R. (2005). The flight paths of honeybees recruited by the waggle dance. *Nature* **435**, 205-207.
14. Roland, J.; McKinnin, G.; Backhouse, C. & Taylor, P. D. (1996). Even smaller radar tags on insects. *Nature* **381**, 120.
15. von Frisch, K. (1967). *The dance language and orientation of bees* (Oxford University Press, London).
16. Wallin, H. & Ekbom, B. S. (1988). Movements of carabid beetles (Coleoptera: Carabidae) inhabiting cereal fields: a field tracing study. *Oecologia* **77**, 39-43.
17. Wallin, H. & Ekbom, B. (1994). Influence Of Hunger Level and Prey Densities On Movement Patterns In 3 Species Of Pterostichus Beetles (Coleoptera, Carabidae). *Environmental Entomology* **23**, 1171-1181.

Recent use of telemetry for assessing home ranges, habitat use and mortality in free ranging vertebrates

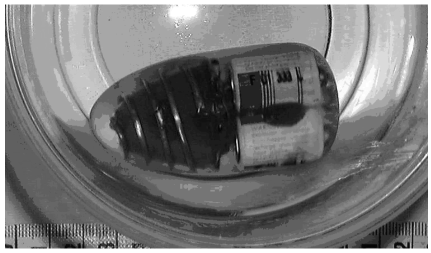
N.W. van den Brink, Jansman, H.A.H., Lammertsma, D.R., Niewold, F.F.J., Müskens, G.J.D.M.
Alterra Wageningen UR, Wageningen,The Netherlands

Abstract
Radio-telemetry is a long proven methodology to assess the whereabouts of vertebrates. Nevertheless, the field of radio telemetry is still evolving, developing new applications and miniaturizing transmitters. In the current paper the use of telemetry is illustrated, for assessing home ranges, habitat use, and dispersal in studies focused on wild ranging animals. At Alterra, or its predecessors, several studies have been conducted in the assessment of behavior of birds, mammals and even amphibians using telemetry. These studies have used implantable transmitters for animals large enough to carry one. In case of smaller animals, especially birds, external transmitters have been applied. The former have the advantage that the animals do not lose them, and that they do not obstruct the animals. The latter are easily applied, but animals may lose transmitter or when applied in a wrong way the transmitter may hamper the animal. In the paper different case studies will be presented, both on mammals and birds.

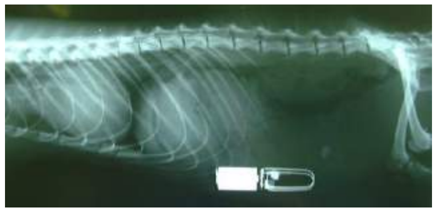
Keywords
Telemetry, home range, habitat use

1 Introduction
Radio telemetry has proven to be very useful in studies assessing home ranges in vertebrates. Studies at Alterra encompass avian studies as well as mammalian and amphibians. For instance for the black-tailed godwit (*Limosa limosa*) the effects of agricultural management on habitat use of birds, and survival of chicks has been assessed based upon telemetric observations (7). Some other currently ongoing case-studies will be addressed in more detail.

Case 1: otter
Currently, a program is conducted, in which otters (*Lutra lutra*) are reintroduced in the Netherlands. Part of this program is to monitor the whereabouts of the reintroduced otters by means of radio-telemetry (5). For this program transmitters have been implanted in otters by surgery by a vet (figure 1). Signals of these transmitters can be received by hand receiver, and based upon manual readings the location of the otter can be assessed. This is a laboring procedure, but based upon these results the home ranges of individual otters can be assessed. In figure 2 the home ranges of several otters have been plotted (5).
In the study on the otter it was chosen to use internal transmitters because external ones would hamper the animals during swimming. In the first seven otters who settled in the area of release, no problems related to the transmitters were encountered during the study. The transmitters lasted for over a year, and as such the otters could be located for a prolonged period of time. The data gathered with the telemetry procedures provided to be very valuable in assessing the home ranges of the otters.



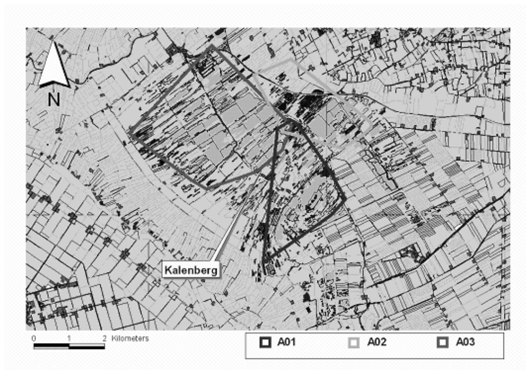
A



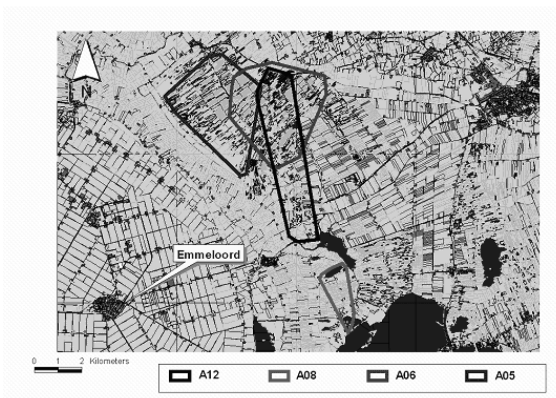
B

Figure 1. *Implantable transmitter, used for the telemetry of otters (total width app. 7 cm), and figure 1B: X-ray of otter with implanted transmitter. Of the transmitter only the antenna and the battery are visible at the x-ray.*

This was very essential in order to evaluate the success and extent of the re-introduction program of the otter. For more info on the re-introduction program of the otter see: www.otter.alterra.nl or www.terugkeer.nl.



A



B

Figure 2. *Home ranges of 7 otters (three in panel A, four in panel B), based upon location assessment by radio telemetry.*

Case 2: little owl

At present, a study on the habitat use of little owls (*Athene noctua*) is focused on small scale habitat use of the birds within their territory. In this study the owls are fitted with a transmitter, and the signals are recorder automatically. This study is using external transmitters, carried on the back of the owl (see figure 3).

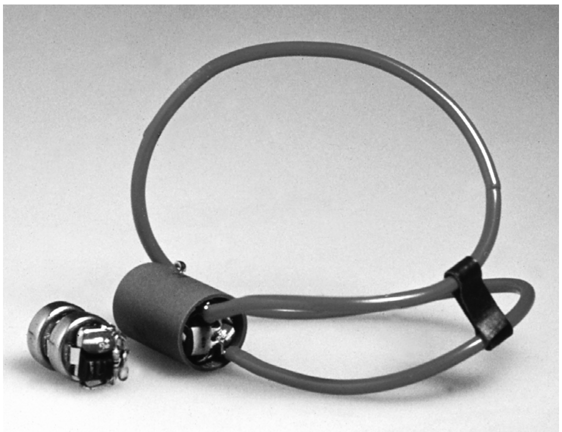


***Figure 3**. Transmitter used with little owls. The harness is used to apply the transmitter, and is the antenna as well. On the left is a transmitter without casing, showing that the transmitter is mostly battery.*

In the field automated recorders are logging the signals from the owls every three seconds. Based upon the relative signals of the different receivers the location of the owl within its territory can be deduced. The scale of this study is much smaller than the one on the otter, maximum 1.5 km instead of several km's. To learn more on the use of telemetry and little owls see: www.berisp.org. In the BERISP program there is also a study on hedgehogs (*Erinaceus Europaeus*), by the University of Antwerp. They used another approach, which is not based on telemetry, but on GPS (8). In that approach a GPS-module (combined with a normal transmitter) is mounted on a hedgehog, which is then released. After a night the hedgehog is located by conventional telemetry, and the GPS data are down-loaded from the GPS module. This GPS module is fitted externally, and taken off after each night. It provides excellent data on the location (5-20 m) of the hedgehog, but the use is currently limited to those applications in which the weight of the device is not of great importance. Application of GPS modules may also be limited do to the possible bad reception in enclosed vegetation. When weight is not a problem GPS can be used in combination with GSM systems, through which the data can be transferred to a computer on-line.

Other recent examples

Other more recent examples of the use of telemetry at Alterra were focused on a wide variety of species: harbour seals (*Phoca vitulina*) (see for 6), hamster (C*ricetus cricetus*) (4), and white-fronted goose (*Anser albifrons*) (2) (this list is not conclusive). In these studies the transmitters have been used externally or internally. In case of the seals they were glued on the skin, and where lost during the molting period. This is not a problem from a scientific perspective, because generally the batteries of the transmitters may have a shorter lifetime than the period that the transmitter is on the animals. In case of the hamster the transmitter were implanted, while with the geese the transmitters were place in a neck collar, which is usually used for color-banding geese.

2 Discussion

Choice of transmitter

The examples mentioned here show that there is a wide variety of applications of telemetry, ranging from larger scale processes as home ranges, or small scale assessment of habitat use. Telemetry is always confined due to several restraints. These restraints are in general: the weight of the transmitter, the size of the transmitter, the energy demand of the transmitter, the range of reception, and the capacity of the battery. These restraints are all related to each other, and are of more importance in smaller animals, and more in flying animals compared to walking animals. In general the weight of a transmitter in birds should not exceed 3-5 percent of the bodyweight (see for examples (1)). When the weight increases, the energy expenditure of the animal is expected to rise, and the behavior may be altered. In walking or swimming animals the weight of external transmitters may not be the first main constraint, however the size and shape may be. One can expect that in swimming animals the streamline of the animal is of great importance, and the transmitter should not result in too much drag. This may be similar for burrowing animals. Internal transmitters will be fitted in the animal for a lifetime. Hence, the transmitter should be inert to biological processes, and should not affect the animal negatively. External transmitters should be designed as such that they are released after a while, for instance by using a drop off system. Another possibility is to glue the transmitters on the skin or feathers. During molt these transmitters will be lost. If a harness is used, see the little owl study, the harness should be designed as such that the birds may get rid of it after a while. The materials used to keep the harness together may be selected for, that they wear after while and may give the animal the opportunity to free themselves of the transmitter. The choice of transmitter type is therefore depending on the type of research question (e.g. the spatial and temporal scales of interest), and of the species of concern. Another topic to be addressed is the frequency of the transmitters. The transmitters used in the otter and little owl studies are transmitting at approximately 30 MHz. This has the advantage that there is limited interference of surroundings like building, trees etc. However, the range of operation is mostly limited from one to several kilometers. Higher frequencies may provide a longer range of operation, but in general there is a greater interference with surroundings. For technical information on radio-telemetry see www.microtes.com.

Other types of information

In clinical studies using small laboratory animals, internal transmitters have been used to automatically gain physiological information on for instance, heart rate, body temperature, and locomotor activity (see for instance (3)). At the moment, the application of such transmitters is limited in free ranging animals. In general the spatial range of transmission is small and the implanted transmitter needs to be connected to organs or located precisely in the body. In free ranging animals, such conditions can generally no be met. Nevertheless, detailed study of the frequency of the signal may provide information on the fact whether the animal is dead or not, on the activity pattern of the animal, or the body temperature. This is only possible with the appropriate experience with animal tracking, and therefore this needs to be conducted by experienced personnel.

References:

1. Brewer, L., Fagerstone, K. (1998). Radio-telemetry applications for wildlife toxicology field studies. *SETAC special publication series*, SETAC Pensacola, Fl, USA.
2. Ebbinge, B., Müskens, G., Oord, J., Beintema, A., Van den Brink, N. (2000). *Stuurbaarheid van ganzen door verjaging en flankerende jacht rondom het ganzenopvanggebied Oost-Dongeradeel (Friesland) in 1999-2000.* Alterra report 128. Alterra, Wageningen, the Netherlands.
3. Kramer, K. (2003). *Applications and evaluation of radio-telemetry in small laboratory animals.* PhD-thesis, University Utrecht.
4. La Haye, M., Müskens, G., Van Kats, K., 2005. *Three years of reintroduction and protection of hamsters in The Netherlands.* De Levende Natuur 106. St. De Levende Natuur, Wageningen, the Netherlands
5. Niewold, F., Lammertsma, D., Jansman, H., Kuiters, L. (2003). *De otter terug in Nederland; Eerste fase van de herintroductie in Nationaal Park De Weerribben in 2002.* Alterra report 852, Alterra Wageningen
6. Reijnders, P., Brasseur, S., Brinkman, A., (2000) *Habitatgebruik en aantalsontwikkelingen van gewone zeehonden in de Oosterschelde en het overige Deltagebied* Alterra report 078. Alterra, Wageningen, the Netherlands.
7. Schekkerman, H., Müskens, G. (2000). *Het gebruik van 'vluchtstroken' door gruttogezinnen.* Alterra report 027. Alterra, Wageningen, the Netherlands.
8. Vermeulen F. 2005. *Modelling accumulation of persistent pollutants through the food chain of the european hedghog to support terrstrial risk assessment.* 15^{th} annual meeting SETAC–Europe 2005. Lille, France.

Laser marking carabid beetles for mark-release-recapture

G.J.K. Griffiths[1], L. Winder[2], C.J. Alexander[1], J.N. Perry[1], J.M. Holland[3], P. Kennedy[4], A. Birt[4] & R. Preston[5]
[1] *Plant and Invertebrate Ecology Division, Rothamsted Research, Harpenden, United Kingdom*
[2] *Department of Biology, University of the South Pacific, Suva, Fiji*
[3] *The Game Conservancy Trust, Fordingbridge, United Kingdom*
[4] *Ecological Sciences, Syngenta, Bracknell, United Kingdom*
[5] *Biological Sciences, University of Plymouth at Seale-Hayne, Newton Abbot, United Kingdom*

Abstract

Marking techniques (e.g. paints, fluorescent dyes, immunoglobin markers and drills) for mark-release-recapture studies are frequently limited either in their ability to provide unique identity codes or to mass mark a sufficient proportion of the population. A technique was developed to etch an alphanumeric code on the elytra of carabid beetles using a 25-Watt CO_2 laser (Fenix). The laser marking system enabled rapid mass-marking of carabid beetles with unique 3-digit identity codes that were permanent and easy-to-read. Marks were applied in the field by establishing a generator-powered marking system in a transit van. The laser marking system was tested by conducting a field-scale capture-recapture study on three carabid beetles, *Pterostichus cupreus, P. madidus* and *P. melanarius*. Although carabid beetles were marked in this study, the technique is applicable to a wide range of organisms with a hard exoskeleton.

Keywords

Laser, mark, recapture, carabid, population

1 Introduction

Studies that utilize data collected from the capture, mark, release and subsequent recapture of individually marked organisms may provide valuable ecological information regarding displacement and population size in combination with appropriate statistical analyses [5, 7, 8, 9]. However, a sufficient proportion of the population must be marked for accurate estimation of dispersal and population parameters [6, 9]. Hence, techniques which facilitate the rapid mass-marking of individual insects can contribute to more robust ecological studies.

Ideally, a marking system should be rapid and provide permanent individual codes that do not affect behavior or survival [9]. Methods available for marking insects have recently been reviewed [3] and commonly involve the use of paints, dyes, immunoglobulin markers, abraded and branded marks, radio isotopes and rare earth elements.

Here, the development of a field-based technique that utilizes an industrial laser marker to lase unique codes on the elytra of ground beetles is reported. In preliminary trials, a scanning electron microscope was used to measure the penetration of the laser into the exoskeleton and the effect of marking on individual beetles was tested in survival trials. The efficacy of the marking system was tested in a field scale capture-recapture study which marked over 8000 ground beetles with individual codes in order to investigate the spatial distribution, movement behavior and population parameters [2, 12]. Ground beetles are ideally suited to capture-recapture studies since they are abundant, ubiquitously distributed in farmland, and readily captured using pitfall traps [10].

2 Method

2.1 Laser-marking system

A mobile laboratory was designed and built to accommodate the laser marking system in a panel van with a floor area of 5.5m^2. Benching (3.5 x 0.6m) was

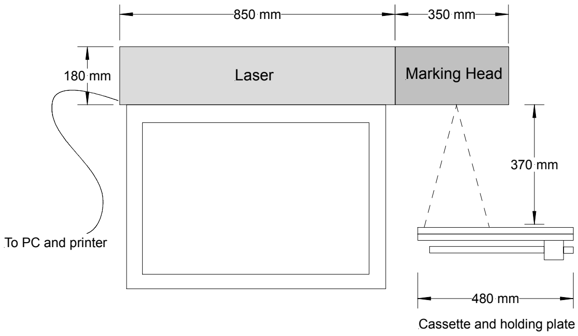


Figure 1. *Schematic diagram of laser-marking system.*

constructed and all equipment securely bolted to the work surface. The system was powered with a small electrical generator (Honda EU 30is). The laser marking system comprised a Synrad Fenix laser marker (Synrad Inc., http://www.synrad.com), a holding plate which restrained beetles during the marking process, a protective shield with interlock for safety purposes and a desktop PC that controlled the system via a VisualBasic program written for this application (Figure 1).

The Fenix 25-Watt, CO_2 laser, was used with a 370 mm focusing lens to provide a suitable incidence angle of the laser beam (19°) and marking area (22 cm x 22 cm). Settings for the laser were determined from preliminary studies [1] to ensure that a readable laser mark was achieved (velocity = 380 mm s^{-1}, resolution = 400 dots inch^{-1}). Power ratings were set for each species marked in the study; *Pterostichus cupreus* L., *P. madidus* Fabricius and *P. melanarius* Illiger were marked at power settings of 24%, 26% and 28% respectively. For juveniles, marking power was reduced by 4% to ensure that the laser did not penetrate the elytra too deeply. A 'simple stroke font' was used to maintain a single laser pass during the lasing process.

For safety purposes a Perspex screen with a sliding door was positioned around the laser marking system. On opening the door, a contact switch was disengaged which disabled a safety interlock. The laser could only be operated after first closing the screen and second turning an interlock key to enable the laser. Additionally, operators wore Photonic Solutions 190-398nm laser safety glasses during use of the system.

Individual beetles were restrained on aluminium cassettes using elastic bands and positioned underneath the laser marker on a holding plate. Elastic bands rested across the pronotum to mimic the ventral pressure carabids experience when burrowing under stones or in soil. Two cassettes (aligned in parallel), each holding 14 beetles, could be loaded onto the holding plate simultaneously. Due to the limited incidence angle of the laser, the holding plate was moveable to allow marking of beetles in two batches of 14 (in two rows of seven beetles).

Control of the laser by a PC required a Fiber Link Controller Card to be installed via an ISA bus slot. The laser was then controlled using WinMark Pro software, supplied as part of the laser marking system. Using WinMark Pro, ".mkh" files may be generated that allow the creation of text or other images to be lased onto an object. Further, WinMark Pro may be automated by the use of ActiveX controls utilized via a suitable development platform. For this application, laser marking and data recording were controlled by a Microsoft Visual Basic program incorporating ActiveX commands. The program controlled four aspects of the marking process. Firstly, information regarding the species, age, sex, and capture location was input for beetles about to be marked. For each of the 14 locations on the cassettes the species- and age-specific power and x-y co-ordinates of the lased mark were set automatically from pre-programmed ".mkh" files. Secondly, a three digit code was drawn from a previously generated Microsoft Access database. The three digit code comprised a series of alphanumeric characters or symbols which were selected to be distinct from each other to maximize readability. In this case, 22 symbols (A, C, F, H, I, N, P, S, T, U, W, X, Z, d, k, m, o, 4, 6, 9, <, >) were used providing 10,648 codes. Thirdly, data was recorded automatically in paper (via an attached printer for data security) and Access database formats. Finally, activation of the laser initiated the marking process (<2 seconds) to etch unique codes readable with the naked eye on each individual. The process was repeated for each batch of beetles.

2.2 Effect of laser marking

A scanning electron microscope (JOEL 5600) was used to examine penetration of the mark into the elytra. Marked elytra were mounted in cold-cure epoxy resin, micro-sectioned and gold-plated to compare marked and unmarked cross-sections.

To determine the effect of marking on survival, cohorts of 30 marked and 30 unmarked *P. melanarius* were observed over a 4-week period. Beetles were placed individually in 9 cm diameter Petri dishes containing damp filter paper, and randomly arranged in a culture chamber. Beetles were fed *ad libitum* on *Lucilea caesar* (L.) maggots and cat food and kept under a natural day/night regime in cool ambient conditions (19-21°C). Food and filter papers were replaced on alternate days.

2.3 Field studies

The efficacy of the laser marking system was tested during 2003 with a capture-recapture study within a 3.75 ha field of conventionally managed winter wheat at Coffinswell, Devon, U.K. A grid, with 12m spacing was arranged within, and fully covered the field with 204 sampling locations. The activity-density of three carabid species, *P. cupreus, P. madidus* and *P. melanarius*, was measured at each sampling location using barrier-connected pitfall traps [11]. Traps were opened for 24 hours on nine weekly occasions between 19 May and 18 July 2003. On each trapping occasion, cohorts of beetles were collected from each of the sampling locations and retained within plastic boxes (with air holes) containing moist crumpled paper towels and food (moist dog food pellets). All unmarked individuals were laser marked using the procedure described above. Any recaptured marked individuals were retained in the same manner and codes were recorded and verified independently by two operators to ensure accurate transcription of information. All beetles were then returned to their point of capture and released within 48 hours.

Spatial pattern was measured and tested with the SADIE 'Red-blue' method. This statistical technique identifies areas in which observed counts are either arranged effectively at random or form clusters of units in local neighborhoods with consistently large (patch cluster index v_i) or small (gap cluster index v_j) counts [4]. Information on the net displacement of individual beetles between the origin and destination trap for each capture-recapture event was calculated using Pythagoras' theorem.

***Figure 2**. Laser marked* P. melanarius *with a 3-digit alphanumeric code 'NNX' etched onto one elytron.*

3 Results

Using the laser marking system, a unique 3-digit alphanumeric code was etched onto the elytra of each beetle (Figure 2). The lased mark penetrated the elytra to a depth of 5.3-29.3 µm whilst elytral thickness ranged from 59.8-126 µm. Hence, the lasing was sufficiently controlled to ensure that the beetles remained unharmed during the marking process. Mortality was identical in marked and unmarked cohorts, in both cases being 2 out of 30 (6.6%). No adverse effects on marked beetles were evident. The time taken to load a cohort of beetles onto a cassette was <5 minutes and the time taken to process and mark the 28 beetles on the holding plate was <3 minutes.

During the field studies, a total of 8,266 beetles were marked. Individuals marked in the first week of the trial and recaptured two months later had codes that remained clear and easy to read and it was concluded that the marks were permanent. The beetle assemblage was dominated by *P. cupreus* which accounted for 98% of captures and only this species was selected for further analysis. For *P. cupreus*, of the 8,046 marked beetles, 2,270 were recaptured. Individuals were recaptured on up to five occasions, although the majority of recaptures (76%) were single events. Those beetles marked in early sampling occasions were caught throughout the study period. Tracks of selected *P. cupreus* recaptured on four sampling occasions show the movement footprints of individuals during their main activity period (Figure 3).

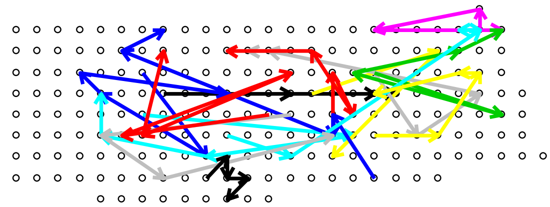


Figure 3. *Movement tracks made by selected P. cupreus recaptured on four occasions*

Analysis of the total activity-density of *P. cupreus* at each location over the entire sampling period (Figure 4) demonstrated that within the central area of the field, catches were up to an order of magnitude larger than those observed at the edges. Median displacement increased with time between release and recapture (Figure 5). The majority of beetles moved between 25 and 100m, but a small proportion of individuals moved considerable distances even within a short time frame. The average distance of all beetles caught at each location was calculated over time. The regression of average displacement on total activity-density was negative over the locations (Figure 6). Hence, there was a large scale region of high activity-density within the field centre where *P. cupreus* tended to exhibit relatively low displacement rates.

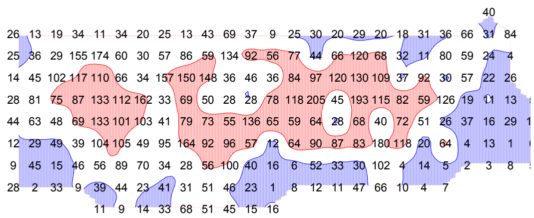


Figure 4*. 'Red-blue' plot of total* P. cupreus *activity-density across the field scale sampling grid. Integers indicate observed data. Red indicates patch clusters of high counts within which interpolated values of $v_i > 1.5$; blue indicates gap clusters of low counts within which interpolated values of $v_j < -1.5$; within white areas counts are arranged effectively at random.*

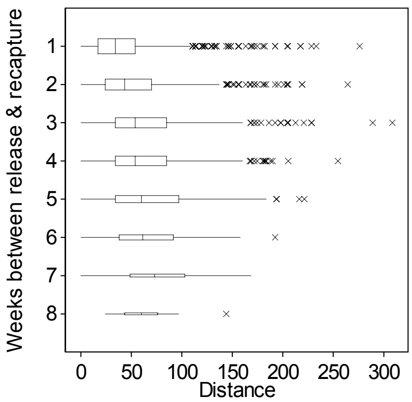


Figure 5 *Distance moved by individuals grouped by number of weeks between release and recapture. Boxes illustrate median and quartiles; the width of the box is scaled by the square root of the number of individuals represented; and the whiskers indicate 1.5 × inter-quartile range beyond the quartiles, distance values further out in the tail of the distribution are displayed as crosses.*

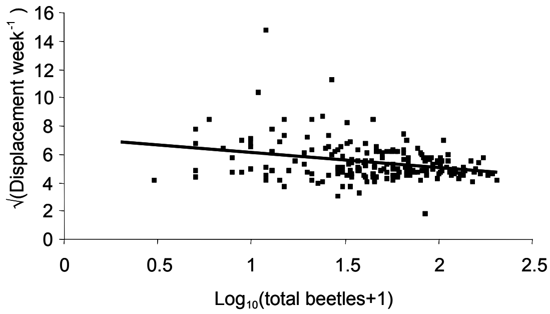


Figure 6*. Regression of average weekly displacement (D) on total beetles caught at each location (B), $\sqrt{D}=-0.08\log_{10}(B+1)+7.2$, $t=16.5$, $P<0.001$.*

4 Discussion

In this study, a laser marking system was used to successfully mark over 8000 field caught beetles to investigate their movement behavior in the field. The marking technique appears to be harmless and highly accurate. For example, laser marks could be positioned on a single elytron to avoid possible penetration of the laser beam into the body cavity through the elytral sutre. The ability to alter power settings, text size, and type of mark provides an adaptable marking system that could be applied to beetles with a wide range of body sizes. Additionally, organisms with a hard shell or exoskeleton may be suitable candidates for laser marking.

The speed and efficacy of the marking system enabled resources to be directed towards the labor intensive retrieval and release of beetles and additional field measurements, rather than the marking process. Consequently, a high proportion of the population could be marked. Unique, permanent and easy-to-read codes enabled multiple recaptures of beetles providing information on movement behavior of individuals throughout their main activity period. Studies conducted over longer time scales would be feasible due to the permanence of the mark.

The field scale capture-recapture study enabled descriptions of spatial distribution, displacement and individual movement footprints of a carabid population. The recapture of 28% of marked individuals also exceeds the 10% threshold required for the accurate estimation of population parameters [6, 9]. Laser marking provides a technique to substantially increase our ability to mass-mark individual organisms and thereby facilitate more accurate estimation of dispersal and population parameters at the field, farm and landscape scale.

We would like to thank Sam Hopwood, Jayne Lambrou, Rebecca Owen, Eugenio Sola, Angie Squires, and the Ag Lab staff for assistance with field work during this study and Roger Wills for generously providing us with a field site. We would also like to thank Ashley Noyce and Keith Jarvis for assistance in building the laser marking system. This work was funded by the BBSRC and supported by Syngenta. Rothamsted-Research is grant aided by the BBSRC.

References:

1. Griffiths, GJ.K., Winder, L., Bean, D., Preston, R., Moate, R., Neal, R., Williams, E., Holland, J.M. & Thomas, G. (2001). Laser marking the carabid *Pterostichus melanarius* for mark-release-recapture. *Ecological Entomology*, **26**, 663-664.
2. Griffiths, G.J.K., Alexander, C., Birt, A., Holland, J.M., Kennedy, P., Perry, J.N., Preston, R. and

Winder, L. (in press). A method for rapidly mass laser-marking individually coded ground beetles (Coleoptera: Carabidae) in the field. *Ecological Entomology*.

3. Hagler, J.R. & Jackson, C.G. (2001). Methods for marking insects: current techniques and future prospects. *Annual Review of Entomology*, **46**, 511-43.
4. Perry, J.N., Winder, L., Holland, J.M. & Alston, R.D. (1999). Red-blue plots for detecting clusters in count data. *Ecology Letters,* **2**, 106-113.
5. Schwarz, C.J. & Seber, G.A.F. (1999). Estimating animal abundance: review III. *Statistical Science*, **14**, 427-456.
6. Seber, G.A.F. (1982). *The Estimation of Animal Abundance*. Griffin, Bucks.
7. Seber, G.A.F. (1986). A review of estimating animal abundance. *Biometrics*, **42**, 267-292.
8. Seber, G.A.F. (1992). Review of estimating animal abundance: 2. *International Statistical Review*, **60**, 129-166.
9. Southwood, T.R.E. & Henderson, P.A. (2000). *Ecological Methods*. Blackwell Science, Oxon.
10. Thomas, C.F.G., Parkinson, L. & Marshall, E.J.P. (1998). Isolating the components of activity-density for the carabid beetles *Pterostichus melanarius* in farmland. *Oecologia*, **166**, 103-112.
11. Winder, L., Holland, J.M., Perry, J.N., Woolley, C. & Alexander, C.J. (2001). The use of barrier-connected pitfall trapping for sampling predatory beetles and spiders. *Entomologia Experimentalis et Applicata*, **98**, 249-258.
12. Winder, L., Griffiths, G.J.K., Perry, J.N., Alexander, C., Holland, J.M., Kennedy, P., Birt, A. (submitted). Large-scale spatially explicit and small-scale localized processes influence the population dynamics of cereal aphids. *Bulletin of Entomological Research*.

Tracing food source use by nectarivorous insects

Felix L. Wäckers
Netherlands Institute of Ecology (NIOO-KNAW), Center for Terrestrial Ecology, PO Box 40, 6666 ZG Heteren, the Netherlands

Abstract

Many insects depend on exogenous sugar sources, such as nectar and honeydew, as a source of energy. The availability of sugar sources can have a strong impact on spatial and temporal dynamics of nectarivorous insects. Whereas sugar feeding has been recognized as an important element in understanding plant-pollinator, plant-herbivore, and plant-carnivore interactions, we know relatively little about the sugar sources utilized by the latter two insect categories. Here we describe two methods that can be applied to study sugar feeding in insects under field conditions. While our research has primarily focused on parasitic wasps, the methods can also be applied to other nectarivorous arthropods.

As a first method, we investigated whether parasitoids could be durably marked through feeding on sucrose solution spiked with strontium or rubidium, or through feeding on extrafloral nectar from *Vicia faba* soil-drenched with aqueous trace elements.

As a second method, we have developed HPLC sugar analysis to study the sugar profile of the parasitoids *Cotesia glomerata* and *Microplitis mediator*. We describe the overall sugar level and the ratio of glucose to fructose as two parameters that in combination unambiguously characterize an individual's nutritional state and feeding history. In addition to revealing past feeding events, HPLC analysis may also provide information about the type of sugar source consumed.

Keywords

Nectar; honeydew; trace elements; HPLC; signature sugars.

1 Introduction

Many arthropods depend on carbohydrate-rich food sources such as (extrafloral) nectar, and honeydew for their development, survival and reproduction. The list of sugar-feeders includes many keystone species such as ants, bees, herbivores and their predators/parasitoids.

The potential fitness benefits of sugar feeding for parasitoids have been established in numerous laboratory studies. Under field conditions, nectarivorous arthropods may exploit various substrates. In addition to floral nectar, their diet may include other sugar sources such as extrafloral nectar, fruits, plant sap, gall secretions, honeydew, Lycaenid dorsal gland secretions, and fungal fluids (Wäckers, 2005). These sugar sources may differ largely with respect to their nutritional suitability and we usually know little about the relative contribution of different sugar sources to the diet of sugar feeders. Here we describe two methods that can be used to assess food source use by sugar feeders.

2 Methods for studying food source use in the field

2.1 Trace elements

Elemental labeling is a true internal labeling technique, because a trace element can be incorporated into tissues of the labeled animals. Trace elements are often used to mark insects for subsequent release ((Hagler & Jackson, 2001). This is analogous to other marking methods used in mark & release studies. An alternative application is the marking of specific food sources in the field to investigate the role of these foods in the diet of insects. In this case insects mark themselves if they feed upon labeled food sources. While this approach primarily addresses food use, it can also be used to study movement and dispersal.

Previous studies had shown that parasitoids can be effectively labeled with Rb when they are reared from herbivore hosts fed diets containing this trace element (Hagler & Jackson, 2001). In a subsequent study (Gu et al., 2001) we wanted to address whether we could also achieve (self-) marking through nectar feeding.

Laboratory experiments were conducted to investigate the feasibility and efficiency of different methods for trace element labeling of the hymenopteran parasitoid *Cotesia glomerata*. We concentrated on labeling parasitoid with Strontium (Sr) or Rubidium (Rb) by: (1) feeding adults on sucrose solution spiked with either element; (2) feeding adults on extrafloral nectar from a plant (Vicia faba) soil-drenched with aqueous Sr or Rb.

Adding markers to sugar solutions had no effect on the acceptance of food solutions by *Cotesia glomerata*. Wasps showed a similar feeding response to sucrose solution spiked with either Rb or Sr at different concentrations. *Cotesia glomerata* had low background levels for both markers (0.43 ± 0.26 µg/g for SR; 0.51 ± 0.25 µg/g for Rb). When feeding adults on sucrose solution spiked with 1000ppm of either element, parasitoids subsequently contained 79 ± 58 µg/g (Sr) or 286 ± 31 µg/g (Rb) (Gu et al., 2001).

Background levels of SR and Rb in the extrafloral nectar of *Vicia faba* were found to be 0.1 and 0.2 µg/g respectively. The content of Rb in Rb-labeled extrafloral nectar increased to 443.6 and 633.9 µg/g as a result of a single soil-drench with this element at 5000 and 15 000 ppm. In the case of Sr, the labeled extrafloral nectar contained an average of 10.9 and 182.6 µg/g following a single soil-drench with this element at 5000 and 15 000 ppm, respectively. Irrespective of the marking method, Sr content in labeled wasps was persistent and did not decline significantly during the 16 days of the experiment.

Due to the transferability of elemental labeling between trophic levels, this technique is particularly suited for studying foraging behavior and trophic interactions in parasitoids and predators (Jackson, 1991).

2.2 HPLC sugar analysis

Rather than marking sugar sources with trace elements, we can also make use of the fact that sugar sources often contain source-specific compounds. If we can detect these compounds in field collected insects, this can be used to establish consumption of this particular food. One example of such food source identification is the tracing of honeydew feeding by honeydew-specific 'signature sugars'. Honeydew often contains specific sugars that are not or only rarely found in other sugar sources (Heimpel et al., 2004). These di- and oligosaccharides are synthesized by the honeydew producer and are believed to have a primary function in osmotic regulation (Wilkinson et al., 1997).

Due to their high specificity, these honeydew-specific 'signature sugars' can be used to establish honeydew-feeding in field collected insects. The most commonly used signature sugar for honeydew is melezitose. This choice is based on the fact that melezitose occurs in substantial amounts in various types of honeydew, while being quite uncommon in other sugar sources. In some cases, the profile of honeydew sugars is specific to the honeydew producing species (Heimpel & Jervis, 2004; Hendrix & Salvucci, 2001) and can be used to distinguish between species of honeydew producers (Wäckers & Steppuhn, 2003).

To obtain information on the food sources used by parasitoids in the field, we determined the sugar profile of field-collected individuals using High Performance Liquid Chromatography (HPLC). We used cabbage as a model system, focusing on the braconid parasitoids *Cotesia glomerata* (L.) and *Microplitis mediator* (Haliday), parasitoids of the large cabbage white (*Pieris brassicae* (L.)) and the cabbage moth (*Mamestra brassicae* (L.)) respectively. These species were chosen as they do not engage in host-feeding and thus are strictly dependent on sugar sources for their nutrition. Parasitoids were collected in Brussels sprouts (*Brassica oleracea*) fields. During the period of the field collections, small populations of the cabbage aphid (*Brevicoryne brassicae* (L.)) and the cabbage whitefly (*Aleyrodes proletella* (L.) were present. Honeydew samples from both phloem feeders on *B. oleracea* were collected to serve as a reference for the interpretation of parasitoid sugar profiles.

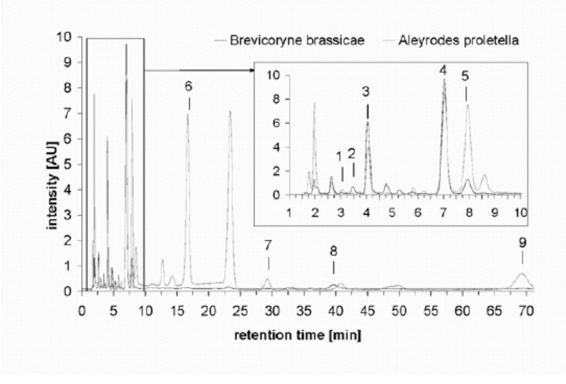


Figure 1*. HPLC chromatograms of honeydew collected from* Brevicoryne brassicae *and* Aleyrodes proletella *feeding on Brussels sprouts [1 - sorbitol; 2- mannitol; 3- trehalose; 4- glucose; 5 - fructose; 6 - sucrose; 7- melezitose; 8 - maltose; 9 - erlose].*

B. brassicae honeydew contained significant levels of trehalose and maltose. *A. proletella* honeydew featured trehalose, melezitose, erlose and high levels of an unidentified sugar (retention time approximately 23 minutes) (Fig 1). The presence of these sugars enabled us to identify honeydew feeding in the Brussels sprouts system, while differences between the two honeydew profiles can even provide an indication as to which honeydew type has been consumed. 80 % of all *C. glomerata* parasitoids and 55% of *M. mediator* contained sugars that are rarely or never found in nectar, while being prominently present in honeydew of cabbage aphids or cabbage whiteflies (Wäckers & Steppuhn, 2003).

References:

1. Gu, H., Wäckers, F. L., Steindl, P., Günther, D. & Dorn, S. 2001. Different approaches to labelling parasitoids using strontium. *Entomologia Experimentalis et Applicata*, **99**, 173-181.
2. Hagler, J. R. & Jackson, C. G. 2001. Methods for marking insects: current techniques and future prospects. *Annual Review of Entomology*, **46**, 511-543.
3. Heimpel, G. E. & Jervis, M. A. 2004. An evaluation of the hypothesis that floral nectar improves biological control by parasitoids. In: *Plant-Provided Food in Multitrophic Interactions* (Ed. by Wäckers, F. L., van Rijn, P. C. J. & Bruin, J.). Cambridge: Cambridge University Press.
4. Heimpel, G. E., Lee, J. C., Wu, Z., Weisser, L., Wäckers, F. L. & Jervis, M. A. 2004. Gut sugar analysis in field-caught parasitoids: adapting methods originally developed for biting flies. *International Journal of Pest Management*, **50**, 193-198.
5. Hendrix, D. L. & Salvucci, M. E. 2001. Isobemisiose: an unusual trisaccharide abundant in the silverleaf whitefly, Bemisia argentifolii. *Journal of Insect Physiology*, **47**, 423-432.
6. Jackson, C. G. 1991. Elemental markers for entomophagous insects. *Southwestern Entomologist Supplement*, **14**, 65-70.
7. Wäckers, F. L. 2005. Suitability of (extra-) floral nectar, pollen, and honeydew as insect food sources. In: *Plant-Provided Food for Carnivorous Insects: a protective mutualism and its applications* (Ed. by Wäckers, F. L., van Rijn, P. C. J. & Bruin, J.). Cambridge: Cambridge University Press.
8. Wäckers, F. L. & Steppuhn, A. 2003. Characterizing nutritional state and food source use of parasitoids collected in fields with high and low nectar availability. *IOBC WPRS Bulletin*, **26**, 203-208.
9. Wilkinson, T. L., Ashford, D. A., Pritchard, J. & Douglas, A. E. 1997. Honeydew sugars and osmoregulation in the pea aphid Acyrthosiphon pisum. *The journal of Experimental Biology*, **200**, 2137-2143.

Testing the adaptive value of behavioral traits in virtual environments

R.P.J. Potting
Laboratory of Entomology, Wageningen University, Wageningen, the Netherlands

Abstract

Individual-based simulation of behavior is becoming a popular tool to study the large-scale consequences of local interactions of members of a population and to study the adaptive value of individual behavioral traits. In these simulations individuals might represent animals in ecosystems, vehicles in traffic or humans in crowds. In an individual-based model behavioral traits and information processing rules of individuals are specified to simulate the behavioral decisions, movements and interactions of individuals in a specified virtual environment. An overview is given on the applications of individual-based models in biology and ecology and artificial human societies.

Keywords

Spatial dynamics, modeling interactions, behavioral ecology

1 Introduction

There is an increasing demand to translate the knowledge of the behavioral ecology of animals to the spatial and temporal dynamics in natural communities, both in theoretical ecology and applied ecology [3]. Although there are some exciting novel techniques for tracking animals in natural environments [14], understanding and predicting animal dynamics in open field systems largely depends on modeling techniques [7].

Traditional analytical equation-based models treat animal populations and their environments as homogeneous entities and usually ignore spatial aspects [9]. However, natural environments of animals are heterogeneous in space and time and insect populations are not homogeneous but consist of locally interacting individuals that differ in their behavioral state [8].

Object-oriented individual-based modeling is an ideal approach for solving complex, stochastic, state-dependent systems that are, in general, impossible to analyze and understand within a purely analytical framework [5, 9]. An individual-based modeling approach is a powerful and flexible way to bridge the gap between individual behavioral ecology and community ecology. An individual-based model can handle biological detail at a low level of organizational complexity (e.g. behavioral and physiological ecology of individuals) and calculates the consequent higher scale trophic dynamics [5, 9]. Thus, these models are perfectly suited to use information from small-scale experiments to identify relevant mechanisms of community dynamics at a field-scale and enable an extrapolation to novel conditions, for example to predict the response of populations to habitat change or management. Here, an individual-based spatially explicit simulation framework is presented with a strong emphasis on state-dependent behavior, cognitive and sensory abilities and movement of insects. The simulations are driven by assigning behavioral rule sets to individuals in the simulation environment. These behavioral rules (i.e.

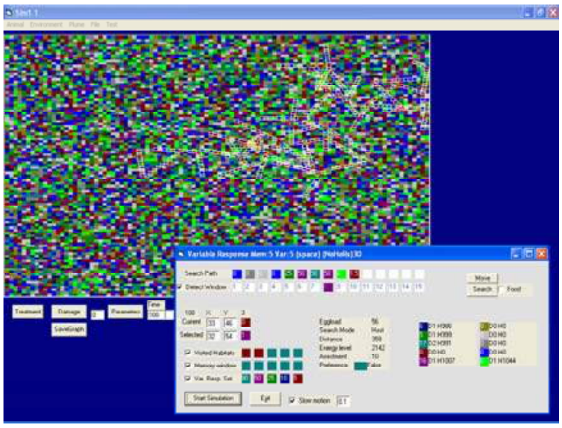


Figure 1. *Snapshot of the interactive simulation framework.*

decisions) enable individuals to respond to stimuli in the computational environment in ways that emulate behavior observed under laboratory or field experiments. The interactive simulation environment (Figure 1) is set up in such a way that it can be easily tailored to particular animal species and environments and can be used to tackle both fundamental and applied ecological questions.

In this paper I will explain the structure and implementation of an individual based model. The use of a behavior-based modeling approach will be demonstrated in behavioral ecological studies and artificial human societies.

2 Behaviour-based modellling

An individual-based simulation framework was build to test the adaptive value of behavioral traits and to investigate the role of behavior in the spatial dynamics of populations in heterogeneous environments. A detailed description of the individual-based spatially explicit simulation framework can be found elsewhere [4, 13]. What follows is a brief description of the general model structure. The simulations are driven by assigning behavioral rule sets to individuals in the simulation environment. These behavioral rules (i.e. decisions) enable individuals to respond to stimuli in the computational environment in ways that emulate behavior observed under laboratory or field experiments (Figure 2).

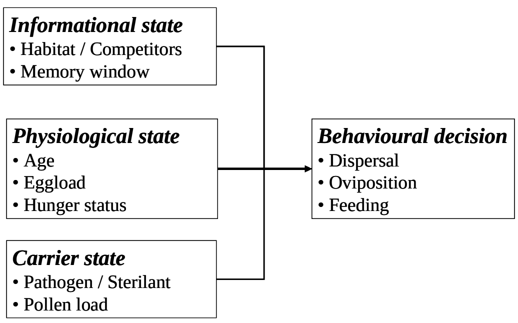


Figure 2. *Overview of state dependent behavioral decisions incorporated in simulation model.*

Basically, the individual-based model continuously keeps track of the physiological state, informational state and position of each individual in the population. Behavioral decisions (e.g. movement) are based on the physiological and informational state of the animal, which generally result in arrestment responses on preferred sites and displacement responses on less or non-preferred sites. The spatial environment is represented by a square lattice of *n* x *n* individual cells. The model is initialized with a specified fraction and spatial arrangement of up to ten habitat types. For each vegetation type several parameters can be set such as preference ranking and initial number of rewards (e.g. prey). Various spatial structures can be specified, see Figure 3 for examples.

The state of each cell (e.g. number of animals, cumulative damage) in the environment is updated for each time step of a simulation run.

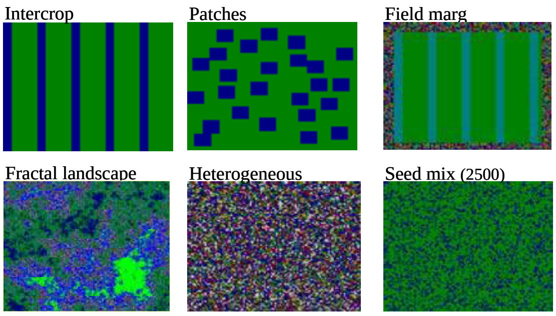


Figure 3. *Examples of environmental setups created in the simulation framework.*

Basically what the simulation does is for each timestep generating state-dependent movements for each animal in the population database and updating the environmental database (with *n* x *n* cells). Decisions and movements are based on stochastic processes. After a simulation run typical results that can be analyzed are the spatial distribution of the population in the virtual landscape and the state of the animals. Typical parameters used in the analysis of adaptive behavior are lifespan and the number of found resources within the given time. The simulation framework has been designed in such a way in that it can be easily tailored to particular animals and environments. It was programmed in Visual Basic where animals were defined as objects with specified properties such as sensory mode. An example of the used pseudocode is given in figure 4.

3 Information processing

Awareness has grown of the importance of learning in the searching behavior of animals. Many insect species can learn visual or olfactory cues associated with successful host location and use these learned stimuli in subsequent foraging decisions [12].

```
For Each Animal in Population
  If Survival = True then
    GetPositionInfo
    GetAnimalState
    GenerateMove
      InflightDetection
        .
  End If
  Update AnimalObject
  Update EnvironmentObject
Next Animal
```

Figure 4. *Pseudocode of the programming language used in the behavior-based simulation framework.*

In heterogeneous environments this may lead to a preference shift towards rewarding habitat types, also termed preference learning [15]. In the simulation environment, movement decisions and perception abilities

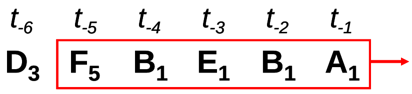


Figure 5. *Graphical representation of the sliding memory window holding the habitat types and their rewards of the last five successful encounters.*

are depended on the informational state of the animal. Animals are able to use information from previous foraging experiences. This is implemented by giving each animal object a memory vector that holds the sequence of habitat types of the previous encountered sites, after the 'sliding memory window' approach of [11]. A memory vector (i.e. array of fixed size) is constantly updated in such a manner that when new information is entered (habitat type of currently visited site) the most distant memory is lost (Figure 5).

Foraging decisions of each individual animal are based upon the composition of their memory vector. Thus, an animal with a memory vector of size 5 will remember and base its foraging decision on the habitat types of the last five visited sites. Animal species with a memory vector of 0 will not use information from previously visited sites. This cognitive ability allows us to investigate the effects of storing (by varying the size of memory window) and processing (by varying decision rules) the stored information. Simulation results show that learning to remember profitable habitat types is adaptive for an animal searching for resources in a relatively unpredictable environment. It was demonstrated that there was a positive correlation between the amount of information stored in memory and the fitness of the animal, estimated by the amount of collected resources (Figure 6).

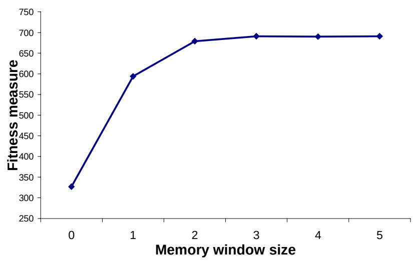


Figure 6. *Mean lifetime fitness of a simulated host searching parasitoids with different numbers of memory units in their decision making. Note that animal with memory window size 0 do not store any information.*

4 Sensory abilities in search behavior

Many insect species, usually active fliers, have good sensory abilities allowing them to use visual or olfactory cues to locate and recognize a mate or a preferred host plant while in-flight [1, 2]. In the simulation framework several sensory modes can be set for the simulated animal. For details see [13]. A nice example to show how a behavior-based model can be used to test the adaptive value of a particular behavioral trait in a particular setting is the flight behavior of mate-seeking moths. Nocturnal moths use species specific odors to communicate with each other. These sex pheromones are usually released by females and create an odor plume that is used by males to locate the calling virgin female. Using a behavior-based model it can be demonstrated that moths use a very simple rule to locate the position of the calling female. First an odor plume is created downwind from the calling female. Then males are released at the downwind opposite end of the female. The movement rule given to the simulated male is: (a) to make a random movement when no pheromone is detected, (b) to make a relatively straight upwind movement if pheromone is detected and (c) to turn 90 degrees left or right from the current flight direction if a pheromone encounter is not followed up by another one. Simulations demonstrate that these simple movement rules are highly adaptive for male moths to locate the position of the calling female (Figure 7).

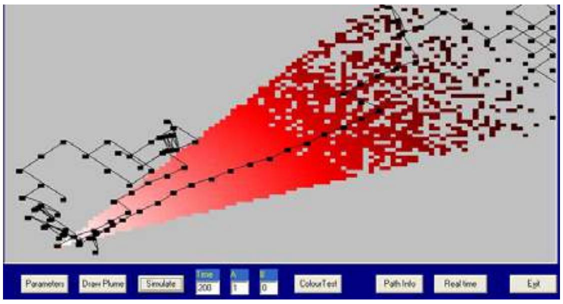


Figure 7. *Example of typical simulation run of male moth orienting in a sex pheromone plume with basic movement rule.*

5 Spatial dynamics of communities

The described simulation framework is not only suitable to investigate individual behavior, but can also be used to study the spatial dynamics and distribution of interacting animals. Understanding the relationship between species abundance and the spatial heterogeneity of the landscape is a primary concern of conservation ecologists and of increasing interest of agro-ecologists [4]. The individual-based modeling approach can be used to upscale individual behavior to population dynamics and to predict how animal populations will react to specified landscapes. For example, to implement agro-ecosystem diversification as a pest management strategy, it is important to have a good understanding of the behavioral ecology of the target pest. Simulation results (for an example see figure 8) demonstrated that the population regulatory effect of the addition of highly attractive or repellent plants in an agro-ecosystem is dependent on the behavioral ecology (e.g. sensory abilities) of the target insect and the spatial arrangement and density of the additional vegetation [13].

6 State dependent behavior in a spatial context

Another example that demonstrates how state-dependent individual behavior may affect spatial dynamics of communities is the switch between host searching and food searching by parasitic wasps of herbivorous insects. enemies of pest insects, for example by providing ample nectar sources in agricultural fields [10].

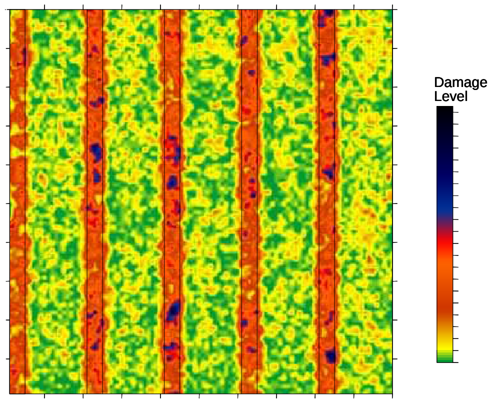


Figure 8. *Example of simulation run of agroecosystem diversification: an attractive intercrop in rows attracts and retains herbivore population. Black intensity indicates cumulative number of visits by herbivores.*

Recently there is an increased interest to manipulate agroecosystems in such a way that they promote natural Parasitic wasps use nectar from flowers as an energy source. Thus a parasitoid needs flowers to feed and herbivorous hosts (e.g. caterpillars) to reproduce. It is assumed that the wasp starts to search for food (i.e. nectar) sources if the energy level drops below a certain threshold. A few nectar feeds restore the energy level and satiated wasps resume host searching (Figure 9). The use of flowering field margins is advocated to enhance the longevity of the parasitoid population. However, results from field experiments show that the effect of a flowering border on the control efficacy of the target pest in the crop is often marginal and highly concentrated near the flowering border. Simulating the searching behavior of parasitic wasps in such an agroecosystem, incorporating an energy state-dependent switching between host searching and food searching, resulted in a similar pattern (Figure 10). The main reason is that wasps do not reach the centre of the crop because they run out of energy and die.

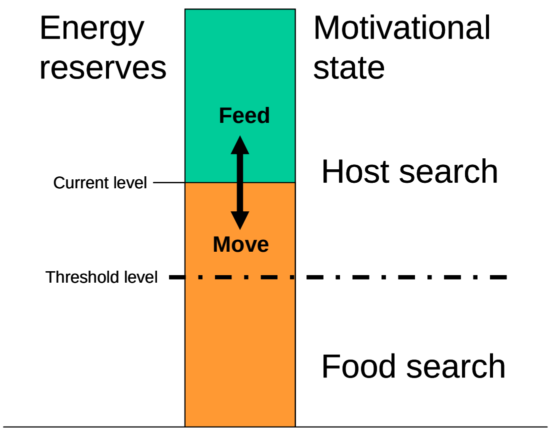


Figure 9. *Graphical representation of state dependent food or host searching in parastic wasps. Note that a feeding bout increases the energy level and each movement decreases the energy level.*

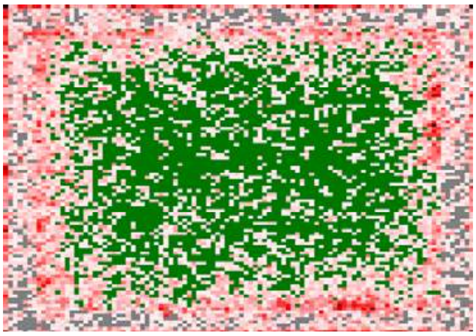


Figure 10. *Activity pattern of parasitic wasps in agroecosystem with a crop (with herbivorous hosts) surrounded by a flowering border (with nectar sources). Note the low activity of parasitoids in the centre of the crop.*

7 Artificial human societies

The individual-based modeling approach is also increasingly used to simulate the response of artificial human societies in different environments. For example, Helbing *et al.* have simulated the dynamics of escape panic of crowds in small rooms with a few exits [6]. They did this by modeling the behavior and movement of each individual in a virtual room. Under normal conditions each pedestrian tries to keep a velocity-dependent distance from other pedestrians and walls. Under panic conditions people are in a rush and contact and counteracting body compression increases resulting in a blocking of the exit. In another simulation they showed that 'herding behavior', the tendency to follow your neighbors within a certain radius, is responsible for the effect that an entire crowd, in a smoke filled room, will eventually move into the same and probably blocked exit. These types of models can be used to test buildings for their suitability in emergency situations.

8 Concluding remarks

Individual-based simulation modeling is a very powerful and flexible way to model the dynamics of complex systems. It is a very suitable technique to upscale the effect of individual behavior to population dynamics in time and space. Furthermore it provides a framework that makes it possible to investigate the adaptive value of behavioral traits in specified environments.

References:

1. Bell, W.J. (1991) Searching behavior patterns in insects. *Annual Review of Entomology*, **35**, 447-467.
2. Bell, W.J. (1991) *Searching behavior. The behavioral ecology of finding resources.* Chapman & Hall, London.
3. Bradbury, R.B.; Payne, R.J.H.; Wilson, J.D.; Krebs, J.R. (2001) Predicting population response to resource management. *Trends in Ecology and Evolution*, **16**, 440-445.
4. Bukovinszky, T.; Potting, R.P.J.; Clough, Y.; Van Lenteren, J.C.; Vet, L.E.M. (2005). The role of pre- and post- alighting detection mechanisms in the response to patch size by herbivores. *Oikos* **109**, 435-446.
5. DeAngelis, D.L.; Gross, L.J. (1992). *Individual-based models and approaches in ecology*. Chapman & Hall, New York.
6. Helbing, D.; Farkas, I.; Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature* **407**, 487-490.
7. Holt, J.; Cheke, R.A. (1997). Modelling. In: *Methods in ecological and agricultural entomology* (eds D.R. Dent & M.P. Walton), pp. 351-378. CABI, Wallingford.
8. Houston, A.; MacNamara, J.M. (1999) *Models of adaptive behaviour: an approach based on state.* Cambridge University Press, Cambridge.
9. Judson, O.P. (1994) The rise of the individual-based model in ecology. *Trends in Ecology and Evolution*, **9**, 9-14.
10. Landis, D.A.; Wratten, S.D.; Gurr, G.M. (2000) Habitat management to conserve natural enemies of arthropod pests in agriculture. *Annual Review of Entomology*, **45**, 175-201.
11. Mangel, M.; Roitberg, B.D. (1989). Dynamic information and oviposition decisions in a fruit fly. *Ecological Entomology*, **14**, 181-189.
12. Papaj, D.R.; Lewis, A.C. (1993). *Insect learning: ecological and evolutionary perspectives*. Chapman & Hall, New York.
13. Potting, R.P.J.; Perry, J.N.; Powell, W. (2005). Insect behavioral ecology and other factors affecting the control efficacy of agro-ecosystem diversification strategies. *Ecological Modelling* **182**, 199-216.
14. Riley, J.R.; Drake, V.A. (2002) Technologies for insect movement and migration research. *Computers and Electronics in Agriculture*, **35**, 57-62.
15. Turlings, T.C.J.; Wäckers, F.; Vet, L.E.M.; Lewis, W.J.; Tumlinson, J.H. (1993). Learning of host-finding cues by hymenopterous parasitoids. In: *Insect learning: ecological and evolutionary perspectives.* (eds D.R. Papaj & A.C. Lewis), pp. 51-78. Chapman & Hall, New York.

Standardized measurement methods for patients with neurological disorders

F.C.T. van der Helm
Delft University of Technology, Delft, The Netherlands

Patients with neurological disorders of the Central Nervous System (CNS) like Parkinson's disease, CVA or Complex Regional Pain Syndrome (CRPS) often exhibit symptoms in their peripheral systems. These symptoms are seen in the motor control system, the thermoregulation system, the autonomous system regulating the bloodflow, etc. Standardized methods to measure these symptoms can be used as diagnostic methods (whether or not to include the patient for a certain patient category), quantifying the severity of the disease, i.e. for monitoring the disease progress or the efficacy of medicine, and/or for exploring the underlying disease mechanism.

Quantifying Bradykinesia in neurological patients

A.C. Schouten, A. Munts, H.E.J. Veeger, F.C.T. van der Helm and J.J. van Hilten

Patients with neurological disorders, like Parkinson's disease (PD), often show bradykinesia. Bradykinesia is demonstrated by slowness and irregularities during fast voluntary movements. Physicians use this sign as an indication for lack of neuromuscular control. Normally bradykinesia is visually scored by a physician. Patients are requested to perform a finger task, i.e. open and close the thumb and index finger as fast as possible, while the performance is scored, grading from 0 (normal) to 4 (can barely perform). This visual score is subjective, physician dependent and has a small resolution. Goal of the project is to develop a method that objectively and accurately scores bradykinesia.

Subjects (PD, n=14; controls, n=36) were requested to perform a finger task for 15 seconds under a high speed digital camera (60 fps). The subjects were instructed to close and open the thumb and index finger as wide and as fast as possible. Markers of colored tape (Leukotape®) were attached to the top of the thumb (green) and index finger (blue). With special software (EthoVision®, Noldus IT) the center of the markers was located with time. As the distance of the hand with respect to the camera was fixed the distance between the fingers could be calculated. The data were further processed using Matlab and two parameters were selected to represent the slowness of the movement: median frequency (MF) and mean absolute acceleration (MA).

Results and discussion: Patients with PD had a significantly lower value for both MF (controls: 3.56 Hz (±0.94); PD: 2.25 Hz (±0.90)) and MA (controls: 21.3 m/s2; (±7.36); PD: 9.76 m/s2 (±5.05)). Statistical analysis revealed that both parameters depended on age and sex, which should be taken into account; handedness was not significant. With a larger dataset we will focus on the most differentiating parameter to describe movement slowness. MA is the primary candidate as acceleration depends on both the distance and speed of the movements, and such is less sensitive to task variations (wide vs. fast). Finally, it is concluded that movement slowness can be objectively quantified, showing the relevance of the method.

Dynamic and Static Thermography and laser Doppler flow measurement in assessing Complex Regional Pain Syndrome type 1

S.P. Niehof

Complex regional pain syndrome type 1 (CRPS1), is a chronic disease of unknown origin. The syndrome is thought to be a damaged nerve disorder that occurs in patients after trauma or surgery at the site of an injury, most often it expresses itself in extremities. It is quite certain that the damage in nerves of the sympathetic nervous system, which are responsible for blood flow and thus partially for tissue temperature, plays an important role in the development of chronic CRPS1 (Huygen et al. 2001). Furthermore, CRPS1 seems to resemble neurogenic inflammation, which produces heat and sympathetic dysfunction (Birklein et al. 2001).

Non-invasive indicators of sympathetic activity were measured using skin temperature and skin blood flow.

Laser Doppler is a feasible method although, until recently, only limited areas could be measured. Several reports on measurement of sympathetic using Videothermography and laser Doppler flow are available in the literature. The value of thermographic as a tool in screening of CRPS1 patients is currently being studied. Static left – right differences in temperature and in deviations in temperature/flow recovery after disturbances are measured using a heat-sensitive infrared camera system and laser Doppler flow. These disturbances are fixed on the sympathetic system by modulating skin temperature.

Observers were used to assess the discriminating value of thermographic pictures.(Niehof, 2005a submitted). In an earlier study we described a promising calculation method to examine the difference between static videothermographic images of CRPS1 patients and healthy controls (Huygen, Niehof et al. 2004). A current study focuses on possible static thermographic differences between patients with a fracture whom do develop CRPS1 and those that don't develop CRPS1.

A study in which the sympathetic system of patients was disturbed using total body cooling followed by total body heating revealed improved diagnosis and some dynamics of the disturbed sympathetic system in CRPS1 patients. (Niehof et al. 2005b submitted). Currently more research is done on patients groups and controls, in a setup where local disturbance are applied using thermographic and laser Doppler flow measurements as indicators of sympathetic activity.

Quantification of Spinal Feedback Gains

E.de Vlugt

Feedback loops are very important for the motion control behavior of human beings. Feedback loops consist of the Central Nervous System as controller, muscles as actuators, position, velocity and force sensors in the muscles and muscle tendons and the skeletal system as the linkage mechanism. During postural tasks the feedback system is very effective to counteract random perturbations that disturb away from the reference position. Robot manipulators are used for imposing force perturbations to the hand, thereby activating the closed loop feedback pathways. By recording signals inside the control loop, the dynamic behaviour of the components of the feedback loop can be estimated using advanced closed-loop system identification techniques. The control settings of the CNS, i.e. the so-called spinal reflex gains, can be estimated based on dynamic transfer functions. These reflex gains are rapidly modulated by healthy subjects, depending e.g. on the frequency content of the perturbation signal and the impedance of the environment. It can be shown that patients with neurological disorders have problems in modulating the reflexive feedback gains.

Disentangling the contribution of the paretic and non-paretic leg to balance control in stroke patients

J.H. Buurke, E.H.F. van Asseldonk, F.C.T. van der Helm and H. van der Kooij

During stroke recovery, both physiological recovery of the paretic leg and adaptative compensation in the non-paretic leg may contribute to improved balance maintenance. Here, we examine a new approach to disentangle these different recovery mechanisms and to objectively quantify the contribution of the paretic and non-paretic leg in balance control.

Balance responses were elicited by random continuous platform movements (forward-backward), and body sway were measured as well as the ground reaction forces below each foot to calculate corrective ankle torques in each leg. From these data the relative contribution of the paretic and non-paretic leg to the total amount of generated corrective torque to correct body sway.

Even in patients with a fairly symmetrical weight distribution (more than 40% on the paretic leg), we observed a clear asymmetry in balance contribution in favour of the non-paretic leg. Overall, the paretic leg made a significantly ($P<0.05$) smaller balance contribution than the non-paretic leg. The presented approach allows for an objective quantification of the contribution of each leg to overall balance maintenance.

Quantifying bradykinesia in neurological patients

A.C. Schouten [1] , A.G. Munts [2] , H.E.J. Veeger [1] , F.C.T. van der Helm [1] , J.J. van Hilten [1]
[1] *Department of Mechanical Engineering, Delft University of Technology, Delft, The Netherlands*
[2] *Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands*

Abstract

Bradykinesia is one of the four symptoms of Parkinson's disease and is characterized by slowness and irregularities during fast voluntary movements. Bradykinesia is also seen in other disorders like complex regional pain syndrome. Normally bradykinesia is visually scored by a physician, who uses this sign as an indication for a lack of neuromuscular control. This score is subjective and has a small resolution. Goal of the project is to develop a method that objectively scores bradykinesia. Subjects were requested to perform a finger tap task under a high speed camera. Markers were attached to the top of the thumb and index finger. Two parameters were selected to represent the slowness of the movement: median frequency (MF) and mean absolute acceleration (MA). Patients with bradykinesia had a lower value for both MF and MA. It is concluded that movement slowness can be objectively quantified, showing the relevance of the method.

Keywords

Bradykinesia, Parkinson's Disease, Complex Regional Pain Syndrome (CRPS).

1 Introduction

Patients with neurological disorders, like Parkinson's disease, often show bradykinesia. Bradykinesia is demonstrated by slowness and irregularities during fast voluntary movements. Parkinson's disease (PD) is a neurodegenerative disorder characterized by: tremor, bradykinesia, rigidity, and impaired postural reflexes. PD is seen predominantly in elderly people. Slowness of voluntary movements is also seen in patients suffering complex regional pain syndrome. Complex regional pain syndrome (CRPS) is a painful condition that typically follows an injury to a limb, which can be minimal or severe (sprain/strain, fracture, contusion/crush injury), although in a number of patients no trauma can be identified. The syndrome manifests with variable combinations of pain, differences in skin color and temperature, edema and sweating (Paice, 1995; Ribbers et al., 1995). The syndrome may spread to other extremities. In addition to the sensory and autonomic signs and symptoms, patients may present or subsequently develop movement disorders (Van Hilten et al., 2001).

Physicians score bradykinesia as an indication for lack of neuromuscular control. Normally bradykinesia is visually scored by a physician. Patients are requested to perform a finger tap task, i.e. open and close the thumb and index finger as fast as possible, while the performance is scored, grading from 0 (normal) to 4 (can barely perform). This visual score is subjective, physician dependent and a clearly has small resolution. Goal of this project is to develop a device that records bradykinesia and a method that objectively and accurately scores bradykinesia.

2 Materials and methods

2.1 Subjects

Eighty four subjects participated in this study. The subjects consists of patients suffering PD (n=14) or CRPS (n=14) and healthy control subjects (n=46). The patients are recruited at the Leiden University Medical Center. The healthy control subjects had no known history of neurological dysfunction and are recruited among partners of the participating patients and among hospital staff. All subjects gave informed consent prior to the experimental procedures.

2.2 Apparatus

The equipment consists of a high speed digital camera (60 fps) and an arm support. The lower arm is supported such that the hand and fingers can freely move. The camera is positioned above the hand, such that movement of the fingers can be recorded, see Figure 1. The arm and camera support assures that the distance between the camera and

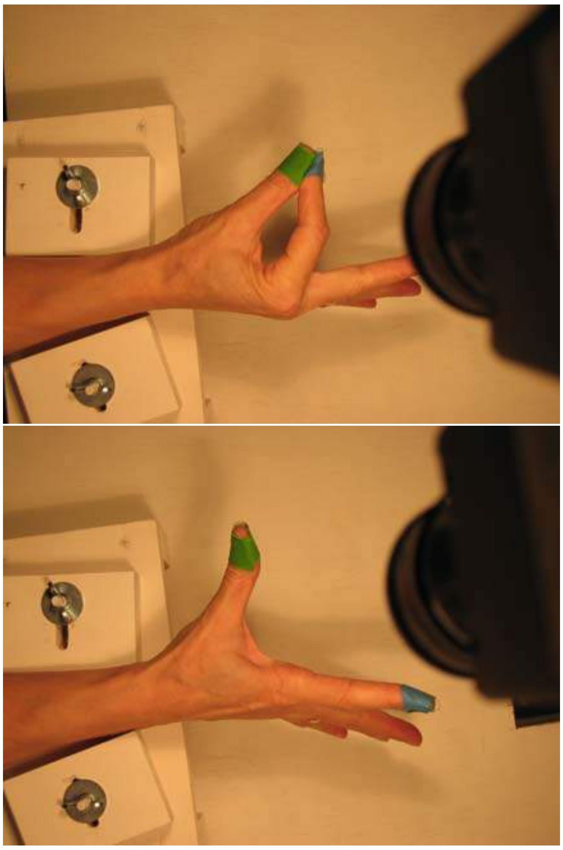


***Figure 1:** Top view of the experimental setup. The upper plot shows the hand of a subject in 'closed hand' position and the lower plot shows the subject in 'open hand' position. At the right of the pictures the high speed digital camera can be seen. The subject's lower arm is supported to prevent movement of the lower arm and to keep the hand at a fixed distance of the camera.*

	Control	PD	CRPS	significance	Post hoc (Tukey)			significance with correction
					PD / controls	CRPS / controls	PD / CRPS	
Number of hands	n=86 L=45; R=41	n=25 L=11; R=14	n=22 L=11; R=11	-	-	-	-	-
Age [years]	45,1 ± 16,6	65,0 ± 12,3	29,6 ± 9,6	<0,001	-	-	-	-
Gender [M/F]	M=23; F=31	M=6; F=8	M=0: F=14	<0,001	-	-	-	-
MF [Hz]	3,56 ± 0,94	2,25 ± 0.90	2,06 ± 1,05	<0,001	<0,001	<0,001	0,674	<0,001
MA [m/s^2]	21,3 ± 7,36	9,76 ± 5,05	10,9 ± 8,07	<0,001	<0,001	<0,001	0,956	<0,001
Trend MA [m/s^3]	-0,59 ± 0,47	-0,25 ± 0,43	-0,23 ± 0,29	<0,001	0,002	0,003	0,952	<0,001

Table 1: *Summarizing the results of this study.The values after the ± denote the standard deviations.In the last column are the values after taking age and gender into account in the analysis.*

the hand are fixed. The camera is connected to a PC through FireWire and the digital recordings are directly stored on hard disk as AVI files.

2.3 Procedures

The subjects sat on a chair in front of the equipment and were required to lay their lower arm onto the arm support. The subjects were instructed to close and open the thumb and index finger as wide and as fast as possible, see Figure 1. Markers of colored tape (Leukotape®) were attached to the top of the thumb (green) and index finger (blue). One trial lasts for 15 seconds, and both hands are recorded once. The procedure including instruction took less than five minutes per subject.

2.4 Data processing

For each subject two recordings are made of 15 seconds at 60 fps, resulting in 900 frames per trial. With special software (EthoVision®, Noldus IT) the centers of the markers were tracked with time. As the distance of the hand with respect to the camera was fixed the spatial orientation of the markers could be calculated. This data were exported and used for further analysis in Matlab (The Mathworks).

First the distance between markers was calculated. Two parameters were extracted from the recordings to represent the slowness of the movement: median frequency and mean absolute acceleration. The median frequency (MF) is obtained by Fourier transforming the signal and calculating the auto spectral density. The median frequency is defined as the frequency at which the power in the auto spectral density is equally divided. The mean absolute acceleration (MA) was calculated by double differentiating the distance and taking the mean of the absolute values. Furthermore the trend of the MA is calculated by taking the MA between 2.5-5.5 s and at 9.5-12.5 s. This parameter is selected to investigate possible ‘fatigue’ during a trial.

2.5 Statistical analysis

All parameters were statistically analyzed with SPSS. Differences between controls and patients are tested with an ANOVA. Differences between groups (CRPS, PD , and controls) were analyzed with a post-hoc test (Tukey).

3 Results

In some recording the markers could not be identified on all frames due to movements of the markers outside the camera arena. These recording were not used for further analysis. All hands of the patients were visually inspected by a physician for bradykinesia. In this study only the affected hand(s) of the patients were used for analysis. In total 67 recording of the left hand were used for analysis (PD, n=11; CRPS, n=11; controls, n=45) and 66 recordings of the right hand (PD, n=14; CRPS, n=11; controls n=41).

The results are summarized in Table 1. It is noticed that a significant difference exists in age between the three patients groups. This was expected as PD primarily affects elderly and CRPS patients affects younger woman.

Patients had a significantly lower value for both MF and MA than controls. Post-hoc testing reveals that both patient groups differ from the controls, but no difference can be seen between the patients groups.

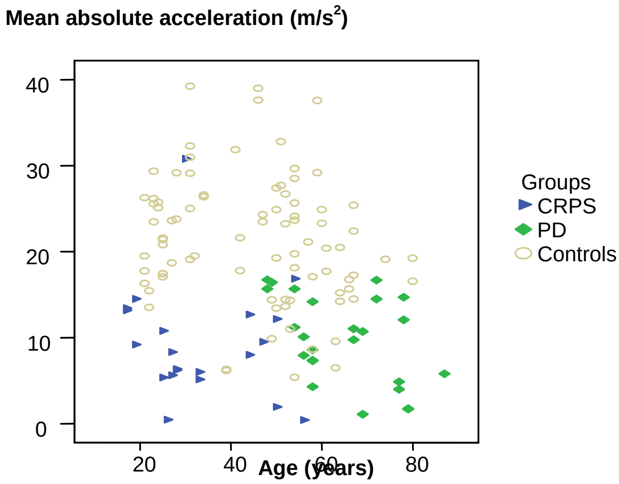


Figure 2: The age of the subjects against the value for the mean absolute acceleration (MA). The control subjects are marked with the open circles, the PD patients are marked with the green squares, and the CRPS patients are marked with the blue triangles.

In Figure 2 is the MA plotted against the age for the two patients groups and the controls. It can be seen that the values for the MA are larger in the controls and the age of this group is spread evenly. Furthermore it can be seen that young controls have a higher score than older controls, i.e. the MA decreases with age. The PD patients score lower than the controls and are older. The CRPS

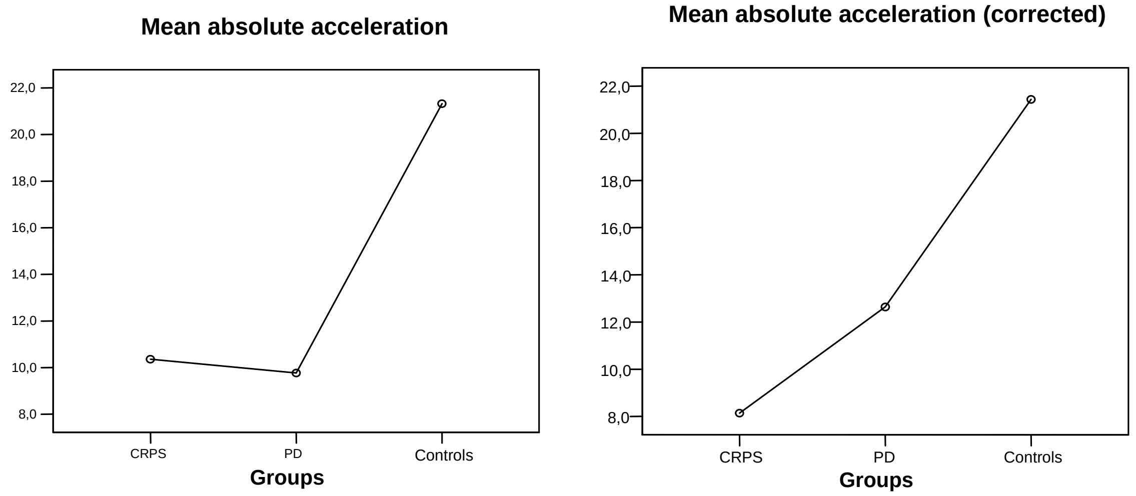


Figure 3: On the left are the values for the mean acceleration for the three group (CRPS, PD, and controls). On the right are the same parameters after age and gender have been taken into account.

patients are younger, and score even lower, than the PD patients.

Statistical testing reveals that age and gender have a significant effect on the parameters and should be taken into account. The results are corrected for age and gender by taking the age and gender as ‘co variable’ and ‘fixed factor’ into the ANOVA. In the left of Figure 3 it can be seen that a clear difference exists between controls and the patients. However the difference between the two patients groups is relatively small. After correction for age and gender the difference in the MA between CRPS and PD increases substantially. This effect is also seen in the MF parameter. By taking gender and age into account in the statistical analysis a significant difference exists between the three groups (CRPS, PD and controls); even between CRPS and PD.

The decrease in MA, i.e. the trend in MA, is in controls approximately twice as large as in patients. Differences are significant although with a higher p-value, compared to MA and MF. The trend in MA being twice as large is not surprising, as the MA itself is also twice as large.

4 Discussion

Bradykinesia, or movement slowness, is a common feature of many neurological disorders. Normally bradykinesia is visually scored by a physician. This study shows that bradykinesia can be scored by analyzing digital recordings of a finger tap task. Both the mean absolute acceleration (MA) and the median frequency (MF) show significant differences between patients and control. The method to record the finger movements takes only a few minutes. This study shows that the method can be useful as a diagnostic tool and future steps will be taken to develop this method towards a diagnostic tool.

4.1 Quantifying slowness in bradykinesia

In patients suffering bradykinesia, or movement slowness, both the values for mean absolute acceleration (MA) and the median frequency (MF) are significantly lower than in healthy control subjects. Although it should be noted that age and gender have an effect on both parameters and should be taken into account. With a larger dataset we will focus on the most differentiating parameter to describe movement slowness. MA is the primary candidate as acceleration depends on both the distance and speed of the movements, and such is less sensitive to task variations (wide vs. fast). Finally, it is concluded that movement slowness can be objectively quantified, showing the relevance of the method.

4.2 Quantifying irregularities in bradykinesia

Bradykinesia is characterized by slowness and irregularities during fast voluntary movements. In this study parameters are sought, and found, that describe the slowness of the movement. However the second aspect of bradykinesia has got no attention: the irregularities of the movement. Especially Parkinsonian patients show a slow ‘saw tooth’ movement pattern. These irregularities seem an important feature of bradykinesia and in future studies we will try the parameterize these irregularities.

4.3 What causes bradykinesia?

Although a clear picture exists of how bradykinesia presents itself, little is known about its origin. Often a lack of neuromuscular control is mentioned. However many pathways exist between the planning of a movement in the CNS and the actual execution of the movement. This study only focused on quantifying the slowness of bradykinesia. Incorporation of the irregularities going with bradykinesia into the analysis may lead to a better classification of bradykinesia and such may contribute to a better understanding of the underlying pathophysiology.

This study is part of TREND (Trauma RElated Neuronal Dysfunction), a consortium that integrates research on epidemiology, assessment technology, pharmacotherapeutics, biomarkers and genetics on Complex Regional Pain Syndrome type 1. The consortium aims to develop concepts on disease mechanisms that occur in response to tissue injury, its assessment and treatment. TREND is supported by a government grant (BSIK03016).

References

1. Paice E. (1995). Reflex sympathetic dystrophy. *BMJ*, **310**, 1645-1648.
2. Ribbers G.; Geurts A.C.H.; Mulder T. (1995). The reflex sympathetic dystrophy syndrome: a review with special reference to chronic pain and motor impairments. *Int J Rehabil Res*, **18**, 277-295.

3. Van Hilten J.J.; Van de Beek W.J.T.; Hoff J.I.; Voormolen J.H.; Delhaas E.M. (2000). Intrathecal Baclofen for the Treatment of Dystonia in Patients with Reflex Sympathetic Dystrophy. *N Engl J Med*, **343**, 625-630.

Computer assisted skin videothermography is a highly sensitive tool in the diagnosis of complex regional pain syndrome type I

S.P.Niehof, F.J.P.M. Huygen , J.Klein, F.J.Zijlstra
Pain Treatment Centre, Erasmus MC, Rotterdam, The Netherlands

Abstract

The use of thermography in the diagnosis and evaluation of complex regional pain syndrome type 1 (CRPS1) is based on the presence of temperature asymmetry between the involved area of the extremity and the corresponding area of the uninvolved extremity. The interpretation of thermographic images is, however, subjective and not validated for routine use. The aim of the present study was to develop a sensitive, specific and reproducible mathematical model based on the results of computer-assisted infrared thermography in patients with early stage CRPS1 in one hand. Temperature asymmetry between the involved and uninvolved hand was calculated using three different methods, namely, the temperature asymmetry factor, the ratio and the average temperature difference. The performance of the mathematical model based on the Receiver Operater Curve (ROC) curve was excellent. The area under the curve (AUC) was 0.97, the p-value was <0.001, the sensitivity 92% and specificity 94%. This objective tool could be used as an additive diagnostic tool in the diagnosis of CRPS.

Keywords

Complex regional pain syndrome type 1, videothermography, diagnosis.

1 Introduction

The complex regional pain syndrome, type 1 (CRPS1), is a chronic disease characterised by severe and constant burning pain, pathological changes in bone and skin, excessive sweating, tissue swelling and allodynia. The syndrome is thought to be a nerve disorder that occurs in patients after trauma or surgery at the site of an injury, most often presented in the extremities. It is established that damaged nerves of the sympathetic nervous system responsible for blood flow and temperature, play an important role in the development of CRPS1[5]. Therefore CRPS1 seems to resemble neurogenic inflammation, sympathetic dysfunction, or a possible interaction between both [2]. Thermographic imaging used to measure skin temperature, which reflects vasomotor activity[13] and inflammation, may therefore be a valuable tool in the diagnosis and monitoring of CRPS1 in the acute stage of the disease.

The main aim of the present study was to assess the sensitivity and specificity of computer-assisted infrared thermography in patients with early stage CRPS1 in one hand. This paper describes a reliable computer-assisted method of collecting thermographic data from complete extremities that eliminates subjective biases and is more suitable then tympanic temperature measurement of spotted regions in the diagnosis of CRPS1.

2 Method

2.1 Patients and controls

The study was approved by the medical ethical committee of the Erasmus MC and all patients gave written informed consent. The study included 18 patients, 16 women and 2 men; mean age 44.8, range 22-68 years diagnosed with unilateral CRPS1 in the hand. CRPS1 was diagnosed according to the criteria defined by Bruehl et al.[3]. Thirteen healthy volunteers with no history of neurotrauma or vascular disease, 10 women and 3 men; mean age 33.3, range 23 –53 years served as controls.

2.2 Measurement with thermography

Skin temperature of both hands was registered with a computer-assisted infrared thermograph (ThermaCAM SC2000, Flir Systems, Berchem, Belgium).

The thermal sensitivity is 0.05 °C at 30 °C, the spectral range 7.5 to 13 µm, and the built-in digital video has 320x240 pixels (total 76800 pixels). Data were obtained through a high speed (50 Hz) analysis and recording system (ThermaCAM Researcher 2001 HS, Berchem, Belgium). coupled with a desktop-PC. The thermograph camera produces a matrix of temperature values, each temperature value represents a pixel in the image measured. The resolution on the hand is 0.8x0.8 mm^2. To obtain only those pixels that represent the hand, the data are filtered by a threshold. For further analyses a frequency table is calculated. The classes consist of temperatures with an interval of 0.1°C. The temperature in both hands was measured in a predefined position with the aid of a Plexiglas curved frame with positioning points between digits 1 and 2, and between digits 3 and 4. The thermal emissive factor of the skin was predefined to be 0.98.

2.3 Calculation methods

For calculation of the difference between the involved and non-involved temperatures of the hands, the asymmetry factor, the ratio factor and the average calculations used. For the asymmetry factor the correlation was calculated between the histograms of the non-involved and involved hand. The ratio factor is calculated by summing the multiplication of each class with its frequency for both the uninvolved and involved side. The two obtained numbers then divided. The average calculation was performed by subtracting the mean temperature calculation of each hand. The background substraction, threshold filtering, and the calculation methods where programmed in Matlab (Matlab Release 12)

2.4 Statistical analysis

The discriminating power of the different calculation methods mentioned above was analyzed by means of a receiver operating curve (ROC) using SPSS (SPSS 11.5). This method calculates the sensitivity and specificity using a range of cut points. The cutoff points are chosen in the range of the lowest and highest value obtained from each calculation method. Thus for each cut point, for each method, a sensitivity and specificity can be calculated. The discriminating power is predicted by the AUC (Area Under the Curve)[12] of the ROC. The AUC can range from 0 till 1, with an AUC greater then 0.80 being considered as significant.

3 Results

In Figure 1 a representative thermographic picture is shown from the involved (see Figure 1a) and uninvolved (see Figure1b) hand in a CRPS1 patient. The difference in temperature is clearly visible.

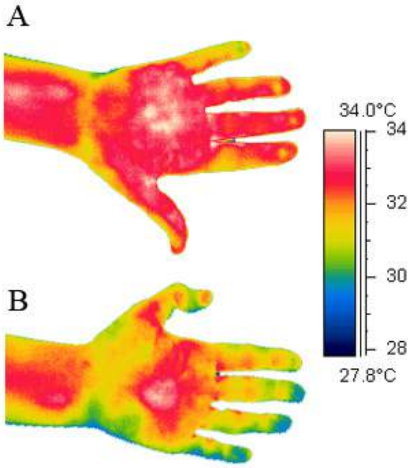


Figure 1. *Thermographic pictures of involved and non-involved hand of a CRPS1 patient.*

On calculation of the histograms derived from the pictures in figure 1 the difference in temperature profile has been made visible by means of the different distribution of temperature (see Figure 2).

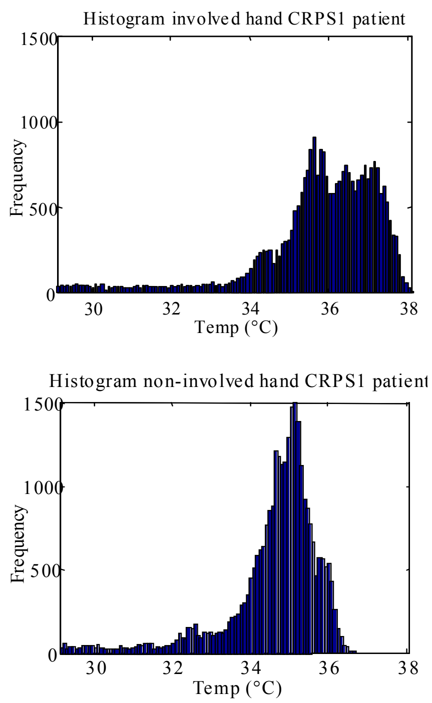


Figure 2. *Temperature histogram of the involved and non-involved hand of a CRPS1 patient.*

The correlation between the histograms in Figure 2 was 0.47. The ratio was 0.61 and the difference in average temperature 0.68°C. Results from calculation based on the thermographic temperature data in the whole group are summarized in Table 1, in which the differences between the groups and standard deviation are shown.

Table 1*. Group comparison of calculations on thermographic temperature data.*

	Controls (N=13)	**Patients (N=18)**
Asymmetry factor	0.91 ± 0.01	0.45 ± 0.01**
Ratio	0.9 ± 0.3	1.0 ± 1.0*
ΔAverage temperature (°C)	0.3 ± 0.1	1.4 ± 1.3

*significant at $p<0.05$ ** significant at $p<0.001$

Using the results from the calculation methods using the thermographic data (see Table1) in ROC analysis, resulted in values of sensitivity, specificity and predictive power. For the Asymmetry factor a sensitivity of 92 % was calculated and a specificity of 94%, the AUC of the ROC curve was 0.97. By applying the ratio method on the thermographic data a sensitivity of 83% and a specificity of 70% was calculated, with an AUC of 0.78. The average calculation did not show a significant discriminating power. These results have been extensively described in a recent publication[6]

4 Discussion

In 1998 Jones already called for a reappraisal of the use of infrared thermal imaging of the skin. Until now it was well known that abnormalities such as malignancies, inflammation and infection cause localised increases in temperature, which are shown as hot spots or as asymmetrical patterns on an infrared thermograph. If thermographs are made under controlled conditions, they may serve to diagnose certain conditions and to monitor the reaction of a patient's physiology to thermal and other stresses[7].

CRPS1 is associated with complex disturbances of the sympathetic nervous system, which also controls the microcirculation of the skin. Circulatory skin changes are in turn reflected by altered superficial thermal emission, which can be reliably imaged by thermography. Such thermographic findings often appear before skin or roentgenographic changes become manifest and may therefore be helpful in early diagnosis[10]. The usefulness of this technique for early detection of CRPS1 has been demonstrated[8].

Although thermography is still believed to be non-specific, one of the most novel advantages of infrared videothermography is the readily available display which exactly detects so far unrecognised problems that affect a patient's physiology. The computerised regression of the number of pixels per interval of 0.1 °C easily displays whether or not symmetry exists between both hands. Using computerised telethermography in healthy individuals, the skin temperature difference between both sides of the body was only 0.24 ± 0.073 °C. In contrast, in patients with peripheral nerve injury, the temperature of the skin innervated by the damaged nerve deviated on average by 1.55 °C[1]. In our population of 18 patients we found a mean temperature difference between both hands, as determined by videothermography, of 1.4 ± 1.3 °C (range 0.3 – 4.7ºC).

In a study by Oerlemans et al.[9] in healthy volunteers, skin temperature differences between both hands were measured with an infrared tympanic thermometer, on four fixed points, to provide insight into the relationship between dorsal and palmar temperature differences. Skin temperature of the hand differed with the site where it has been measured; differences between sites changed over time. The mean absolute differences in skin temperature between dorsal and palmar aspects of the hands were similar (0.30 ºC and 0.25 ºC, respectively) and were comparable with differences found by Uematsu[13] and colleagues. However, in the present study, the temperature differences between involved and non-involved hands were significantly greater. This could be due to the disadvantages of contact thermography and infrared tympanic thermometry which are limited by the number of spots which can be monitored following a standardised matrix protocol[11]. Neither the surface area of a certain skin temperature nor the temperature range are considered using these methods.

In 1987 Sherman et al. reported that the stability and symmetry of thermographic patterns over time among both healthy subjects and subjects whose pain remained at the same intensity across several recordings were found to be both high and consistent. They found thermography to be an excellent tool for monitoring changes in pain related to variations in near surface blood flow and found a good relationship between changes in pain intensity and changes in symmetry of heat patterns. Skin temperature measurements at all finger tips under resting conditions and continuously monitored during controlled modulation of sympathetic activity, have been described recently in CRPS1 patients, in patients with painful limbs of other origin, and in healthy individuals[14]. The results showed only minor skin temperature asymmetries between both limbs under resting conditions in most CRPS1 patients. During controlled thermoregulation, however, temperature differences between both sides increased dynamically in CRPS1 patients. Although measurements were only performed at finger tips, the authors concluded that skin temperature differences in the distal limbs are capable of reliably distinguishing CRPS1 from other extremity pain syndromes. In another study, 185 patients considered to have CRPS1 were subjected to thermal stress, after which temperature patterns were evaluated and the probability of CRPS1 was compared with clinical diagnostic criteria[4]. Based on clinical criteria and an estimated 50% prior probability, the positive predictive value was 90% and the negative predictive value 94%. Using the asymmetry factor derived from regression curves of temperature-related pixels of left and right sided hands, we also found a positive predictive value of 92% (ROC value 0.923). The value of thermographic as a tool in screening of CRPS1 patients is currently being studied. Static left – right differences in temperature and in deviations in temperature/flow recovery after disturbances are measured using thermography, as described here, and laser Doppler flow. These disturbances are focused on the sympathetic system by modulating skin temperature and measured over time.

This study was financially supported by the Algesiological Research Foundation, Erasmus MC, Rotterdam. The authors thank Dirk Stronks and Nevin Rhamdhani for their technical assistance, and Laraine-Visser Isles for editorial assistance.

References:

1. Baron, R., A. Binder, et al.,2002. *[Complex regional pain syndrome. Reflex sympathetic dystrophy and causalgia]*. Nervenarzt, **73**(4): p. 305-18; quiz 319.
2. Birklein, F., M. Schmelz, et al.,2001. *The important role of neuropeptides in complex regional pain syndrome.* Neurology, **57**(12): p. 2179-84.
3. Bruehl, S., R.N. Harden, et al.,1999. *External validation of IASP diagnostic criteria for Complex Regional Pain Syndrome and proposed research diagnostic criteria. International Association for the Study of Pain.* Pain, **81**(1-2): p. 147-54.
4. Gulevich, S.J., T.D. Conwell, et al.,1997. *Stress infrared telethermography is useful in the diagnosis of complex regional pain syndrome, type I (formerly reflex sympathetic dystrophy).* Clin J Pain, **13**(1): p. 50-9.
5. Huygen, F.J., A.G. de Bruijn, et al.,2001. *Neuroimmune alterations in the complex regional pain syndrome.* Eur J Pharmacol, **429**(1-3): p. 101-13.
6. Huygen, F.J., S. Niehof, et al.,2004. *Computer-assisted skin videothermography is a highly sensitive quality tool in the diagnosis and monitoring of complex regional pain syndrome type I.* Eur J Appl Physiol, **91**(5-6): p. 516-24.
7. Jones, B.F.,1998. *A reappraisal of the use of infrared thermal image analysis in medicine.* IEEE Trans Med Imaging, **17**(6): p. 1019-27.
8. Karstetter, K.W. and R.A. Sherman,1991. *Use of thermography for initial detection of early reflex sympathetic dystrophy.* J Am Podiatr Med Assoc, **81**(4): p. 198-205.
9. Oerlemans, H.M., M.J. Graff, et al.,1999. *Reliability and normal values for measuring the skin temperature of the hand with an infrared tympanic thermometer: a pilot study.* J Hand Ther, **12**(4): p. 284-90.
10. Pochaczevsky, R.,1987. *Thermography in posttraumatic pain.* Am J Sports Med, **15**(3): p. 243-50.
11. Sherman, R.A., A.L. Woerman, et al.,1996. *Comparative effectiveness of videothermography, contact thermography, and infrared beam thermography for scanning relative skin temperature.* J Rehabil Res Dev, **33**(4): p. 377-86.
12. Swaving, M., H. van Houwelingen, et al.,1996. *Statistical comparison of ROC curves from multiple readers.* Med Decis Making, **16**(2): p. 143-52.
13. Uematsu, S.,1985. *Thermographic imaging of cutaneous sensory segment in patients with peripheral nerve injury. Skin-temperature stability between sides of the body.* J Neurosurg, **62**(5): p. 716-20.
14. Wasner, G., J. Schattschneider, et al.,2002. *Skin temperature side differences--a diagnostic tool for CRPS?* Pain, **98**(1-2): p. 19-26.

Quantification of Spinal Feedback Gains

E. de Vlugt[1], A. C. Schouten[1], F. C. T. van der Helm[1,2]
[1] *Mechanical Engineering, Delft University of Technology, Delft, The Netherlands*
[2] *Mechanical Engineering, University of Twente, Enschede, The Netherlands*

Abstract
Feedback loops are very important for the motion control behavior of human beings. Feedback loops consist of the Central Nervous System (CNS) as the controller, muscles as actuators, position, velocity and force sensors in the muscles and muscle tendons and at last the skeleton as the linkage system. During postural tasks the feedback system is very effective to counteract random perturbations that disturb away from the reference position. To quantify those feedback gains, robot manipulators have been used at the Delft University of Technology for imposing force perturbations to the hand, thereby activating the closed loop feedback pathways. By recording signals from inside the control loop, the dynamic behavior of the components of the feedback loop can be estimated using advanced closed-loop identification techniques. The control settings of the CNS, i.e. the so-called spinal reflex gains, can be estimated based on dynamic transfer functions. These reflex gains appeared to be modulated by healthy subjects, depending on the frequency content of the perturbation signal and the impedance of the environment. It is shown that patients with neurological disorders have problems in modulating their reflexive feedback gains.

Keywords
Posture, Spinal Reflex, Force Disturbances, Identification

1 Introduction
Skeletal muscles can be activated by voluntary commands and by peripheral sensory feedback. There has been numerous studies dedicated to the physiology of reflexes. As it is known that the most important reflexes originate from the muscle spindles and Golgi tendon organs, their role during natural movement tasks is still far from clear.
This paper describes recent experimental approaches to identify the gains of the reflexive feedback system during posture maintenance of the human arm. The crux is the application of random force disturbances allowing a natural posture task facilitating functional analyses. In this case, resistance from muscular effort is effective to maintain posture. Differences in reflexive feedback behavior were found in the pathological case (CRPS and Parkinson's Disease (PD)) compared to normal and will be discussed. A new two DOF (Degree Of Freedom) robot manipulator made it even possible to quantify reflex gains of multiple joints simultaneously. This paper gives a brief overview of recently performed experiments.

2 Methods
Continuous force disturbances were applied consisting of a number of sinusoids with random mutual phase shifts. The resulting force disturbance was unpredictable to eliminated anticipation. The subjects were instructed to minimize the displacements of their hand as good as possible. Figure 1 shows a 1DOF experimental setup for perturbing the shoulder joint.

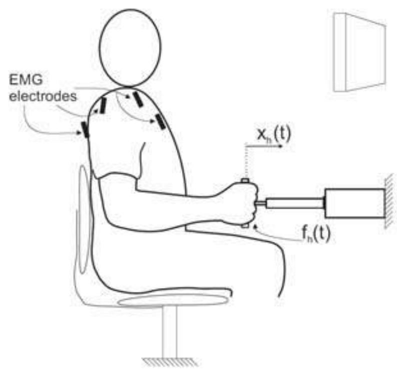


Figure 1. *Example of an (1DOF) experimental setup. The subject is attached to a handle of the manipulator that acts as a virtual admittance (i.e. mass-damped-spring system) while continuous random forces were applied.*

From the measured hand reaction force, handle position and EMG signals, FRFs (Frequency Response Functions) were estimated representing important input-output relationships of the arm (at endpoint level). These were the *mechanical admittance* (force to position) and *reflexive impedance* (position to EMG) as displayed in Figure 2.

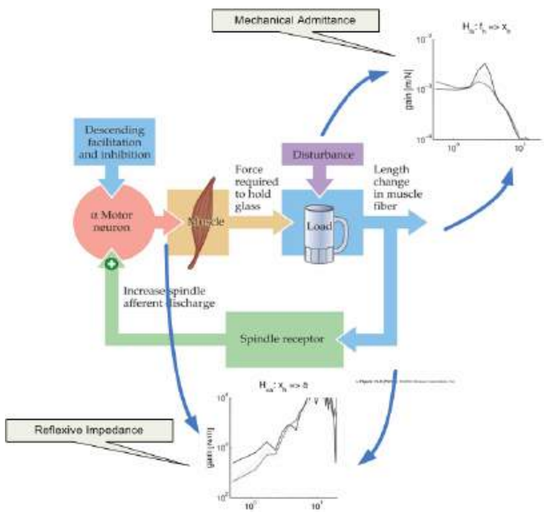


Figure 2*. Schematic representation of the feedback mechanism and the estimated FRFs (see Text).*

Model parameters were found from fitting the modeled FRFs onto the estimated FRFs using standard optimization routines. The models were all linear and expressed in the frequency domain as function of the Laplace operator and were not given here for concision. The following experimental conditions that have been carried out will be presented here:

1. 1DOF shoulder, NB1, NB2 (CRPS)
2. 1DOF shoulder, External Damping (PD)
3. 1DOF wrist, NB1, NB2, External Damping (PD)
4. 3DOF (shoulder-elbow-wrist) External damping, Configuration

NB1 (Narrowband type 1) and NB2 disturbances are schematically explained in Figure 3. In this paper, only the typical results are presented. For concision, the results from the wrist study were omitted. Furthermore, since we are interested in neural control we mainly report on the reflex gains instead of the intrinsic muscle visco-elasticity.

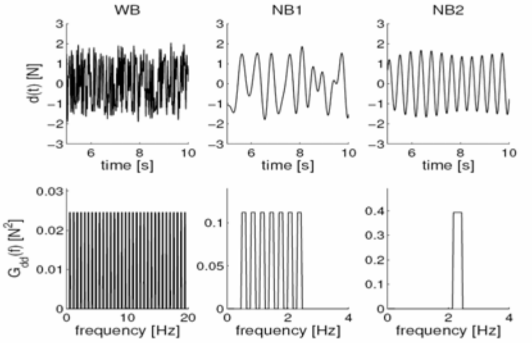


Figure 3*. Examples of force disturbance signals as applied at the hand. Wide bandwidth (WB) containing frequencies from zero to 20 Hz (left column), narrow bandwidth type one (NB1) having a fixed lowest frequency and variable highest frequency (middle column) and narrow bandwidth type two (NB2) having a fixed small bandwidth with varying center frequency. Bottom row shows the corresponding power spectra.*

External damping was increased to increase the stability margins thereby provoking larger reflex gains without making the overall system instable. Values ranged from zero to several times the intrinsic joint viscosity.

3 Results

Figure 4 shows the estimated FRFs of the mechanical admittance and the reflexive impedance of the shoulder for NB1 type disturbances. Clearly, the gain of the admittance decreases with decreasing bandwidth of disturbance. This means that the overall rotational stiffness was larger for small bandwidth disturbances. The opposite was found for the impedance: small bandwidth showed enhanced reflex gain. High coherence values indicated high linear behavior between force disturbance and hand position and EMG respectively.

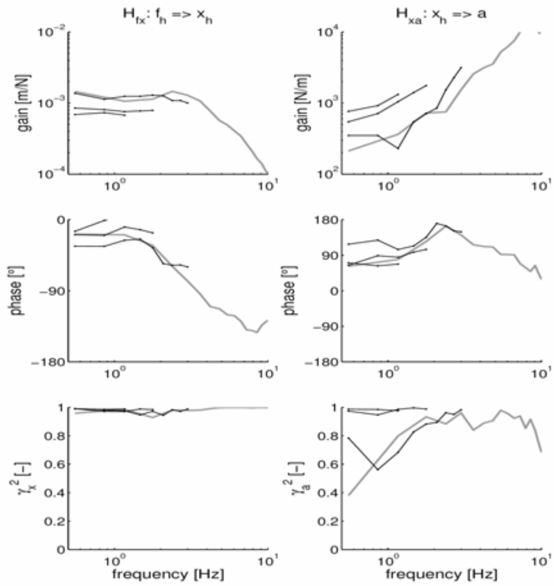


Figure 4*. Estimated FRFs (gain, phase and coherence) for different NB1 disturbances (dark lines) applied to the shoulder. Mechanical admittance (left column) and reflexive impedance (right column). Light grey lines correspond to WB disturbance.*

Figure 5 shows the results for NB2 type disturbances. Again the admittance was smaller for low frequency disturbances. However, the phase did increase less with frequency compared to NB1 disturbances.

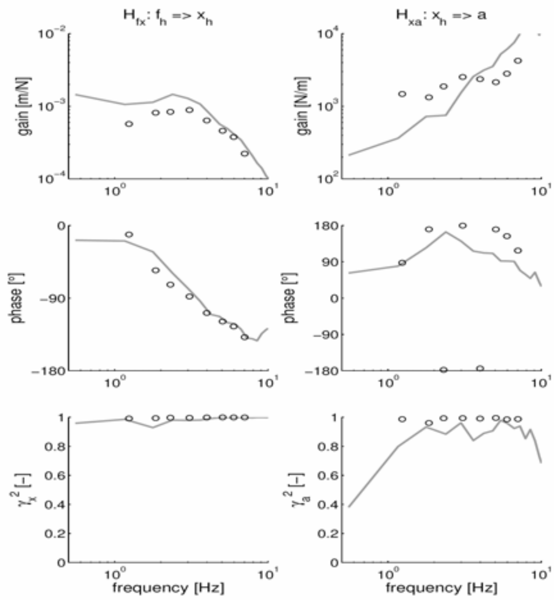


Figure 5*. Estimated FRFs (gain, phase, coherence) for different NB2 disturbances (circles). Mechanical admittance (left column) and reflexive impedance (right column). Solid light grey lines correspond to WB disturbance.*

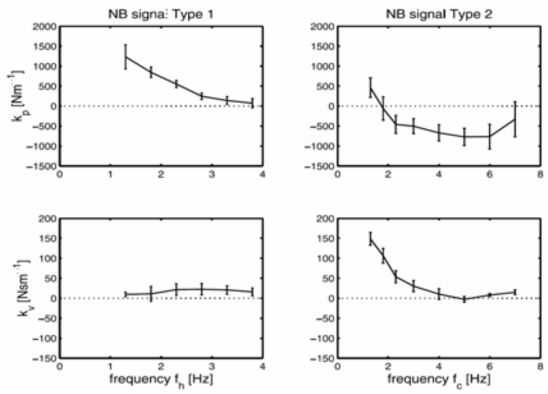


Figure 6*. Estimated reflex gains from model fits onto the FRFs. Length feedback gain (k_p) and velocity feedback gain (k_v).*

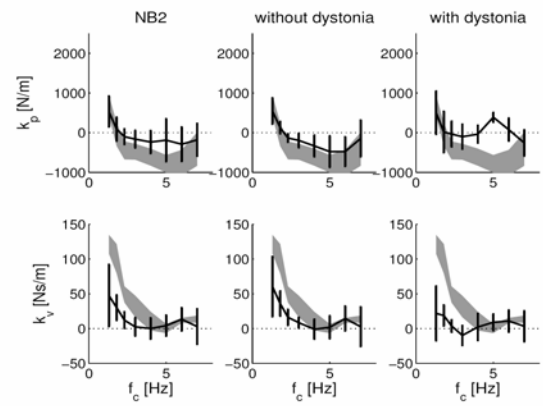


Figure 7*. Estimated reflex gains for CRPS (left), divided into lack of tonic dystonia (middle) and presence of dystonia (right). Shaded areas denote the mean ± SD of healthy subjects.*

Figure 6 summarizes the reflex gain modulation with frequency content of both NB1 and NB2 disturbances. Generally, the gains decrease with frequency. Markedly, Length feedback became negative during NB2. Figure 7 shows the comparable results for CRPS patients for NB2 where the differences were significant for the velocity feedback gain at lower frequencies and around 5 Hz for CRPS with tonic dystonia.

The estimated FRFs during application of different external damping of the robot are shown in Figure 8. Generally, we found an increased peak in the admittance with increasing external damping, that is around 3 Hz (the eigenfrequency of the arm for this configuration).

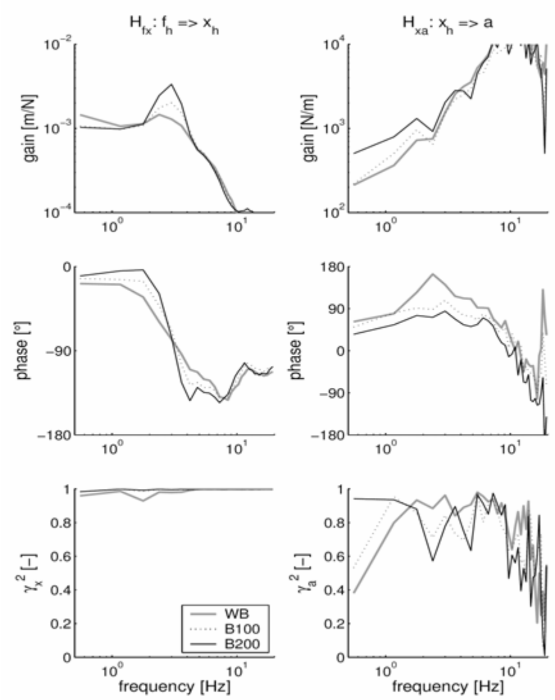


***Figure 8**. FRFs of the mechanical admittance (left) and reflexive impedance (right) for different external damping. Force perturbations were all type WB.*

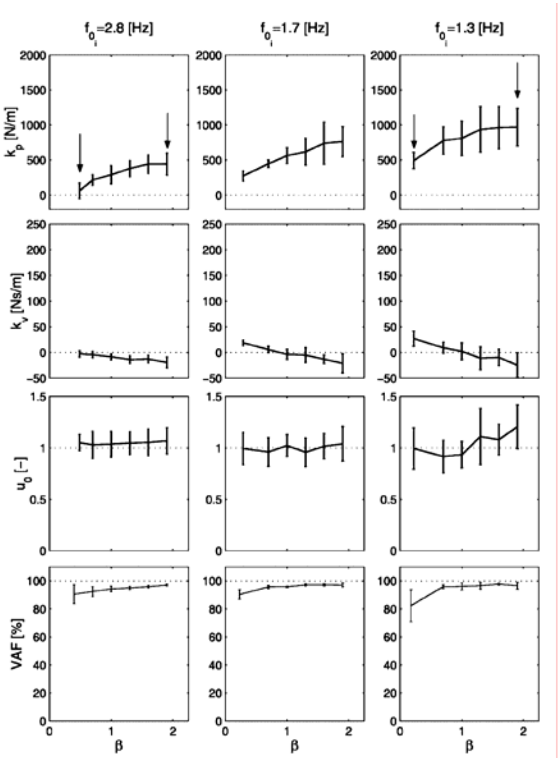


***Figure 9.** Reflex gains (average of five subjects) for increasing external damping (upper two rows). U_0 is the mean estimated cocontraction level. VAF scores of the model (bottom row).*

Figure 9 summarizes the results from the damping study. Position feedback gain increased while velocity feedback slightly decreased. The estimated co-contraction level (represented by the parameter u_0) was found to be rather constant and coincides with mean EMG (not shown).

In another study, a linearized three joint model was fitted onto multivariable FRFs (MFRFs) describing the arm admittance in a horizontal plane (Fig. 9). In this study, EMG was not yet incorporated for quantification of model parameters.

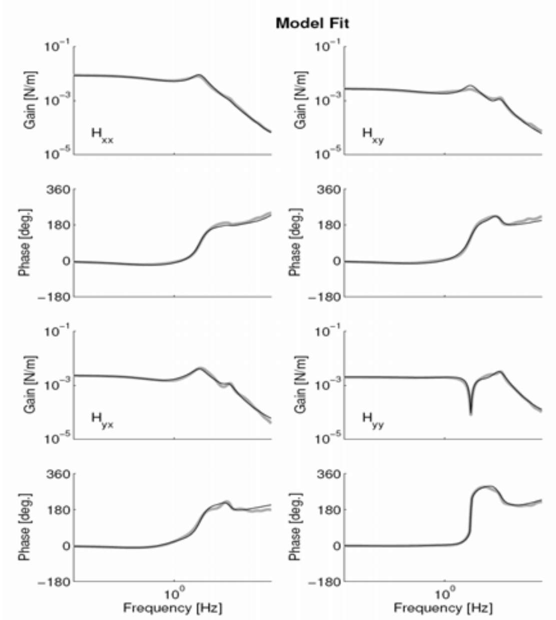


***Figure 9**. Estimated multivariable FRFs (light grey) and a fitted three joint model (dark) for a typical subject.*

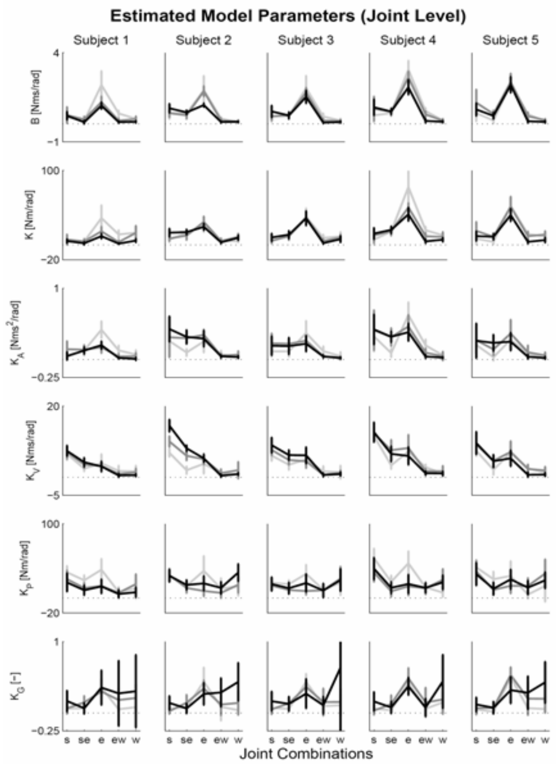


***Figure 10**. Intrinsic and reflexive model parameters from fitting onto estimated multivariable FRFs (MFRFs). Intrinsic damping (B), stiffness (K), acceleration feedback (K_A), velocity feedback (K_V), position feedback (K_P) and force feedback (K_G). Joint combinations on the horizontal axis: single joint shoulder (s), two joint shoulder to elbow (se), single joint elbow (s), two joint elbow to wrist (ew) and single joint wrist (w). Hand positions: Left (black lines), Central (dark grey lines) and Right (light grey lines).*

Figure 10 shows the 3DOF model parameters. There was large variation of the parameters over the different joints and joint combinations. Generally, shoulder intrinsic visco-elasticity is smaller than for the elbow. On the contrary, shoulder reflexes were larger. Contrary to the 1DOF shoulder studies, consistent parameter modulation with external damping and configuration was not seen.

4 Discussion

In the past, the experimental approach was to forcibly move a joint by a prescribed trajectory and measure the resulting force and EMG signals. Under such an experimental paradigm, sensory input to the CNS could not affect the movement but only the contact forces at the interface with the apparatus. To avoid such an unnatural approach, we applied continuous force disturbances while subjects had to minimize the position deviations that resulted from these disturbances. In fact, the task was a disturbances rejection tasks such as one is regularly performing e.g. holding a steering wheel during lane keeping on irregular roads, positioning a drilling machine, or body posture maintenance.

Application of force disturbances involve interaction where the subjected limb operates in a closed control loop with its mechanical environment. Force controlled robots and closed loop identification techniques were developed and applied to estimate the human neuromuscular properties during posture tasks [3,5].

The results showed that reflex gains increased when the frequency of the force disturbance decreased. We have demonstrated by model optimization that this adaptive behavior is close to optimal in the sense that reflexes increase the mechanical resistance without causing oscillations that typically can arise in delayed feedback systems [1]. When external damping was increased, higher feedback gains can be used without causing such a risk for instability, like those found experimentally (Fig. 9). High position reflex gains for the damping protocol were also found to be near optimal [2]. From the optimizations however, the velocity feedback gains were predicted larger than experimentally found. Perhaps this discrepancy is due to nonlinear muscle spindle sensitivity to stretch and stretch rate, such as a “winner takes all principle” at the spindle output. Modulation of reflex gains was also supported by the variation of the estimated reflexive impedance FRFs (Figs 4 and 8).

Tonic dystonia is characterized by an increase in muscle tone at the site of an injury. Until now there is some evidence that directs to impairment of inhibitory interneuronal circuits at the level of the brainstem or spinal cord. The disability of CRPS patients with dystonia to switch the sign of the feedback loop to negative (Fig. 7) supports the idea of reduced inhibitory control [4].

The 3DOF study showed that reflex gains for the wrist, elbow and shoulder could be quantified simultaneously. A striking feature was that intrinsic muscle visco-elasticity appeared to be substantially smaller compared to the 1DOF studies while reflexive feedback of the shoulder joint was relatively large compared to those of the elbow. Because the distance from shoulder to the hand was larger than in the 1DOF study and the moment arms of shoulder muscles were smaller in the multijoint study, we concluded that this was due to decreased effectiveness of the shoulder to the task performance (at the hand). Therefore we hypothesized that shoulder reflexes compensated for this so called “intrinsic loss” [6].

5 Conclusions

Important parameters of the neuromuscular system can now be quantified using force disturbances and closed loop identification techniques. For this purpose, powerful robotic manipulators have been developed for the application of single and multiple joint studies. Reflex gain modulation of the CNS can now be studied during natural posture tasks. The gains can directly be interpreted on a functional basis in terms of stability and performance. This experimental paradigm is of great value to the fundamental understanding of neurological control mechanisms and revealing the effect of many diseases on human motion control.

References

1. De Vlugt, E.; Van der Helm F.C.T.; Schouten A.C.; Brouwn G.G. (2001) Analysis of the reflexive feedback control loop during posture maintenance. *Biol Cybern* 84: 133-141
2. De Vlugt E.; Schouten A.C.; Van der Helm F.C.T. (2002) Adaptation of reflexive feedback during arm posture to different environments. *Biol Cybern* 87: 10-26
3. De Vlugt E.; Schouten A.C.; Van der Helm F.C.T.; Teerhuis P.C.; Brouwn G.G. (2003) A force-controlled planar haptic device for movement control analysis of the human arm. *J Neurosci Methods* 129: 151-168
4. Schouten A.C.; Van de Beek W.J.T.; Van Hilten J.J.; Van der Helm F.C.T. (2003) Proprioceptive reflexes in patients with Reflex Sympathetic Dystrophy. *Exp Brain Res* 151: 1-8
5. Van der Helm F.C.T.; Schouten A.C. De Vlugt E.; Brouwn G.G. (2002) Identification of intrinsic and reflexive components of human arm dynamics during postural control. *J Neurosci Methods* 119: 1-14
6. De Vlugt E.; Schouten A. C.; Van der Helm F. C. T. Quantification of intrinsic and reflexive properties during multijoint arm posture. Submitted to the *J Neurosci Methods*

Disentangling the contribution of the paretic and non-paretic leg to balance control in stroke patients

E. H.F. van Asseldonk[1], J. H. Buurke[2], F. C.T. van der Helm[1,3], H. van der Kooij[1].
[1] *Dept. of Biomechanical Engineering, University of Twente, Eschede, The Netherlands*
[2] *Roessingh Research & Development, Enschede, The Netherlands.*
[3] *Department of Mechanical Engineering, Delft University of Technology, Delft, The Netherlands*

Abstract
During stroke recovery, both physiological recovery of the paretic leg and adaptive compensation in the non-paretic leg may contribute to improved balance maintenance. Here, we examine a new approach to disentangle these different recovery mechanisms and to quantify objectively the contribution of the paretic and non-paretic leg in balance control. Balance responses were elicited by random continuous platform movements (forward-backward), and body sway were measured as well as the ground reaction forces below each foot to calculate corrective ankle torques in each leg. From these data the relative contribution of the paretic and non-paretic leg to the total amount of generated corrective torque to correct body sway. Even in patients with a fairly symmetrical weight distribution (more than 40% on the paretic leg), we observed a clear asymmetry in balance contribution in favour of the non-paretic leg. Overall, the paretic leg made a significantly ($P<0.05$) smaller balance contribution than the non-paretic leg. The presented approach allows for an objective quantification of the contribution of each leg to overall balance maintenance.

Keywords
Balance, stroke, system identification control, motor system

1 Introduction
The precise mechanisms underlying the clinical recovery of functions during the rehabilitation of stroke patients remain insufficiently understood. This is particularly true for tasks where both the non-paretic and the paretic side contribute to task execution, as occurs for example in postural control or gait. For such activities, a net improvement of task execution following an acute stroke is usually ascribed to recovery of the impaired paretic leg [1]. However, alternative explanations are possible, because secondary adaptations of the non-paretic leg can occur that may compensate for the impairments in the paretic leg. The presence of such compensatory processes is often neglected in research studies of patients with a chronic disease [2]. However, careful separation of disease recovery from compensatory strategies has obvious clinical implications, for example for the design of rehabilitation strategies.

In the field of postural control, it has hitherto proved difficult to reliably disentangle these entirely different processes. For this purpose, it is necessary to identify and quantify the contribution of each leg to overall task performance. The most straightforward way to achieve this is to determine the contribution of the generated activity (i.e., EMG activity or joint torque) in each leg to the stabilization of the centre of mass (CoM) relative to the base of support.

Theoretically speaking, improvements in task performance during rehabilitation that are accompanied by a greater contribution of the paretic leg should index that the paretic leg has recovered (parts of) its efficacy. Conversely, an increased contribution of the non-paretic leg might point towards compensatory adaptation in the non-paretic leg in order to take over part of the original function of the paretic leg.

Here, we will study the merits of this new method in determining the individual contribution of the paretic and non-paretic leg to postural control. For this purpose, we elaborated upon the original approach [3] and incorporated two ankles instead of one, so the stabilising contribution of generated corrective torque in both the paretic and the non-paretic ankle could be determined. The contribution of both legs to balance maintenance can be expressed as a fraction, much like the amount of weight distribution. The new approach was evaluated in a group of chronic stroke patients whose balance was challenged by horizontal platform translations. We hypothesized that an asymmetry in balance contribution in stroke patients is not a mere reflection of an asymmetry in weight distribution.

2 Methods

2.1 Subjects
Seven patients with hemiparesis secondary to a single and first ever unilateral stroke in the territory of the anterior or middle cerebral arteries participated. Subjects were at least one year post-stroke, had no musculoskeletal or neurological diseases in addition to stroke and were able to remain standing without help of support for at least 90 s. The experiment was approved by the local medical ethical committee. All subjects gave their written informed consent prior to the start of the experiment.

2.2 Theoretical background
The method is developed based on a commonly used model of postural control [4, 5] (see Fig 1). In this model, the human body is considered to act as an inverted pendulum, which is stabilized by corrective torques acting on the human body around the ankles.
In previous balance control studies that used the inverted pendulum model only the net corrective torque from both ankles together was considered. However, in this study we will consider the activity generated around both ankles separately as we expect that in stroke patients the sensory and motor impairments of the hemiparetic side will negatively affect its ability to generate a corrective torque. In this case, the non-paretic leg should compensate for this deficiency by generating more corrective torque. In order to quantify the contribution of the paretic and non-paretic leg in maintaining upright stance, we split up the total stabilizing mechanism into two parallel stabilizing

mechanisms: the paretic and non-paretic stabilizing mechanisms (see Fig 1).

The contribution of the paretic and non-paretic leg to postural control can be determined by quantifying the proportion of the total corrective torque generated by each separate leg in response to a deviation of upright posture. This requires that the relation between sway and the corrective torques is established, what comes down to identifying the content of the stabilizing mechanisms. The stabilizing mechanisms are dynamical, which means that the amplification and timing of their response, the corrective torque, depends on the frequency content of their input, the sway angle. In order to determine the dynamical characteristics of the stabilizing mechanisms the Frequency Response Function (FRF) can be estimated by using the joint input-output approach [3]. The joint input-output approach can only be applied when external mechanical and/or sensory perturbations are applied to the balancing human and not during quiet stance. In this study horizontal platform movements in the forward-backward directions were used to perturb the balance of stroke patients, which made it possible to determine the contribution of paretic and non-paretic leg in balance control.

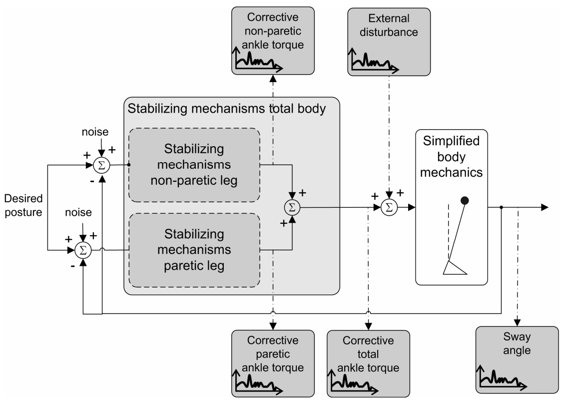


Figure 1. *A model of postural control of a stroke patient. The simplified body mechanics represent the dynamics of an inverted pendulum with the total corrective torque as input and the sway angle as output. The sway angle is compared to the desired posture, which is upright stance. The deviation of upright stance with the addition of noise serves as input for the stabilizing mechanisms. The stabilizing mechanisms represent the dynamics of the combination of active and passive feedback pathways of the concerned body(part) and generate a torque to correct for the deviation of upright stance. Noise represents the unknown noise on the reference signal, errors in the perceived sway angle and insufficiencies of the model. The graph icons denote which signals are measured or can be calculated from measured signals in this study.*

2.3 Apparatus and recordings

The participants stood with their arms folded in front of their chest or hanging at their sides on a force plate, embedded in a computer controlled 6 degrees of freedom motion platform (Caren, Motek, Amsterdam, The Netherlands). The feet were placed with their medial sides and heels against a fixed foot frame resulting in a separation of the medial sides of the heels of +/- 20 cm and a outward rotation of the feet of 9° with respective to the sagittal midline. The subjects faced a visual scene (4.80 × 3.70 m at a distance of 3.20m) consisting of a light grey background with a life-sized dummy just in front of it.

The custom made force plate consisted of four 6 degrees of freedom force sensors (ATI-Mini45-SI-580-20), mounted in a rectangular configuration on an aluminum plate. Each sensor was covered with a rectangular aluminum plate with dimension of 15 x 17.5 cm. The foot frame ensured that each foot was placed solely on the cover of two force sensors. Forces and torques of each sensor were sampled at a frequency of 360 Hz.

Reflective spherical markers were attached to heel, toe, malleolus, tibia, knee, femur, of both legs as well as the sacrum, head and shoulders. Moreover, a cluster of 3 markers was attached to the iliac crest of both legs and 3 markers were attached to the platform. The positions of the markers were recorded at a sampling rate of 120 Hz by means of a three-dimensional passive registration system, consisting of six video cameras and a control unit (Vicon Oxford Metrics, Oxford, UK).

Subjects wore a safety harness suspended from the ceiling, which prevented the subject from falling, but did not constrain the movements necessary to maintain balance or provide any support or orientation information.

2.4 Procedures

In the experiment subjects had to stand on the platform and were instructed to 'maintain their balance without moving their feet' while continuous random platform movements were applied in the forward backward direction. The platform movements consisted of a multisine signal. The multisine had a period of 34.13 s (equal to 2^{12} = 4096 samples at a sample rate of 120 Hz) and contained power in the range of 0.06-2.37 Hz. The amplitudes in the power spectrum at each frequency decreased logarithmically with logarithmically increasing frequency. The power of the signal at a given amplitude was optimized by crest optimization. The perturbation trials lasted 90 seconds, in which the above-described multisine was repeated time after time.

Before data recording, the disturbance was presented to the subject in order to let them get acquainted to the signal and to determine the amplitude of the disturbance during the recordings. The amplitude for each subject was set to the maximal amplitude of which the patient and an accompanying physical therapist were confident that the patient could withstand it for 90 seconds. This amplitude was used in three trials of 90 s each in which the response of the subjects was measured.

2.5 Data analysis

The signals denoted with a graph icon in Figure 1 were calculated from the measured data. From the recorded marker positions, the position of the CoM was calculated [6]. The length of the pendulum (l_{CoM}) was determined from the static trial as the average distance in the sagittal plane from the ankle to the CoM. The sway angle (θ) was calculated from the length of the pendulum and the horizontal distance from the CoM to the mean position of the ankles.

The measured forces and torques from the four sensors were first resampled to a frequency of 120 Hz and subsequently low pass filtered with a second order recursive Butterworth filter with a cut-off frequency of 6 Hz. Subsequently the forces and torques were corrected for the inertia and mass of the top cover. The corrected forces and torques from both sensors below each foot were used to calculate the vertical and horizontal forces below each foot and their point of application, the CoP. The ankle torque in the paretic (T_{ank-P}) and the non paretic leg (T_{ank-NP}) were calculated with inverse dynamics [6]. The total generated corrective torque (T_{ank-T}) is the sum of the ankle torques at both ankles.

The vertical forces below both feet were also used to calculate the static (SW) and dynamic weight-bearing (DW) asymmetry by dividing the average vertical force below the paretic foot by the average of the sum of the vertical forces below both feet during static and the perturbation trials, respectively.

The perturbation signal of 90 s consisted of approximately 2.5 cycles of the multisine in each of the three trials. From each trial's perturbation (dext) and response (θ, ,T_{ank-T}, T_{ank-P} and T_{ank-NP}) trajectories 2 synchronized, succeeding, complete cycles (2^{12}=4096 samples≈34.13 s) were extracted. Subsequently the perturbation and response cycles were decomposed in their sinusoidal component parts by using Fast Fourier Transform (FFT). From the obtained Fourier coefficients the following Cross Spectral Densities (CSD) were calculated and averaged over the 6 (2 cycles x 3 trials) cycles: CSD of the external disturbance torque with each of the ankle torques and CSD of the external disturbance torque with the body sway. A joint input-output approach [3] was adopted to obtain the estimate of FRF of the stabilizing mechanisms. The FRF of the total stabilizing mechanisms (C_{PA-T}), the paretic stabilizing mechanisms (C_{PA-P}) and the non paretic stabilizing mechanisms (C_{PA-NP}) were estimated by dividing the averaged CSD of the external disturbance torque and the body sway with the CSD of the external disturbance torque and the corresponding corrective ankle torque.

The frequency dependency of the characteristics of the stabilizing mechanisms can be shown by calculating the gain and phase for each frequency from the estimated FRFs. The gain represents the ratio of the amplitude of the corrective torque to the amplitude of the sway angle and the phase gives the relative timing of the corrective torque with respect to the sway angle.

The gain of the stabilizing mechanisms of the total body indicates the total amount of corrective torque generated in response to a deviation of upright stance at each frequency. In order to determine the contribution of each leg to the generation of the total corrective torque, the contribution of the gain and phase of each leg to the gain of the total body was calculated. The total stabilizing mechanisms (C_{PA-T}) is the sum of the stabilizing mechanisms of the paretic (C_{PA-P}) and non-paretic leg (C_{PA-NP}). In the imaginary plane, C_{PA-T} is the vectorial sum of C_{PA-P} and C_{PA-NP} (see Fig (2)). The contribution of the paretic and non-paretic stabilizing mechanisms to the gain of the total stabilizing mechanisms was determined by projecting the vector of both stabilizing mechanisms on the vector of the total stabilizing mechanism. Division of the result by the total gain led to the contribution of each of the stabilizing mechanisms to the total expressed as a proportion.

In the foregoing, the contribution was calculated for each frequency, by averaging the contribution over the different frequencies, the Dynamic Balance Contribution (DBC) of the paretic and non-paretic leg to the overall balance is determined.

2.6 Statistical analysis

A Wilcoxon signed rank test was used to compare the paretic DBC with the non-paretic DBC, the paretic DBC with the paretic DW, and the paretic SW with the paretic DW.

3 Results

3.1 Frequency Response Functions

The gain profiles for all patients showed more or less the same pattern, with a gain that is lowest for the smallest frequencies and that gradually increases with higher frequencies (see Fig 2 for a typical stroke patient). The gain of the C_{PA-NP} was higher for all frequencies than the gain of the C_{PA-P}. The difference in magnitude of the gains of the paretic leg and non-paretic leg was reflected on their contribution, the paretic leg showed a much smaller contribution for all frequencies than the non-paretic leg. In Figure 2 the FRFs of the stabilizing mechanisms of a typical stroke patient are depicted.

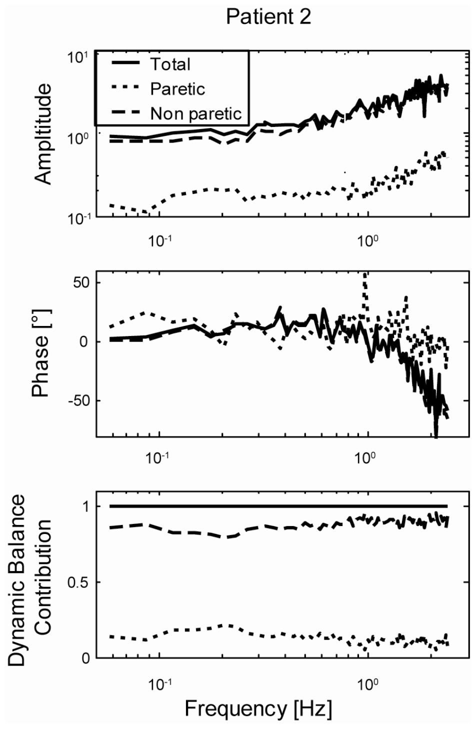


Figure 2. *Stabilizing mechanisms of a representative stoke patients. From the top to the bottom panel the gain, phase and the contribution to the total gain of the total, the paretic and non-paretic stabilizing mechanisms are depicted. The gains are scaled to the gravitational stiffness. Hence, a gain of one indicates that in response to deviations of erect stance, the generated torque by the stabilizing mechanisms was exactly equal to the torque resulting from gravity. A gain greater than one indicates that the generated corrective torque was larger than the gravitational torque at the particular frequency. Consequently, the stabilizing mechanisms "push" the body back to erect posture. A gain of less than one points to a smaller corrective torque than the gravitational torque. In this case, the stabilizing mechanisms failed to generate enough torque to compensate for the gravitational torque. Whenever this occurs for the total stabilizing mechanisms this would theoretically mean that the body falls over.*

3.2 Dynamic Balance Conribution

In order to determine the overall contribution of the paretic and non-paretic leg to balance maintenance, the contributions were averaged over all frequencies. The resulting values are shown in Figure 3, together with the static and dynamic weight distribution. The weight distribution was not significantly different when the patients were standing on the motionless platform or when they were withstanding the platform perturbations (p=0.735), indicating that patients maintained their normal posture during the perturbation trials. The group of patients could roughly be divided into two groups. The first group showed a rather symmetrical weight distribution during the perturbations (dynamic weight bearing on paretic leg>0.4, range: 0.43-0.54 weight on paretic leg), while the second group had a marked asymmetry in dynamic weight bearing (range: 0.39-0.28 weight bearing). Notwithstanding the symmetry in weight distribution, all patients except one of the first group showed a clear asymmetry in balance contribution (range: 0.11-0.45 DBC of the paretic leg). The patients of the second group had an asymmetry in their balance contribution, which was even more pronounced than the asymmetry in weight distribution (range: 0.07-0.2 DBC on the paretic leg). Overall the dynamic balance contribution of the paretic leg was significantly smaller (p=0.018) than the contribution of the non-paretic leg. Furthermore the paretic leg contributed significantly less (p=0.018) to balance than to weight bearing.

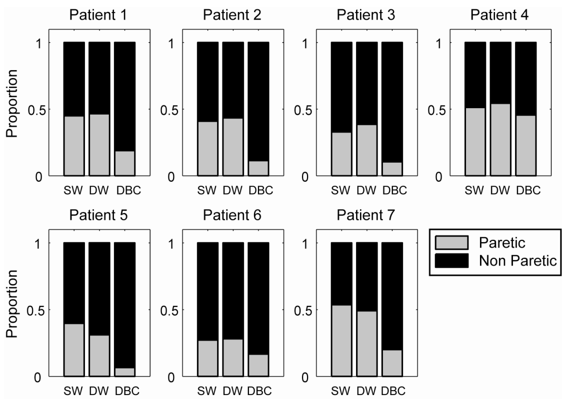


Figure 3. *Weight distribution during quiet stance (SW), weight distribution during the perturbation trials (DW), and the dynamic contribution to balance maintenance (DBC) for the paretic and the non-paretic leg for 7 patients. All values are expressed as proportions and consequently sum up to 1.*

4 Discussion and conclusion

The goal of this paper was to assess a new method to determine the contribution of the paretic and non-paretic leg to balance maintenance. For this purpose, it was necessary to relate the overall balance performance, in terms of the sway angle, to the torque generated in each of the two ankles. To achieve this, we had to consider the total balance control mechanisms and applied – for the first time –theoretical models and methods [3] to identify such stabilizing mechanisms in a patient population. The results showed that, even in patients with a fairly symmetrical weight distribution, a clear asymmetry in balance contribution could be observed in favour of the non-paretic leg. This asymmetry was even more pronounced in patients with an asymmetrical weight distribution. Overall, the paretic leg made a significantly smaller balance contribution than the non-paretic leg.

Our method assumes that all corrective torques needed to stabilize the CoM were generated at the ankle (the so-called inverted pendulum of human postural control). However, in real life the human body almost never acts like a perfect inverted pendulum. Therefore, other joints likely contributed to the maintenance of balance [7]. Such action of other joints can explain why the subjects were able to maintain their balance even when the gain of the total stabilizing mechanisms was smaller than one, indicating that less corrective ankle torque was generated than required to compensate for the gravitational torque. A gain lower than one occurred at some of the lower frequencies for the patient depicted in Figure 2 and also for some of the other patients. As the drops below zero were very minor, the stabilising contribution of joints other than the ankles seems fairly low.

In this study, the dynamic balance contribution was evaluated in a group of chronic stroke patients. Application of this method in longitudinal surveys of balance rehabilitation in stroke patients makes it possible to disentangle physiological recovery of the affected leg from compensation in the unaffected leg. Such insights will be critical for the development and evaluation of individually tailored rehabilitation strategies. Furthermore, the same method has great potential value for other patient populations, which are also characterised by a clear asymmetry in the impairments of both legs. Possible target populations include, among others, patients with Parkinson's disease [8], patients with a prosthetic leg. Furthermore

This study was supported by a Vedi grant (016027011) of the Netherlands Organisation of Scientific research.

5 References

1. Geurts, A.C.H., De Haart, M., Van Nes, I.J.W., and Duysens, J. A review of standing balance recovery from stroke. *Gait and Posture* (in press).
2. Bloem, B.R., Visser, J.E., and Carpenter, M.G. (2001). Compensatory *postural strategies in patients with longstanding balance impairment*, in *Balance at all times*, J.C.E. Van der Burg, et al., Eds. Utrecht: Digital Printing Partners BV: Utrecht. p. 51-73.
3. Van der Kooij, H., Van Asseldonk, E., and Van der Helm, F.C. (2005). Comparison of different methods to identify and quantify balance control. *Journal of Neuroscience Methods*, **145**, 175-203.
4. Peterka, R.J. (2002). Sensorimotor integration in human postural control. *J. Neurophysiol.*, **88**, 1097-118.
5. Fitzpatrick, R., Burke, D., and Gandevia, S.C. (1996). Loop gain of reflexes controlling human standing measured with the use of postural and vestibular disturbances. *J. Neurophysiol.*, **76**, 3994-4008.
6. Koopman, B., Grootenboer, H.J., and de Jongh, H.J. (1995). An inverse dynamics model for the analysis, reconstruction and prediction of bipedal walking. *J. Biomech.*, **28**, 1369-76.
7. Park, S., Horak, F.B., and Kuo, A.D. (2004). Postural feedback responses scale with biomechanical constraints in human standing. *Exp. Brain Res.*, **154**, 417-27.
8. Bloem, B.R., Hausdorff, J.M., Visser, J.E., and Giladi, N. (2004). Falls and freezing of gait in Parkinson's disease: a review of two interconnected, episodic phenomena. *Mov. Disord.*, **19**, 871-84.

Innovative use of the water maze to assess neurocognitive abilities in laboratory rodents

R. D'Hooge, H. Goddyn and Z. Callaerts-Vegh
Laboratory of Biological Psychology, University of Leuven, Belgium

The water maze was described more than twenty years ago to assess spatial learning and memory in laboratory rodents (Morris, 1981), and it has become one of the most frequently used laboratory tools in behavioral neuroscience (D'Hooge and De Deyn, 2001). Work using the water maze has shown that the integrity of the hippocampal formation is essential for spatial learning, but water maze performance and spatial learning in general actually depend on the coordinated action of different brain regions in a functionally integrated neural network (Cain and Saucier, 1996). The water maze has been used in studies on the involvement of various brain regions and the neurobiological basis of complex learning processes, but it has been instrumental in many reports on the pathophysiology and possible treatment of neurocognitive disorders as well.

Many authors continue to stress the need for methodological innovation and alternative approaches. Novel training protocols and measuring procedures have been proposed, including standardized spatial memory and delayed-matching-to-place protocols. Recent studies are addressing how spatial and contextual memory can provide a framework for remembering events. Richard Morris and associates at Edinburgh University (UK) recently developed a new paired-associate task for investigating episodic-like and semantic-like memory in animals. In this event arena, rats are trained to find a specific flavor of food in a particular location (Day et al, 2003). Learning may take place in a single trial (episodic-like), or over several trials and days (semantic-like).

Keywords

Water maze, hippocampus, cerebellum, neurocognitive disorders

Hippocampus-dependent learning

Hippocampus seems to be involved with tasks in which memories have to be acquired and retrieved during performance (Knowlton and Fanselow, 1998). Hippocampal place cells were suggested as the primary substrate of spatial memory processes. It is necessary for acquisition and retrieval of spatial information as well as for consolidation/storage (Poucet et al., 2000). Hippocampus may be dedicated to path integration, and representing and transforming ideothetic information (Whishaw et al., 1997). According to Redish and Touretzky (1998), it is involved in self-localization and route replay, which depend on the combination of hippocampal space fields and synaptic plasticity. Distal extramaze cues are important in guiding place navigation in the watermaze. Behavioral and electrophysiological studies indicate that proximal intra-maze cues may also serve as place cues. Etienne Save, Carole Parron and Bruno Poucet of the Laboratory of Neurobiology and Cognition (Université Aix-Marseille I, France) have recently shown that, although distal and proximal cues both support mapping strategies, processing these two kinds of cues involves activation of different neural circuits. Their results suggest that the processing of distal cues requires the hippocampus and the entorhinal cortex, whereas the processing of proximal landmarks involves the hippocampus and the parietal cortex. This is consistent with the idea of a dual processing system whose activation depends on the distance of landmarks. Save and colleagues argue that these two systems are mediated by distinct cortical areas but eventually converge on the hippocampus which is assumed to form an integrated spatial map.

Ales Stuchlik and Jan Bures from the Czech Academy of Sciences (Prague, Czech Republic), recently developed variants of the dry-land place avoidance test, which can be an alternative to classical water maze procedures. In their place avoidance task, an animal is taught to avoid an unmarked prohibited sector in a circular arena. They demonstrated that animals create separate and autonomous representations of the arena and room coordinate frames, when it is required to avoid a room frame-fixed sector on the stable arena. An active allothetic place avoidance (AAPA) task (a modification of the place avoidance paradigm) was demonstrated to require greater hippocampal integrity than the classical water maze. In the AAPA, animal has to separate spatial stimuli from the environment into coherent representations of the room and arena frames, and to navigate according to the room frame while ignoring the arena frame.

Role of cerebellum in water maze learning

The cerebellum receives input from brain areas involved in many aspects of water maze learning including visual cortex, superior colliculus and hippocampus, and, apart from motor control and acquisition/retention of conditioned reflexes, its specific functions may also include specific cognitive processes (Lalonde, 1997). It has been shown that mice with cerebellar damage do display impaired water maze learning (Lalonde, 1994). Petrosini et al. (1996) suggested that the role of the cerebellum in spatial learning is primarily that of controlling the procedural aspects of the task. It involved in procedures required to find an object in space, rather than knowing its spatial location. Hemicerebellectomized rats were impaired in executing exploratory behaviors and acquiring spatial information during hidden-platform acquisition training. Such rats were also unable to change from a non-spatial searching strategy acquired before surgery to a spatial strategy (Leggio et al., 1999).

Also through observation, cerebellar circuits contribute to the acquisition of spatial procedures. Laura Mandolesi, Maria G. Leggio, Francesca Federico & Laura Petrosini of the Fondazione Santa Lucia (Rome, Italy) developed a new paradigm to analyze the role of the cerebellum in observational learning of spatial tasks (Leggio et al., 2000). Observer animals observed companion actor rats performing a water maze tasks. After the observational training, observer animals were hemicerebellectomized, and were tested in the water maze task they had previously observed. Observers displayed exploration abilities closely matching the previously observed behaviors, indicating that they did learn complex navigational strategies by observation. When the cerebellar lesion preceded the observation training, complete lack of spatial observational learning was observed.

Use of the water maze in studies on the pathophysiology and treatment of neurocognitive disorders

Water maze experiments have been used in many studies on the pathophysiology and treatment of cerebrovascular and neurodegenerative disorders. Cerebrovascular disease remains one of the three most importance causes of death and morbidity in the industrialized world. Due to plastic changes in the brain outside the ischemic lesion, functional improvement does occurs in stroke survivors, which can be further enhanced by rehabilitation. Cognitive impairment is a common and highly disabling symptom after stroke. Rodent models of focal, partial and global cerebral ischemia have been described in literature, and have been tested in water maze tasks (D'Hooge and De Deyn, 2001). Focal lesions of the frontal cortex and partial cerebral ischemia impair water maze acquisition (Rogers and Hunter, 1997; Uchiyama et al., 1995). Several authors also demonstrated water maze deficits and selective neuronal damage in global ischemia models (Block, 1999). Neurons in hippocampal CA1 region are vulnerable to ischemia, and the loss of these neurons probably underlies the water maze deficits (Nunn et al., 1994; Olsen et al., 1994).

Rats reared in an enriched social environment, consisting of a large cage equipped with a variety of toys (running wheel, tunnels, rubber balls, etc.) show improved water maze learning, enrichment-induced neuroprotection, and prevention of spontaneous apoptosis (Young et al., 1999). Neonatal handling and environmental enrichment attenuate the age-associated decline in water maze performance and hippocampal atrophy (Fernández-Teruel et al., 1997). Enriched housing also reduced water maze deficits following neonatal anoxia (Iuvone et al., 1996). Rather than enhancing the proper recovery of function, these beneficial effects of environmental enrichment appear to be due to an enhancement of compensatory mechanisms (Rose et al., 1993; Rose, Al-Khamees et al., 1993).

Per Dahlqvist, Annica Rönnbäck & Tommy Olsson at Umeå University Hospital (Sweden) reversed the learning impairment after focal cerebral ischemia in rats by environmental enrichment. Environmental enrichment consists of a combination of social, cognitive and physical elements. Dahlqvist and associates demonstrated that rats housed in an enriched environment display enhanced brain plasticity and improved recovery of sensorimotor functions after experimental stroke. Middle cerebral artery occlusion resulted in major difficulties in finding the hidden platform in the water maze with increased thigmotaxic behavior. Post ischemic housing in an enriched environment significantly improved function in the water maze, but did not reduce infarct volume.

References

1. Block, F (1999). Global ischemia and behavioural deficits. *Prog Neurobiol*., **58**, 279-295.
2. Cain, D.P.; Saucier, D (1996). The neuroscience of spatial navigation: focus on behavior yields advances. *Rev. Neurosci*., **7**, 215-231.
3. D'Hooge, R; De Deyn, P.P. (2001). Applications of the Morris water maze in the study of learning and memory. *Brain Res. Rev*., **36**, 60-90.
4. Day M.; Langston, R.; Morris, R.G. (2003). Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature*, **424**, 205-209.
5. Fernandez-Teruel, A.; Escorihuela, R.M.; Castellano, B.; Gonzalez, B.; Tobena, A. (1997). Neonatal handling and environmental enrichment effects on emotionality, novelty/reward seeking, and age-related cognitive and hippocampal impairments: focus on the Roman rat lines. *Behav. Genet*., **27**, 513-526.
6. Iuvone, L.; Geloso, M.C.; Dell'Anna, E. (1996). Changes in open field behavior, spatial memory, and hippocampal parvalbumin immunoreactivity following enrichment in rats exposed to neonatal anoxia. *Exp. Neurol*., **139**, 25-33.
7. Knowlton, B.J.; Fanselow, M.S. (1998). The hippocampus, consolidation and on-line memory. *Curr. Opin. Neurobiol*., **8**, 293-296.
8. Lalonde, R. (1997). Visuospatial abilities. *Int. Rev. Neurobiol*., **41**, 191-215.
9. Lalonde, R. (1994). Cerebellar contributions to instrumental learning. *Neurosci. Biobehav. Rev*., **18**, 161-170.
10. Leggio, M.G.; Neri, P.; Graziano, A.; Mandolesi, L.; Molinari, M.; Petrosini, L. (1999). Cerebellar contribution to spatial event processing: characterization of procedural learning. *Exp. Brain Res*., **127**, 1-11.
11. Leggio, M.G.; Molinari, M.; Neri, P.; Graziano, A.; Mandolesi, L.; Petrosini, L. (2000). Representation of actions in rats: the role of cerebellum in learning spatial performances by observation. *Proc. Natl. Acad. Sci. USA*, **97**, 2320-2325.
12. Morris, R.G.M. (1981). Spatial localization does not require the presence of local cues. *Learn. Motiv*., **12**, 239-260.
13. Nunn, J.A.; LePeillet, E.; Netto, C.A.; Hodges, H.; Gray, J.A.; Meldrum, B.S. (1994). Global ischaemia: hippocampal pathology and spatial deficits in the water maze. *Behav. Brain Res*., **62**, 41-54.
14. Olsen, G.M.; Scheel-Kruger, J.; Moller, A.; Jensen, L.H. (1994). Relation of spatial learning of rats in the Morris water maze task to the number of viable CA1 neurons following four-vessel occlusion. *Behav. Neurosci*., **108**, 681-690.
15. Petrosini, L.; Molinari, M.; Dell'Anna, M.E. (1996). Cerebellar contribution to spatial event processing: Morris water maze and T-maze. *Eur. J. Neurosci*., **8**, 1882-1896.
16. Poucet, B.; Save, E.; Lenck-Santini, P.P. (2000). Sensory and memory properties of hippocampal place cells. *Rev. Neurosci*., **11**, 95-111.
17. Redish, A.D.; Touretzky, D.S. (1998). The role of the hippocampus in solving the Morris water maze. *Neural. Comput*., **10**, 73-111.
18. Rogers, D.C.; Hunter, A.J. (1997). Photothrombotic lesions of the rat cortex impair acquisition of the water maze. *Pharmacol. Biochem. Behav*., **56**, 747-754.

19. Rose, F.D.; Al-Khamees, K.; Davey, M.J.; Attree, E.A. (1993). Environmental enrichment following brain damage: an aid to recovery or compensation? *Behav. Brain Res.*, **56**, 93-100.
20. Rose, F.D.; Davey, M.J.; Attree, E.A. (1993). How does environmental enrichment aid performance following cortical injury in the rat? *Neuroreport*, **4**, 163-166.
21. Uchiyama-Tsuyuki, Y.; Kawashima, K.; Araki, H.; Otomo, S. (1995). Prostacyclin analogue TTC-909 reduces memory impairment in rats with cerebral embolism. *Pharmacol. Biochem. Behav.*, **52**, 555-559.
22. Whishaw, I.Q.; McKenna, J.E.; Maaswinkel, H. (1997). Hippocampal lesions and path integration. *Curr. Opin. Neurobiol.*, **7**, 228-234.
23. Young, D.; Lawlor, P.A.; Leone, P.; Dragunow, M.; During, M.J. (1999). Environmental enrichment inhibits spontaneous apoptosis, prevents seizures and is neuroprotective. *Nat. Med.*, **5**, 448-453.

Use of the Water Maze to Assess Recovery of Cognitive Function after Focal Cerebral Ischemia

P. Dahlqvist, A. Rönnbäck and T. Olsson
Department of Medicine, Umeå University Hospital, Umeå, Sweden

Abstract

Housing in an enriched environment (EE), i.e. group housing in a large cage with various objects, improve recovery of sensorimotor functions after experimental stroke in rats. This is not associated with reduction in brain infarct size, but instead related to plastic changes in the brain outside the ischemic lesion.

We have used the Morris water maze to investigate whether cognitive impairment after ischemic stroke in rats, by middle cerebral artery occlusion (MCAo), can also be alleviated by postischemic EE-housing.

After MCAo Male Sprague Dawley rats were randomized to EE or a deprived environment. Sensorimotor function was tested weekly for four weeks and on day 31-34 after MCAo spatial learning was tested in a classic Morris water maze set-up using four trials per day for four consecutive days, followed by a probe test on day five.

MCAo resulted in major difficulties in finding the hidden platform, with increased thigmotaxis behavior. The learning impairment was positively correlated with the brain infarct volume. Post ischemic housing in EE significantly improved function in the water maze, without reduction in infarct volume.

We conclude that the Morris water maze is a useful tool to evaluate therapies directed to alleviate cognitive impairment after focal cerebral ischemia in rats.

Keywords

Morris water maze, spatial memory, focal cerebral ischemia, recovery.

Introduction

Measuring cognitive impairment in stroke models

To investigate the pathophysiology of stroke and to evaluate potential therapies several animal models of stroke have been developed, the most common being different variants of middle cerebral artery occlusion (MCAo) in rodents. Efficacy of different therapies has mainly been evaluated by reduction of infarct volume, and increasingly by reduced impairment in sensorimotor tests. Except for sensorimotor disabilities stroke patients often suffer from long-term memory impairment [2]. This can be reproduced in animal stroke models as impaired function in cognitive tests, including the radial 8-arm maze [3], passive avoidance [4,5] and the Morris water maze [5-7]. However, there are diverging results using different stroke models and different tests of cognitive function, highlighting the need to validate the ability for a specific behavioral test to measure an impairment following a specific brain lesion.

Enriched environment and stroke recovery

Housing in an enriched environment (EE), i.e. 8-12 rats together in a large cage with a variety of objects that are exchanged or moved daily, enhances the recovery of sensorimotor function after brain injury, including focal ischemia [1,8-10]. This is linked to plastic processes outside the lesion, such as increased number of neurons, dendritic arborization and increased spine density [9,10]. Whether housing in EE can improve recovery of the memory impairment after focal cerebral ischemia has not been established. If a positive effect could be clearly established this could provide a useful model for studies of the neurobiological mechanisms of recovery of cognitive functions after ischemic stroke.

Materials and Methods

Surgery and housing

Male Sprague-Dawley rats were group housed 5 per cage in standard laboratory cages (595 x 380 x 200 mm), temp 20-21°C, lights on 06.00 to 18.00 h., with free access to water and food (standard rat chow). One week prior to surgery rats were accustomed to handling and for two days trained and tested in the sensorimotor tests described in [1]. MCAo was induced by the intraluminal suture model, adopted from Longa et al. [11], described in detail in [1]. Briefly, non-fasted rats were anaesthetized with halothane. Body temperature, arterial blood pressure, blood gases and blood glucose levels were controlled and kept within physiological ranges during surgery. The right MCA was occluded by insertion of a poly-L-Lysine coated, sandpaper blunted 3-0 nylon monofilament suture through the carotid arteries. Paresis was confirmed by Bederson score 75 minutes after induction of ischemia [12], and the suture was removed after 90 minutes to allow reperfusion.

Postoperatively, animals were housed individually in standard cages (425 x 266 x 185 mm). Two days after surgery, rats with confirmed sustained hemiparesis were randomized to a deprived environment (n=13), i.e. individual housing in standard cages (425 x 266 x 185 mm) or EE (n=12). In the EE 8-10 rats were housed in a large cage (820 x 610 x 450 mm) with elevated horizontal and inclined boards and ladders and equipped with several objects, such as wooden tunnels, a running-wheel, a chain and a swing [8,13]. The objects were moved around once daily and some objects were exchanged with new ones.

Sensorimotor tests

Recovery of sensorimotor functions was evaluated by the ability to traverse a rotating pole and in two forelimb placement tests at day 7, 14, 21 and 28 after MCAo, details in [1].

Water maze set-up

The water maze was a black circular pool 180 cm in diameter and 60 cm in height, and was filled to a depth of 32 cm with water at 25°C. A circular transparent Plexiglas platform, 10 cm in diameter, was permanently placed in the middle of the northeast quadrant, 40 cm into the pool, and 1.5 cm below the water surface. The maze was placed in a semi-dark room (2.7 m x 3.0 m) with white walls and with black and white distal visual extra maze cues present on the walls around the pool. The experimenter and the equipment were located in an adjacent room, more details in [14]. A video camera connected to an image analyzer (HVS Image, Hampton, UK) recorded latency and swim path length to find the platform. It also calculated the average and cumulative distance to the platform (i.e. the mean or the sum of the distance to the platform every second of the trial, also known as Gallagher measures) and time spent near the pool wall (within 9 cm).

Water maze procedure

On each morning before the start of the experiment, rats were taken from their home cage (enriched or deprived) to new small individual cages (425 x 266 x 185 mm) and left for 30 minutes in the experimental room for acclimatization. Two days before the first learning trial, the animals were habituated to swimming for 60 seconds in the pool without a platform. On day 31-34 after surgery one block of four trials were given for four consecutive days. A trial consisted of gently placing the rat by hand into the water, facing the wall of the pool. A different starting point was used on each of the four daily trials. The order of starting points was pseudorandom, but the same for all animals. When finding the platform the rats were allowed to remain there for 30 seconds, then held gently and dried for another 30 seconds until the next trial. If the rat failed to locate the platform within 120 seconds, it was guided by hand to the platform. At day five the animals were given one 60 seconds retention probe test in which the platform was removed from the pool. During retention, the total time and distance each rat swam in the former platform quadrant were recorded, as well as the time spent in the former platform position. All water maze tests were performed from 8.00-12.00 h.

Water maze statistics

Data from the four trials for each block in the water maze was averaged. Morris water maze data were analyzed by a two-way mixed ANOVA (Group [3] x Trial block [4]), with repeated measures on the trial block variable, followed by Tukey's post-hoc test. Infarct volumes between the ischemic groups were compared by a t-test. To investigate a possible association between larger infarct volume and longer mean latency/swimpath (day 1-4), we used Pearson (one-tailed) correlations. $p<0.05$ was considered significant.

Histology

Indirect infarct volumes were quantified by measuring ipsilateral and contralateral areas of intact brain tissue in haematoxylin and eosin-stained cryostat sections 1000 µm apart (covering levels from +4.5 mm to -7.5 mm from Bregma).

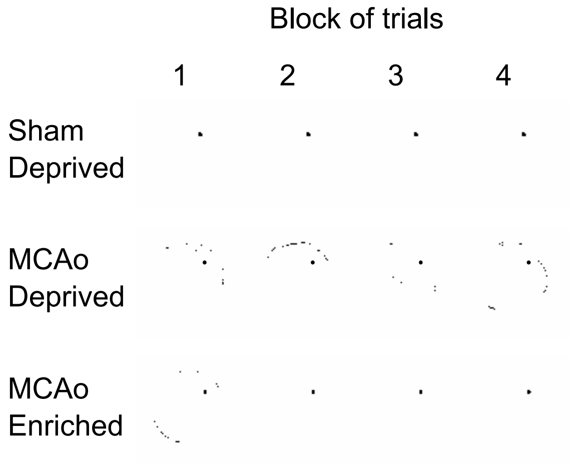


Figure 1. Swim paths from the last of four swims in each blocks of trials, day 31-34 after MCAo. Note thigmotaxis behavior in ischemic deprived rat. Reproduced with permission. from [1].

Results

Histology

Indirect infarct volume was 35.6 ± 8.5% of the contralateral hemisphere in deprived rats and 36.5 ± 11.2% in rats housed in EE after MCAo. Infarcts did not include the hippocampus, but in five rats (one deprived and four enriched) there was a subtotal loss of hippocampal neurons (within the CA1 subregion in three and within CA1-CA3 in two rats).

Sensorimotor function

The MCAo groups showed significant left-sided impairments in the sensorimotor tests compared to sham operated rats. Postischemic EE-housing improved recovery of function in the rotating pole test and the whisker-elicited forelimb placement test over the first four weeks of recovery in differential housing [1].

Results from water maze

MCAo significantly impaired the ability to find the hidden platform. Postischemic EE-housing substantially attenuated this impairment. Thus, the MCAo+deprived group was significantly inferior to sham+deprived, i.e. longer latency ($p = 0.0020$) and swim path length ($p = 0.0017$) before finding the platform and also longer average distance ($p = 0.013$) and cumulative ($p = 0.0026$) distance to the platform during the swims. The MCAo+enriched were superior to the MCAo+deprived with significantly shorter latency ($p = 0.0077$), swimpath length ($p = 0.019$), average distance to platform ($p = 0.0019$) and cumulative distance to platform ($p = 0.0075$). The MCAo+enriched group did not differ significantly from sham+deprived in any of these parameters (Figure 2A).

The rats in the MCAo+deprived group spent more time swimming near the pool wall than sham+deprived ($p < 0.001$). This swimming pattern was significantly reduced by postischemic EE housing ($p = 0.030$ vs. MCAo+deprived) (Fig 2B). The sham+deprived and MCAo+enriched groups significantly reduced their time spent near the wall over the four training blocks whereas the MCAo+deprived group did not.

In the probe test the sham+deprived and MCAo+enriched groups spent more time swimming in the correct quadrant (sham+deprived, 31%; MCAo+enriched, 35%; MCAo+ deprived, 28%) and in the correct platform position (sham+deprived, 4.2%; MCAo+enriched, 4.2%; MCAo+deprived, 1.5%) but this did not reach statistical significance.

Correlation between histology and water maze

In the deprived group spatial memory (i.e. mean swim path length during trial blocks 1-4) was significantly correlated with infarct volume (r=0.58, p=0.020). Thus rats with larger ischemic lesions swam a longer distance before finding the platform. Rostro-caudal subdivision of the brains showed stronger correlation with lesion size in the caudal parts of the brain. In contrast rats housed in EE showed no association between water maze performance and infarct volume.

Water maze performance was not significantly correlated with occurrence of hippocampal damage. There was no correlation between time spent near the pool wall and infarct volume (total, cortical or striatal) in either groups.

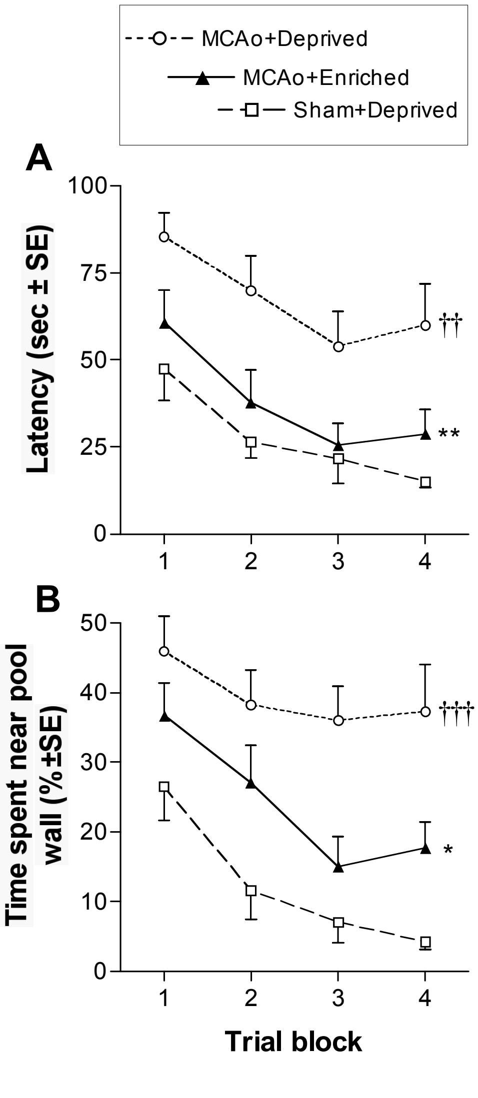


*Figure. 2. Learning curves of trial blocks 1-4 during the Morris water maze test, days 31-34 after middle cerebral artery occlusion (MCAo). (A) Latency to reach the hidden platform. (B) percentage of time spent near the pool wall. Data are presented as mean ± SE. Data from four swims/day from different starting positions are averaged for each trial block. *p< 0.05, **p< 0.01 MCAo+deprived versus MCAo+enriched; ††p< 0.01, †††p< 0.001 MCAo+deprived versus sham+deprived. Modified from [1].*

Discussion

In our studies we find that a classic Morris water maze test with four blocked trials for four consecutive days is a useful test for evaluating therapies directed to reduce functional impairment after MCAo in rats. We have found that postoperative housing in EE reduce the spatial memory impairment after MCAo in rats [15]. This effect of EE was not related to differences in size or distribution of the brain infarct.

Using the water maze in animals with large brain lesions may introduce difficulties with the interpretation of the results. For example, differences in swim speed due to the motor impairment after a large cortical stroke and treatment-induced differences in motor recovery create a problem with the interpretation of the latency data. This has been the case in previous, similar studies [16]. Also in the present study swim speed was significantly higher in the EE compared to the deprived ischemic group [15]. Thus, other water maze parameters, not influenced by swim speed have to be evaluated. In the present study postischemic EE significantly reduced swim path length to platform and Gallagher measures (average and cumulative distance to platform). Thus our interpretation is that the observed differences in water maze performance after MCAo are not entirely due to differences in motor impairment but reflects an EE-induced decrease in cognitive impairment. Apart from motor impairment, visual impairment, i.e. heminanopsia or even neglect of the left side may influence performance in the water maze, not directly related to spatial memory impairment. To control for this confounder a visible platform test may be useful. In the presented study this is not included. We have however in a separate pilot found no impairment after MCAo in a visible platform test (unpublished data).

The water maze is a complex test, requiring different cognitive skills for optimal performance, depending on the details of the set-up. In our set-up the rats are not pretrained in the maze and are first required to learn the procedural parts of the test. The MCAo+deprived group spent a lot of time circling around the pool edge (thigmotaxis). This behavior has previously been described after MCAo, and has been associated with striatal damage [5,17]. In the present study thigmotaxis was not significantly associated with lesion size, neither total, cortical or striatal. Persistence of thigmotaxis indicates failure to learn the procedural parts of the place navigation, which may represent a non-spatial learning component of the water maze task; i.e. the probability of finding the platform increases by swimming in the inside of the maze [17]. Postischemic EE significantly reduced thigmotaxis induced by MCAo, which is likely to contribute to the decreased latency and swim path length to find the platform. It has also been suggested that anxiety in the test situation may increase thigmotaxis, thus hampering the results of the test and interfering with the aim of the test to study spatial memory.

Statistical methods used to analyze water maze data vary and several considerations may be made. Due to the fairly large trial-to-trial variability we choose to average data from the four trials in each daily block. This gives a nice description of the gradually increasing ability to find the platform for each block of trials, with acceptable variability. The groups were compared using a mixed two-way ANOVA, with repeated measures on the trial block variable, and as main outcome we considered the over-all effect by group, i.e. superior performance over the entire test. Some use significant interaction in the ANOVA to show group difference in the slope of the learning curve. This could be useful in some studies, but would be misleading in ours. In our study there is no interaction because the sham and MCAo+enriched groups are superior to the MCAo+deprived group already at day one, and over days 2-4 there is a sustained, stable difference in performance. Moreover, the probe test is often used as primary result of the water maze to show differences in spatial learning. In our study there was a trend for increased time in correct quadrant and passes over platform position in the 60 seconds probe test, but this was not significant. A clear impression while testing the rats was a decreasing interest to go to the platform over the four days, more jumping off, etc. This is supported by the learning curves, i.e. increased latency and swim path on day 4 compared to day 3. Thus a possible explanation for no effect in the day 5 probe test may be that over many trials the rats realize that the platform offers no long-term escape from swimming and start searching for alternative escapes. This also suggests the use of fewer trials.

Conclusion

The Morris water maze test is a useful tool to evaluate recovery of cognitive impairment after MCAo in rats. MCAo results in large brain lesions, with multimodal functional impairments. This has to be taken into account when designing and interpreting the results of a water maze test.

Acknowledgements

The authors wish to thank Dr. Barbro B Johansson, Dr. Per Wester, Dr. Inga-Maj Johansson, Dr Ingegerd Söderström and Wei Jiang for valuable help and advice on the stroke model, the enriched environment and the behavioral tests. This work was supported by the Swedish Medical Research Council (grant no 04x-12237 and 71p-11769), the Swedish Heart and Lung foundation, Söderbergs Foundation, the Medical Faculty of Umeå University, the Swedish Stroke Foundation and the Swedish Association for the Neurologically Disabled.

References

1. Dahlqvist P.; Ronnback, A.; Bergstrom, S.A.; Soderstrom, I.; Olsson, T. (2004). Environmental enrichment reverses learning impairment in the Morris water maze after focal cerebral ischemia in rats. *Eur. J. Neurosci.*, **19**, 2288-2298.
2. Asplund, K.; Stegmayr, B.; Peltonen, M. (1998), *From the twentieth to the twenty-first century: A public health perspective on stroke.* In: Ginsberg, MD; Bogousslavsky, J. (Eds.) Cerebrovascular Disease: Pathophysiology, Diagnosis, and Management. Malden, MA: Blackwell Science, pp. 901-918.
3. Okada, M.; Tamura, A.; Urae, A.; Nakagomi, T.; Kirino, T.; Mine, K.; Fujiwara, M. (1995). Long-term spatial cognitive impairment following middle cerebral artery occlusion in rats. A behavioral study. *J. Cereb. Blood Flow Metab.*, **15**, 505-512.
4. Borlongan, C.V.; Cahill, D.W.; Sanberg, P.R. (1995). Locomotor and passive avoidance deficits following occlusion of the middle cerebral artery. *Physiol Behav.*, **58**, 909-917.
5. Yonemori, F.; Yamaguchi, T.; Yamada, H.; Tamura, A. (1999). Spatial cognitive performance after chronic focal cerebral ischemia in rats. *J Cereb. Blood Flow Metab.*, **19**, 483-494.
6. Yonemori, F.; Yamada, H.; Yamaguchi, T.; Uemura, A.; Tamura, A. (1996). Spatial memory disturbance after focal cerebral ischemia in rats. *J. Cereb. Blood Flow Metab.*, **16**, 973-980.
7. Markgraf, C.G.; Green, E.J.; Hurwitz, B.E.; Morikawa, E.; Dietrich, W.D.; McCabe, P.M.; Ginsberg, M.D.; Schneiderman, N. (1992). Sensorimotor and cognitive consequences of middle cerebral artery occlusion in rats. *Brain Res.*, **575**, 238-246.
8. Ohlsson, A.L.; Johansson, B.B. (1995). Environment influences functional outcome of cerebral infarction in rats. *Stroke*, **26**, 644-649.
9. Will, B.; Galani, R.; Kelche, C.; Rosenzweig, M.R. (2004). Recovery from brain injury in animals: relative efficacy of environmental enrichment, physical exercise or formal training (1990-2002). *Prog. Neurobiol.*, **72**, 167-182.
10. Johansson, B.B. (2004). Brain plasticity in health and disease. *Keio J. Med.*, **53**, 231-246.
11. Longa, E.Z.; Weinstein, P.R.; Carlson, S.; Cummins, R. (1989). Reversible middle cerebral artery occlusion without craniectomy in rats. *Stroke*, **20**, 84-91.
12. Bederson, J.B.; Pitts, L.H.; Tsuji, M.; Nishimura, M.C.; Davis, R.L.; Bartkowski, H. (1986). Rat middle cerebral artery occlusion: evaluation of the model and development of a neurologic examination. *Stroke*, **17**, 472-476.
13. Dahlqvist, P.; Zhao, L.; Johansson, I.M.; Mattsson, B.; Johansson, B.B.; Seckl, J.R.; Olsson, T. (1999). Environmental enrichment alters nerve growth factor-induced gene A and glucocorticoid receptor messenger RNA expression after middle cerebral artery occlusion in rats. *Neuroscience*, **93**, 527-535.
14. Johansson, I.M.; Birzniece, V.; Lindblad, C.; Olsson, T.; Backstrom, T. (2002). Allopregnanolone inhibits learning in the Morris water maze. *Brain Res.*, **934**, 125-131.
15. Dahlqvist, P.; Rönnbäck, A.; Bergström, S.-A.; Söderström, I.; Olsson, T. (2004). Environmental enrichment reverses learning impairment in the Morris water maze after focal cerebral ischemia in rats. *Eur. J. Neurosci.*, **19**, 2288-2298.
16. Puurunen, K.; Jolkkonen, J.; Sirvio, J.; Haapalinna, A.; Sivenius, J. (2001). An alpha(2)-adrenergic antagonist, atipamezole, facilitates behavioral recovery after focal cerebral ischemia in rats. *Neuropharmacology*, **40**, 597-606.
17. Devan, B.D.; Goad, E.H.; Petri, H.L. (1996). Dissociation of hippocampal and striatal contributions to spatial navigation in the water maze. *Neurobiol. Learn. Mem.*, **66**, 305-323.

A new paradigm to analyze observational learning in rats

L. Mandolesi[1,3], M.G Leggio [2,3], F. Federico [2,3] and L. Petrosini [2,3]
[1] *University of Naples "Parthenope", Italy*
[2]*Dept. of Psychology, University of Rome "La Sapienza", Italy*
[3] *IRCCS Fondazione Santa Lucia, Rome, Italy*

Abstract
A new paradigm of learning by observation was developed through an observational training in which rats observed companion rats performing different spatial tasks. Observer animals were separately housed in small cages suspended over a water maze tank. They repeatedly observed companion actor rats performing spatial tasks differing according to the experimental requirements. After the observational training, observer animals were or not surgically hemicerebellectomized. This surgical ablation was performed to block any further acquisition of new behavioral strategies during actual performance of swimming task. When cerebellar symptomatology stabilized, observer animals were actually tested in the Morris water maze task they had previously only observed. The observer rats displayed exploration abilities that closely matched the previously observed behaviors. The results obtained indicate that it is possible to learn complex behavioral strategies by observation using this new protocol. Furthermore, acquisition of the single facets that form the behavioral repertoire can be separately studied.

Keywords
cerebellum; imitation; explorative strategies; Morris Water Maze.

1 Introduction
There is a wide psychological evidence that the actual execution of a task is not the only means to acquire an ability. In fact, even observation of actions provides an effective means of learning new skills. From an evolutionary point of view, the observation of motor acts performed by others and the comprehension of the meaning of these acts appear to be essential for humans as well as for animals, because they allow making inferences about others' behaviors [2, 3, 6]. Also, the observation of someone else performing an action is the first necessary step for imitating him and then acquiring the competencies to reproduce similar motor skills [7]. This highly adaptive strategy is such a powerful and widespread way of acquiring new behaviors from others (parents, teachers...), that it is recognized as a specific learning paradigm, described in the literature as learning by imitation [1]. As part of the broader phenomenon related to recognizing, intending and preparing a movement, the observational learning can be thus considered somehow related to motor physiology. However, its mechanisms are rather controversial and little is known on the neural structures involved.
By taking advantage of the specific role of the cerebellum in the procedural acquisition [12], we developed a paradigm of observational learning apt for rats, to analyze if they are able to learn by observation behaviors linked to environment exploration. It was observed that, in the presence of a cerebellar lesion, namely a hemicerebellectomy (HCb), it is impossible to learn any new exploration strategy. In our observational learning paradigm, normal rats observed other conspecifics exploring a water maze to find the escape platform. After this observational training, "observer" rats underwent an HCb and were then tested in the Morris water maze (MWM).
As behavior to be learned we used the exploration of a new environment put into action in the water maze. It is an acquired behavior; it can be optimally learned through observational learning; it is a cerebellum-dependent behavior and thus freezable in the moment the cerebellar lesion occurs; it is based on sequenced explorative strategies that are used, separately or concurrently, to reach the escape solution with progressively shorter latencies and more direct navigational trajectories.

2 Materials

2.1 Animals
Adult Wistar rats (250-300g) were used. The animals were housed two to a cage with free access to food and water throughout the experiment and a standardized dark/light schedule (10/14 h). All efforts were made to minimize animal suffering and all procedures were conducted in accordance with the Guide for the Care and Use of Laboratory Animals (National Academy Press, Washington, 1996).

2.2 Special equipment
For construction of the observation equipment:
Ten small wooden roofed cages (30 x 15 x 40 cm) with black inside walls were assembled together in two rows of five cages. The cage floors consisted of a metallic grid with open spaces of 1 x 1 cm, through which each rat observed the scene below. The cages leaned on a support located 60 cm over the water maze tank, where companion rats swam according to specific paradigms (Fig. 1).

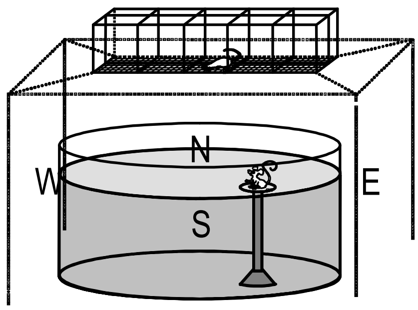


Figure 1. *Schematic diagram of the experimental setting used in the present observational learning paradigm.*

For monitoring spatial performances in the MWM:

The paths taken by observer animals in the pool were monitored by a video camera (Sony CCD-IRIS Black and White) mounted on the ceiling. The resulting video signal was relayed to a monitor (Sony), that allowed both on- and off-line analyses, and to an image analyzer (EthoVision 3, Noldus Information Technology, Wageningen, The Netherlands). The x- and y- coordinates of the rat's position were sampled and stored on disk.

2.3 Detailed procedure

Observational training. The animals were housed separately in small cages suspended over the water maze tank. Through the cage floor they repeatedly observed companion rats performing spatial tasks that differed according to the experimental requirements. Each suspended animal observed 200 trials performed by conspecifics. During observational training, at each trial start observer animals were acoustically (radio music, whistle) and mechanically (slightly touching the animals inside the cages, moving the cage support, water sprays from the grid floor) stimulated to maintain a high level of arousal.

MWM equipment. A circular plastic pool (diameter 140 cm) with white inside walls was located in a normally equipped laboratory room, uniformly lighted by four neon lamps (40 W each) suspended from the ceiling (3 m). No effort was made to enhance (or vice versa to impoverish) extra-maze cues, which were held in constant spatial relations throughout the experiments. The pool was filled with water (24°C), 50 cm deep, made opaque by the addition of 2 liters of milk. A white steel escape platform (10 cm in diameter) was placed in the middle of one cardinal quadrant (NW, NE, SW, SE), 30 cm from the side walls; it was either submerged 2 cm below or elevated 2 cm above the water level. Each rat was gently released into the water from a wall point facing the center of the pool. The animal was allowed to swim around to find the platform.

MWM protocols performed by "actor" animals and observed by "observer" animals. The observer animals watched MWM trials performed by actor animals executing a specific explorative strategy, according to experimental requirements. The first group of animals observed 200 MWM sessions carried out by intact animals performing a basic MWM paradigm. The actor animals learned to reach the escape platform that was hidden in the NW pool quadrant in the first 16 trials, visible in the NE quadrant in the successive 8 trials, and again hidden in the NE quadrant in the final 16 trials. Another group of animals observed 200 MWM sessions performed by intact animals searching for (and not finding) the platform. The actor animals were put in the MWM tank without the platform for 60 sec. Thus, observer animals repeatedly observed animals that scanned the whole tank in search of an escape solution. Another group of animals observed 200 direct findings of the platform performed by well-trained intact animals that reached without hesitation a hidden platform whose position they had previously learned. At the end of the observational training, the observer animals were or not hemicerebellectomized, according to their specific protocol.

Surgery. A right hemicerebellectomy (HCb) was performed with an i. p. solution of ketamine (90 mg/kg) and xylazine (15 mg/kg). The dura was excised and the right cerebellar hemisphere and hemivermis were ablated by suction; care was taken not to lesion extracerebellar structures. The cavity was filled with sterile gel foam and the wound edges were sutured. After recovery from anesthesia, the animals were housed two per cage.

Motor assessment. Testing was performed two weeks after the HCb, when no further change in cerebellar symptomatology was observed. The following aspects were taken into account: head and body tilts, position of hindlimbs, presence of ataxia, tremor, rearing behavior, falls to lesion side, wide-based locomotion, collapsing on the belly, pivoting, vestibular drop reactions and the ability to traverse a narrow path and to be suspended on a wire [9].

MWM protocols performed by observer animals. After the specific observational training, observer animals were tested in the MWM task according to a basic protocol. Each rat performed blocks of four trials, two blocks of trials per day. On reaching the platform, the rat was allowed to remain on it for 30 sec before being again placed in the water for the next trial. If a rat failed to locate the platform within 120 sec, it was guided there by the experimenter and allowed to stay on it for 30 sec. In the first four sessions (trials 1-16) the platform was hidden in the NW pool quadrant (Place I), in the successive two sessions (trials 17-24) the platform was visible in the NE quadrant (Cue phase) and in the final four sessions (trials 25-40) the platform was hidden in the NE quadrant (Place II) [10, 12].

Recording of MWM performances. To analyze the MWM performances, the parameters taken into account were successes in finding the platform, finding latencies and swimming trajectories. By considering spatial and temporal distribution of swimming trajectories, path length, swimming speed, percentage of time spent in inner or outer annuli, and headings (deviation between the rat's actual direction when leaving the edge of the tank and a straight line from the start location to the tank point containing the platform), exploration behavior was divided into four main categories: - circling, peripheral swimming at tank wall, without entering in the inner sectors of the arena; - extended searching, swimming in all pool quadrants, visiting the same areas more than once; - restricted searching, swimming in some pool quadrants, not visiting some tank areas at all; - direct finding, swimming towards the platform without any foraging around the pool.

Histological controls. After completion of behavioral testing, the HCbed animals were anaesthetized with Nembutal and perfused with saline followed by 10% buffered formalin. The extent of the cerebellar lesion was determined from Nissl-stained 40 μm frozen sections. Surgical lesion of the cerebellar structures of the right side was considered appropriate if the right cerebellar hemisphere, the right hemivermis and the deep nuclei were ablated, while the left side of the cerebellum and all extracerebellar structures were completely spared. *Statistical Analysis.* Metric unit results of the different experimental groups were compared by using one-way or two-way "p x q" analyses of variance (ANOVAs) with repeated measures, eventually followed by multiple comparisons using Tukey's tests.

3 Results

The present protocol allows for different kinds of observational training, based on the observation of single explorative strategies, as well as on the observation of the entire explorative repertoire, to study their learning power in the acquisition of complex spatial behaviour. After the observational training, observer animals were tested in the same MWM apparatus they had previously only observed. By comparing the frequency of actually using strategies that matched the previously observed strategies, it was possible to verify whether observational learning has taken place.

The explorative strategies of the observers were significantly influenced by the strategies the animals had previously observed. Observers of the entire repertoire of explorative strategies put into action by intact animals exhibited a competent exploration of the tank from the very first sessions, avoiding almost completely the repetitive circling in the pool peripheral areas characteristic of the initial steps of tank exploration. Of course, this efficient behaviour resulted in escape latencies significantly reduced in comparison to those of animals without observational training. Of course, as they explored the tank, the observer animals progressively learned the actually performed explorative strategies. Thus, the influence of the observational learning was detectable only in the first trials, when the learning based on actual performance was not yet present.

To analyse the influence of observational learning of single explorative strategies and single it out from the effects of learning through actual performance, it was thus necessary to block any further learning during the actual MWM performances. Of course, it was necessary that the animals maintained intact swimming abilities. Hemicerebellectomy fulfilled this double requirement. In fact, hemicerebellectomized animals are rather competent at swimming, and in water exhibit the same performances of controls in turning abilities, forepaw inhibition, and maintenance of nose above water level. Conversely, the cerebellar lesion blocks any further procedural acquisition, although it does not affect the procedural competencies learned before the cerebellar lesion. Thus, the spatial performances actually put into use by the observer animals, that underwent cerebellar lesion, had to have been learned before the cerebellar lesion and thus by observation.

As paradigm of observational learning, we used a group for which the observational training consisted of observing intact animals searching for (and not finding) the platform, called Os+H (Observation of searching behavior + Hemicerebellectomy) group. When actually tested in the MWM, this observationally trained group displayed performances biased by the previously observed performances. In fact, in spite of the cerebellar lesion, which, in the absence of pre-training, typically provokes repetitive circling at the pool periphery, they circled at the pool periphery only in the very first trials; then, throughout the testing, they persisted in swimming around the pool looking for the platform. Such searching behavior, although effective for searching (and then finding) the platform, was not an adaptively modifiable strategy, as indicated by the flattened slope of the latency time course and by the almost complete lack of evolution from one kind of exploration behavior to another.

Another paradigms of observational learning consisted of observing only direct finding of the platform, (group Of+H (Observation of finding behavior + Hemicerebellectomy). These animals displayed long-lasting circling with a greatly reduced percentage of searching. Interestingly, they exhibited direct finding, as their second most frequent strategy. In fact, the animals either circled at the pool periphery or initially tried to detach from the pool walls towards the central sectors possibly containing the platform. If the escape solution failed, they immediately returned to circle at the pool periphery. This pattern of behavior resulted in a very high percentage of peripheral circling, paradoxically accompanied by a rather high percentage of direct finding of the platform.

4 Discussion

The paradigm of observational learning is made possible by the rats' natural curiosity about the surrounding environment. The advantage of the present procedure is that, during the observational training the animals are free to move without constraint. Therefore, since they are unstressed, the observer animals are interested in the surrounding environment and in the actors' behavior. Furthermore, the suspension of the observers in single cages prevents social interactions that could distract them. Furthermore, the possibility of simultaneously training ten subjects on one hand speeds up experimental procedures and, on the other minimizes experimental variability linked to environmental factors.

The described observational learning protocol allows for acquisition of spatial strategies by observation. The observational training shapes the successive actual performances in the MWM task, improving the exploration of the arena in the search for the platform.

Paradigms of actually practiced learning, based on trial and error, have heavy limits to allow analyzing single behavioral steps, since the physical experience itself activates their sequential organization, making it almost impossible to single out their acquisition [5, 7]. Conversely, in the present protocol, by exploiting the potential offered by the observation of each single behavioral step in combination with the block of learning elicited by the cerebellar lesion [8, 11], behavioral steps can be singularly acquired. Thus, one very attractive feature of this new protocol is that the numerous facets of a complex performance can be singly analyzed. For example, this protocol can be not only used to analyze the capacity to learn single exploration strategies by observation but also to study whether it is possible to acquire localizatory knowledge about specific goal positions by observation.

Considering the results as a whole, it seems that the newly developed protocol of learning by observation was successful in shaping the exploration strategies according to the kind of observational training [4].

Besides these aspects, the described protocol has the potential to study whether it is possible to bias actual MWM performances through the observation of "wrong" exploration behavior, as the repeated observation of animals circling in the peripheral pool regions [4]. This approach can allow investigating the effects of negative models on different learning paradigms.

Of course, the present protocol may be also used to test the capacity to learn by observation in contexts and experimental settings that are completely different from the MWM task. Many behavioral tasks can be learned by observation, and by means of this protocol it is possible to fractionate their organization and study their sequential components. Radial maze, plus-maze, and open field with spatial or object change are behavioral tests that can be studied using this new approach.

References:

1. Byrne, R.W., Russon, A.E. (1998). Learning by imitation: a hierarchical approach. *Behav. Brain Sci*. **21**, 667-721.
2. Decety J., Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends Cogn. Sci.* **3**,172-178.
3. Gallese, V., Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* **2,** 493-501.
4. Graziano A., Leggio M.G., Mandolesi L., Neri P., Molinari M. and Petrosini L. (2002) Learning power of single behavioral units in acquisition of a complex spatial behavior: an observational learning study in cerebellar-lesioned rats. *Behav. Neurosci*. **116**, 116-125.
5. Heyes C.M., Dawson G.R. (1990) A demonstration of observational learning in rats using a bi-directional control. *Quart. J. Exp. Psicol.* **42,** 59-71.
6. Heyes, C.M. (1998). Theory of mind in nonhuman primates. *Behav. Brain Sci.* **21**, 101-148.
7. Heyes, C.M., Jaldow, E., Nokes, T., Dawson, G.R. (1994). Imitation in rats: the role of demonstrator action. *Behav. Process.* **32**, 173-182.
8. Leggio M.G., Neri P., Graziano A., Mandolesi L., Molinari M. and Petrosini L. (1999) Cerebellar contribution to spatial event processing: characterization of procedural learning. *Exp. Brain Res*. **127**, 1-11.
9. Leggio, M.G., Molinari, M., Neri, P., Graziano, A., Mandolesi, L., and Petrosini L. (2000). Representation of actions in rats: the role of cerebellum in learning spatial performances by observation. *Proc. Natl. Acad. Sci. USA* **97**, 2320-2325.
10. Molinari M., Petrosini L., and Gremoli T. (1990) Hemicerebellectomy and motor behavior in rats. II. Effects of cerebellar lesion performed at different developmental stages. *Exp. Brain Res*. **82**, 483-492.
11. Petrosini L., Leggio M.G. and Molinari M. (1998). The cerebellum in the spatial problem solving: a co-star or a guest star?, *Progr. Neurobiol*. **56** 191-210.
12. Petrosini, L., Molinari, M., & Dell'Anna, M.E. (1996). Cerebellar contribution to spatial event processing: Morris Water maze and T-maze. *Eur. J. Neurosci.* **9**, 1882-1896.
13. Schenk, F., and Morris, R.G.M. (1985). Dissociation between components of spatial memory in rats after recovery from the effects of retrohippocampal lesions. *Exp. Brain Res.* **58**,11-28.

Using proximal and distal cues in the water maze: strategies and neural bases

E. Save, C. Parron, B. Poucet
Laboratory of Neurobiology and Cognition, UMR 6155 CNRS, Université Aix-Marseille I, Marseille, France.

Abstract

Rats were trained in two versions of a place navigation task in the Morris water maze. In the proximal condition, rats had to use nearby objects placed in the pool and in the distal condition they had to use room cues. Path analysis was made by a Videotrack 2.6 tracking system (View Point, Champagne-au-Mont-D'or, France). Hippocampal lesions affected learning in the two conditions. Rats with parietal lesions were impaired in the proximal condition only whereas rats with entorhinal lesions were impaired in the distal condition only. The results suggest the existence of two processing systems, mediated by different neural networks, one devoted to the processing of proximal cues and the other to the processing of distal cues.

Keywords

Spatial cognition, Navigation, Water maze, Hippocampus, Entorhinal cortex, Parietal cortex.

1 Introduction

Animals are able to form spatial representations (or spatial maps) of their environment which allow them to perform flexible spatial behaviors [7]. Navigation based on a spatial map is classically tested in the Morris water maze [5]. In the place navigation version of the water maze task, the animal is required to use distal room cues to locate and navigate toward a submerged, i.e. not visible, platform. The possibility of using local cues (olfactory, visual cues), closely associated to the platform location, is virtually eliminated due to the water. That the animals exclusively rely on distal room cues to locate the goal is tested during a probe trial with the platform removed. Rats usually spend more time in the area that contained the platform during training than in other equivalent areas of the pool, thus exhibiting clear place learning ability.

A number of behavioral, lesion and electrophysiological studies have suggested that animals are also able to rely on proximal cues to form spatial representations [2,3,10] However, it has been often reported that it is much more difficult to train animals to perform spatial tasks on the basis of proximal cues than distal cues. Thus, in the place navigation task in the Morris water maze, only a few number of trials are needed to obtain asymptotic performance level [4]. In contrast, tens of trials are necessary for the animal to use objects directly placed in the arena to locate a reward location [2]. This suggests that, although proximal and distal cues can be used by the animals to form a spatial map enabling place navigation, these two kinds of cue are processed by distinct systems.

These two systems may correspond to activation of different neural networks. The central role of the hippocampus in place navigation using distal cues has been repeatedly demonstrated since the seminal work of Morris et al. (1982) [6]. However, it is now acknowledged that the processing of spatial information is distributed over a large network of brain structures including cortical and subcortical regions [1]. In particular, the hippocampus receives inputs from several associative cortices, including retrosplenial, parietal, postrhinal, perirhinal cortices, via the entorhinal cortex, suggesting that it deals with highly processed spatial information. The role of these cortical areas in the encoding of environmental cues remains to be determined. We thus investigated the possibility that distal and proximal cues are processed by different networks and examined the effects of hippocampal, parietal and entorhinal cortical lesion on place navigation in the Morris water maze based of two kinds of cue, proximal and distal.

2 The use of proximal cues in the water maze

2.1 Experiment design

Proximal cues were 3 large objects, a gray and white vertically stripped cylinder, a gray and white horizontally stripped cylinder, and a gray cone (41, 61, and 45 cm high, respectively). They were smooth enough so as to prevent the rats to cling to them and use them as resting places. The objects were directly placed in the pool (1.40 m in diameter) but could not be used as beacons. They were located at the periphery, against the pool wall at a distance from the platform (Figure 1). To prevent the use of room cues, the pool was surrounded by an opaque curtain. In addition, the objects and the platform were rotated as a rigid set 120° with respect to the room, in a pseudorandom direction between days. A radio set placed above the apparatus allowed to mask possible directional auditory cues.

Thus, it was expected that in the proximal cue condition, the animals would need to form an allocentric representation to locate the goal as they do in the distal cue condition. In this latter condition, the rats were required to use room cues, such as posters, tables, cabinets etc., to locate the platform. Independent groups of rats were trained in the proximal and distal cue conditions. Daily 6-trial sessions in the proximal condition and 4-trial sessions in the distal condition were conducted. At the end of training, in the two conditions, a 60 sec probe trial with the platform removed was conducted allowing place learning response evaluation.

In the distal cue condition, rats with entorhinal cortical lesions received additional post training trials. To examine the possibility that these rats did not rely on distal cues, they received two successive trials from different starting points with an opaque curtain placed around the pool (« curtain trials »).

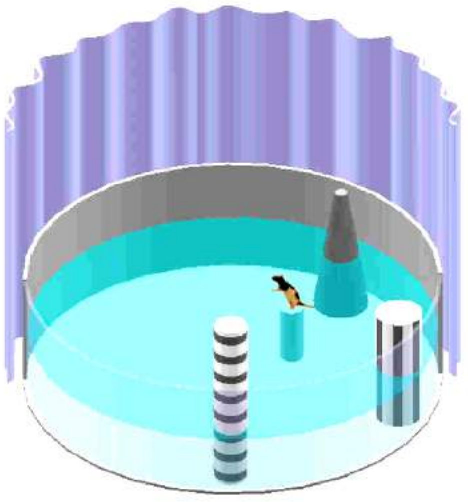


Figure 1. Proximal cue condition. Water maze containing three objects and surrounded by a curtain to prevent the use of room cues.

2.2 Data analysis

Trials were processed on-line by a Videotrack 2.6 tracking system (View Point, Champagne-au-Mont-D'or, France). This produced files consisting in a list of x-y coordinates (from 0 to 512). Sampling was done at 25 Hz. Custom-made computer programs allowed to calculate a number of parameters. In the two conditions, escape latency and total distance swam to the platform were used to quantify place navigation performance. Swim speed measured general locomotor activity.

Presence of the objects in the pool yielded a particular behavior combining exploration and navigation. The rats frequently swam by the object, touching them, before they search for the goal. Exploratory activity has been previously shown to reflect the acquisition of spatial information and is likely crucial for building a representation of space [11]. To quantify object exploration and to examine possible deficits due to entorhinal cortex lesions, we measured the amount of contacts made by the animals with the objects.

3 Effects of hippocampal and cortical lesions on navigation based on the use of proximal and distal cues

In a first study, we showed that rats with hippocampal lesions were impaired in using either proximal or visual cues to navigate accurately toward the platform [9]. This result is fully compatible with the key-role of the hippocampus in the formation and use of an allocentric representation, would it be based on distant or nearby cues. Rats with parietal cortex lesions were impaired like hippocampus-lesioned rats in using proximal cues to exhibit accurate place navigation (Figure 2).

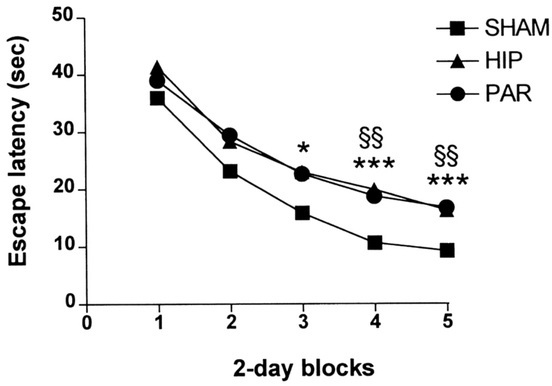


*Figure 2. Proximal cue condition. Escape latencies for each group during training. * and *** indicate statistical differences between control rats (sham) and hippocampal rats (Hip) (p<0.05 and p<0.01, respectively). §§ indicates statistical differences between control rats and parietal rats (Par) (p<0.025).*

In contrast, they were able to use distal cues. These results suggest that there are two processing systems, one dealing with proximal cues and the other dealing with distal cues. Each system may require activation of different brain structures but our results seem to indicate that the hippocampus contributes to both systems.

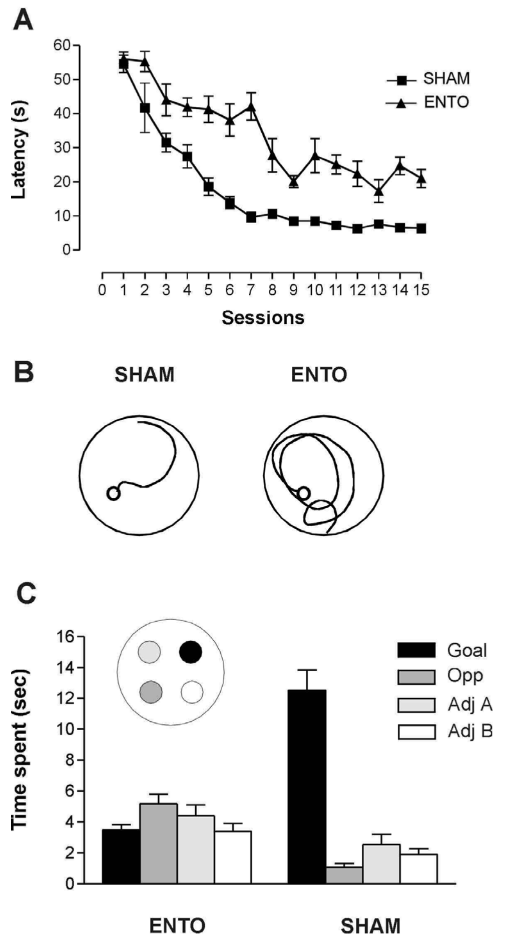


Figure 3. Distal cue condition. A. Escape latency during acquisition. B. Examples of trajectories at the end of acquisition. C. Time spent in the goal area and the 3 equivalent areas during the probe trial.

We then sought to identify other structures that would be involved in the processing of proximal or distal cues. The entorhinal cortex likely plays a crucial role in cue processing since it interfaces the hippocampus with associative cortical areas. Would it be involved in both systems like the hippocampus or only in one of them ?
Surprisingly, rats with entorhinal cortex lesions were found to be impaired in the distal cue but not in the proximal cue condition [8]. They displayed deficits in acquisition and in the probe trial (Figure 3).

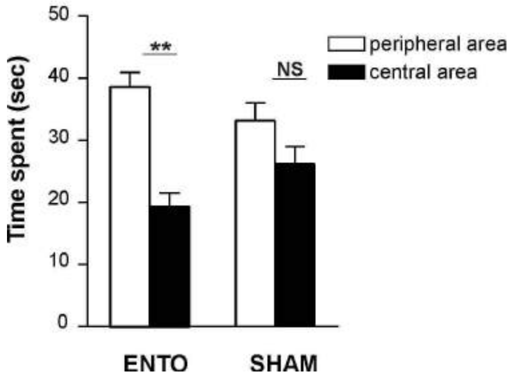


*Figure 4. Time spent in a peripheral annulus including the platform and in the central area of the pool. ** indicates significant difference (p < 0.01) and NS non significant.*

Observation of the trajectories suggested that rats with entorhinal cortex had adopted a looping strategy during acquisition. They swam at a distance from the wall corresponding roughly to the goal distance until they come across the platform (Figure 3B). This was confirmed by measuring the time spent in a peripheral annulus during the probe trial (Figure 4).

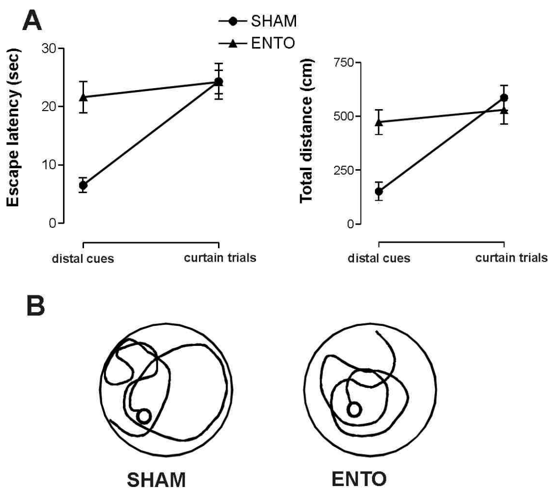


Figure 5. A. place navigation performance in the last training day with distal cues and in the averaged curtain trials for control and entorhinal lesioned rats. B. Examples of trajectories.

In addition, analysis of the curtain trials revealed that the performance of rats with entorhinal cortex lesions was not affected in absence of distal cues (Figure 5).
Overall, this suggests that the animals did not rely on distal cues but used the pool as a cue to reach the platform.
In the proximal cue condition, rats with entorhinal cortex lesions were not different from control rats (Figure 6A).
As shown in Figure 6B, rats frequently explored the objects before searching for the platform. To examine whether entorhinal-lesioned rats would differ from control rats in exploring the objects, we measured the amount of contacts with the objects. However, no difference was found. In addition, lesioned rats did not exhibit impaired locomotor activity (no difference in swimming speed).

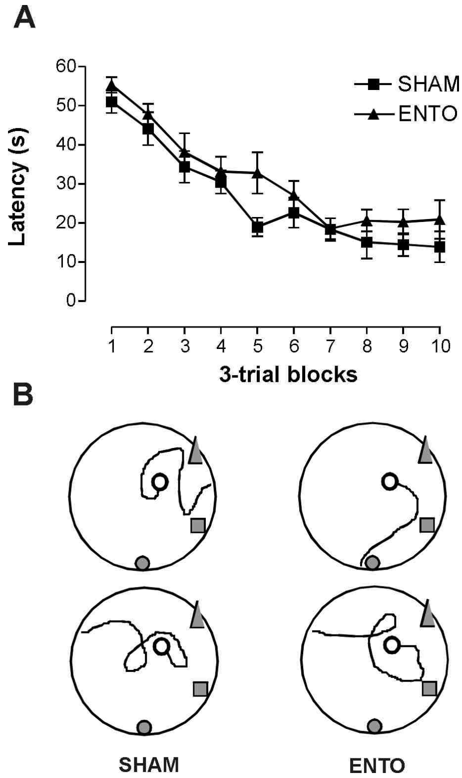


Figure 6. Proximal cue condition. A. Escape during acquisition. B. Examples of trajectories at the end of acquisition.

4 Conclusion

The results support the hypothesis that there are two processing systems, one activated when the animal has to use proximal cues and the other when it has to use distal cues. These systems probably interact very closely one with another but it is not known yet how they are related to each other. Such interaction may take place in the hippocampus. Further experiments must be conducted to identify the structures and therefore the neural pathways involved in either cue processing system.
Using the place navigation task with proximal cues also allows to draw parallels between the coding of space at a brain area level and at a single cell level. Indeed, in a number of studies, we have recorded hippocampal place cell activity in rats exploring a circular arena containing 3 objects. We are currently investigating how place cell location-specific activity is controlled by proximal objects relative to distal cues (Renaudineau, Poucet, and Save, unpublished data).
These two approaches are complementary and may provide some indications of how animals organize their perception of space and how spatial information is processed from the cortical sensory areas to the hippocampus via multiple associative cortical areas, and from global information flow in the brain to activation of a single cell.

Acknowledgments: supported by the CNRS and the Ministère de la Recherche et de la Technologie.

References

1. Aggleton, J.P.; Vann, S.D.; Oswald, C.J.P.; Good, M. (2000). Identifying cortical inputs that subserve allocentric spatial processes: a simple problem with a complex answer. *Hippocampus*, **10**:466-474.
2. Gothard, M.G.; Skaggs, W.E.; Moore, K.M.; McNaughton B.L. (1996). Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *Journal of Neuroscience*, **16**:823-835.
3. Gould-Beierle, K.L.; Kamil, A.C. (1996). The use of local and global cues by Clark's nutcrackers, Nucifraga Columbiana. *Animal Behavior*, **52**:519-528.
4. Harker, K.T.; Whishaw, I.Q. (2002). Impaired spatial performance in rats with retrosplenial lesions: importance of the spatial problem and the rat strain in identifying lesion effects in a swimming pool. *Journal of Neuroscience*, **22**:1155-1164.
5. Morris, R.G.M. (1981). Spatial localization does not require the presence of local cues*.* *Learning and Motivation*, **12**:239-260.
6. Morris, R.G.M.; Garrud, P.; Rawlins, J.N.P.; O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, **297**:681-683.
7. O'Keefe, J.; Nadel, L. (1978). *The hippocampus as a cognitive map*. London : Oxford University Press.
8. Parron, C.; Poucet, B.; Save, E. (2004) Entorhinal cortex lesions impairs the use of distal but not proximal landmarks during navigation in the rat. *Behavioural Brain Research.*, **154**:345-352.
9. Save, E.; Poucet, B. (2000). Involvement of the hippocampus and associative parietal cortex in the use of proximal and distal landmarks for navigation. *Behavioural Brain Research*, **109**:195-206.
10. Save, E.; Poucet, B.; Foreman, N.; Buhot, M.C. (1992). Objects exploration and reaction to a spatial and a non-spatial change in the rat following damage to the posterior parietal cortex or the dorsal hippocampus. *Behavioral Neuroscience*, **106**:447-456.
11. Thinus-Blanc, C.; Save, E.; Poucet, B. (1998). Animal spatial cognition and exploration. N Foreman, R Gillett (Eds.). *Interacting with the environment: a handbook of spatial research paradigms and methodologies vol. 2*. (pp. 59-86) Hove : Psychology press.

The application of Active Allothetic Place Avoidance (AAPA) task in the study of cognitive deficit and hyperlocomotion following MK-801 administration in rats

A. Stuchlik[1], V. Bubenikova[2], K. Vales[1]
[1] *Institute of Physiology, Academy of Sciences, Prague, Czech Republic*
[2] *Prague Psychiatric Center, Prague, Czech Republic*

Abstract
Blockade of brain NMDA receptors with MK-801 results in specific behavioral alterations, which can be ameliorated by antipsychotics. Nonetheless, the effect of antipsychotics on the cognitive deficit following MK-801 application is not elucidated yet. The present study tested whether clozapine alleviates the cognitive deficit following administration of two doses of MK-801 (0,1 a 0,3 mg/kg, s.c.).in the AAPA task. In this task, rats are required to avoid a room-frame fixed shock sector on the rotating arena, and this task requires the rats to separate spatial stimuli into coherent representations and use only the room-frame system for the navigation. Application of MK-801 increased locomotion and decreased spatial efficiency. Clozapine decreased total distance walked in the session but it was ineffective in alleviating decreased spatial efficiency following MK-801. It is summed up that in the AAPA task, clozapine is capable of decreasing hyperlocomotion but incapable of alleviating deficits in spatial cognition.

Keywords
MK-801, dizocilpine, animal model of schizophrenia, spatial cognition

1 Introduction
Application of non-competitive NMDA receptor antagonists is used as an animal model of schizophrenia with a relatively high face and predictive validities. Blocking NMDA receptors with MK-801 leads to specific behavioral changes, including hyperlocomotion, stereotypies, social deficit and impaired sensorimotor gating (measurable as deficit in the prepulse inhibition of acoustic startle reflex). These changes can be alleviated by application of classical and atypical antipsychotics. Cognitive deficit following systemic administration of MK-801 in the AAPA task was described recently [1]. However, the effect of antipsychotics on the cognitive deficit following NMDA receptor blockade is not understood yet. Aim of the present study was determine the effect of clozapine on the cognitive deficit following administration of two doses of MK-801 (0.1 and 0.3 mg/kg, s.c.).in the Active Allothetic Place Avoidance (AAPA) task.

2 Methods

2.1 Animals
Forty-eight naive male adult rats of the Wistar strain (5-month-old, weighing 350-450g) obtained from the Institute's breeding colony were used in the experiment. Animals were accommodated two per cage in 25 x 30 x 45-cm transparent plastic cages in a laboratory air-conditioned animal room with constant temperature (22^{o}C) and 12:12 light/dark cycle (with the lights on at 7:00 a.m.). Under light diethylether anesthesia, all animals were implanted with a low-impedance connector made from a hypodermic needle, which pierced the rat's skin between the shoulders. The sharp end of the needle was cut off and bent with tweezers to form a small loop, which prevented the connector from slipping out and provided anchor for an alligator clip, which was connected to a shock-delivering wire. Water and food were available *ad libitum* throughout the experiments. All manipulations were done in accord with the Law on Animal Protection of Czech Republic and with the appropriate directive of the European Communities Council (86/609/EEC).

2.2 Drugs
MK-801 (Dizocilpine maleate; [5R,10S]-[+]–5-methyl-10,11-dihydro-5H-dibenzo[a,d] cyklohepten-5,10-imine; Sigma Aldrich) was dissolved in saline (0.1 mg/ml and 0.3 mg/ml) and injected 30 min prior to behavioral training (0.1 and 0.2 mg/kg). Clozapine (Sigma; Aldrich) was dissolved in saline (5 mg/ml) acidified with 20 microl of concentrated acetic acid and injected 60 min prior to behavioral testing (5 mg/kg).

2.3 Apparatuses and behavioral procedures
The AAPA apparatus was described in detail elsewhere [2]. Briefly, it consisted of a smooth metallic circular arena (80 cm in diameter) enclosed with a 30 cm high transparent Plexiglas wall and elevated 1 m above the floor of the dimly lighted 4x5m room with the abundance of extramaze landmarks. At the beginning of each training session, a rat was placed on the rotating arena (1 rpm), where a directly imperceptible 60-degrees to-be-avoided sector (shock sector) was defined by the computer-based tracking system (iTrack, Biosignal Group Corp., USA), located in an adjacent room. The location of the shock sector could be determined exclusively by its spatial relations to distal orienting cues located in the room. The rat wore an infrared light-emitting diode (LED) fixed between its shoulders with a light latex harness, and its position was tracked every 40 ms and recorded onto a computer track file, allowing subsequent reconstruction of the track with an off-line analysis program (TrackAnalysis, Biosignal Group Corp., USA). Whenever a rat entered the shock sector for more than 0.5 s, mild electric shocks (50 Hz, 0.5 s) were delivered at intervals of 1.5 s until the rat left the shock sector for at least 0.5 s (Fig. 3). The shocks were delivered through a thin subcutaneous low-impedance nichrome wire implant on the back of the rat standing on the grounded floor. The appropriate shock current (ranging between 0.2-0.7 mA) was individualized for each rat to elicit a rapid escape reaction but to prevent freezing. Since the arena was rotating, the rat had to move actively away from the shock

in the direction opposite to the arena rotation, otherwise it was passively transported to the shock sector.

2.4 Design of experiments and data analysis

Rats (N=48) were divided into six experimental groups. The animals from the group SAL (n=8) were injected with 1ml/kg of saline (60 min and 30 min prior to behavioral testing), rats from the group SAL-CLOZ (n=8) were injected with clozapine (5 mg/kg) 60 min prior to testing and saline (1ml/kg) 30 min prior to testing, animals from the groups MK01 (n=8) and MK03 (n=8) were injected with 1ml/kg of saline 60 min prior to testing and with 0.1 mg/kg and 0.3 mg/kg of MK-801 30 min prior to testing, respectively. Subjects from groups MK01-CLOZ (n=8) and MK03-CLOZ (n=8) were injected with 5mg/kg of clozapine 60 min prior to the testing and with 0.2 mg/kg and 0.3 mg/kg of MK-801 30 min prior to the testing, respectively. All animals received injections of the same volume of liquid per kg weight.

Animals were trained for four consecutive daily sessions in the AAPA task, with the shock sector located in the north of the room. Experimental sessions in AAPA lasted 20 min and each rat had one session every day, carried out during daylight hours.

The following parameters were recorded and analyzed in order to assess rats‘ behavior in the AAPA: the total distance traveled in a session (DISTANCE) measured in the arena frame (which only takes into account active locomotion) reflected the locomotor activity of rats and the number of entrances to the shock sector (ERRORS) reflected the efficiency of spatial performance in the AAPA task. Latency to the first entrance into the shock sector TIME1ST reflected between-session learning. Maximum time a rat spent in the safe part of the arena between two ERRORS in a particular session was also recorded (MAX T). It reflected the ability to remember the shock sector location and to avoid it. Data from the four days of training in the AAPA task were analyzed using a two-way ANOVA (Treatments x Sessions) with repeated measures on Sessions. Tukey’s HSD test was used when appropriate. The significance was accepted on probability level of 5%.

3 Results

No vocalizations, increased defecations or ataxia were observed during and after the injections of MK-801 and clozapine. On the contrary, visual inspection revealed hyperactivity after injection of higher dose of MK-801, which was later confirmed by measuring the total DISTANCES. Rats were able to maintain correct postural positions and their main neurological reflexes were preserved.

As concerns DISTANCE, it was increased significantly in the groups MK01 and MK03 with respect to the group SAL, whereas clozapine blocked the MK-801 induced hypermotility. Analysis of spatial performance using the parameter ENTRANCES revealed that lower dose of MK-801 was without effect on ENTRANCES, whereas higher dose significantly increased the number of ENTRANCES. Clozapine failed to block the MK-801-induced increase of ENTRANCES.

As concerns the MAXT parameter, it was significantly increased in the groups MK01, MK03, CLOZ, CLOZ-MK01 and CLOZ-MK03 with respect to the group SAL, i.e., both clozapine and MK-801 impaired spatial navigation efficiency analyzed with this parameter.

4 Discussion

The results demonstrate that application of MK-801 increases locomotor activity and decreases spatial efficiency, measured as the increased number of shock sector entrances and decreased maximum time avoided. Administration of clozapine decreased total path traveled in the session but was ineffective in alleviating decreased spatial efficiency following MK-801.It is concluded that in the AAPA task, clozapine is capable of decreasing hyperlocomotion but incapable of alleviating deficits in spatial cognition.

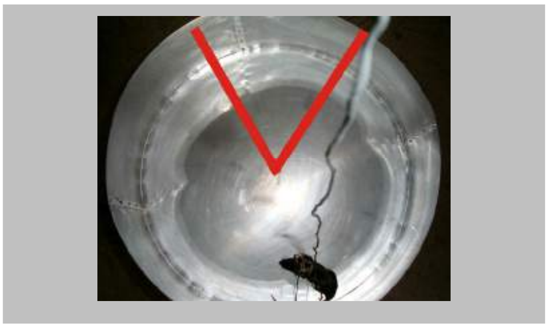


Figure 1. *Overhead photograph of the experimental AAPA arena. The rat walked on the featureless rotating arena and avoided an unmarked sector (denoted by a dashed line) fixed in the coordinate frame of the room. It must be pointed out that the sector is unmarked (defined only in the computer-based tracking system). The rat is connected to a shock-delivering wire, which delivers mild shocks whenever the rat enters the shock sector. Thick arrow denotes the sense of arena rotation (1 rpm).*

Supported by grants MZ ČR-NL-7684-3, GACR No . 309/03/P126 and GACR No 309/03/0715 and by MSMT CR project 1M0002375201.

References:

1. Stuchlik A., Vales, K. (2005) Systemic administration of MK-801, a non-competitive NMDA-receptor antagonist, elicits a behavioural deficit of rats in the Active Allothetic Place Avoidance (AAPA) task irrespectively of their intact spatial pretraining *Behav. Brain Research* **159**(1):163-171.
2. Cimadevilla JM, Wesierska M, Fenton AA, Bures J. (2001) Inactivating one hippocampus impairs avoidance of a stable room-defined place during dissociation of arena cues from room cues by rotation of the arena. *Proc Natl Acad Sci U S A*; **98**:3531-36.

Annotating and measuring meeting behavior

B. van Dijk, A. Nijholt and D. Heylen
Human Media Interaction Group, Department of Computer Science, University of Twente, Enschede, The Netherlands

Abstract
Within the AMI (Augmented Multi-party Interaction) project technologies will be developed that can facilitate human interaction in the context of instrumented meeting rooms, which includes remote participant support and the possibility to browse through past meetings. The project collects data on people engaged in meetings in order to describe, analyze and theorize about meeting behavior and collaborative work. This symposium addresses research methods and techniques used in the AMI project. It deals with the collection of a richly annotated corpus, the design of annotation tools that are focused on specific annotation tasks, methods for user requirements elicitation and browser evaluation, and the observation and interpretation of behavior to obtain recognized behavioral aspects that are visualized and evaluated in a virtual meeting room.

Keywords
Meetings, human interaction, annotation, multimodal, user requirements

1 Introduction
The focus of the Augmented Multi-party Interaction (AMI) project is on meeting research and on assisting participants of meetings with technology. When people meet they interact and hence the project needs to understand the details of human communicative behavior in the realistic and comparatively well-structured setting of a meeting. The overall objective of AMI is to develop technologies that are able to understand and support multimodal, multiparty, human communication in meetings. The primary focus is on the development of meeting browsers that enable users to navigate and search through past meetings and on remote meeting assistants that enable remote participants to have a richer interaction with the meeting or to monitor the meeting while they are engaged in other activities in parallel.

AMI is a multi-disciplinary research project which includes modeling of human-human interaction and group dynamics, multimodal (speech and vision) processing and recognition, content abstraction and human-computer interaction.

2 Analysis of human interactions
Automatic analysis of human interactions is an emerging domain in meeting research [2]. Building upon the findings of social psychologists (amongst others [4]) computer understanding of human interactions is used to accomplish more natural human-machine interaction and computer enhanced human-human communication [3]. For the automatic analysis of natural interactive communicative behavior in meetings, between humans as well as between humans and machines, the availability of large annotated databases of multimodal meeting recordings is essential. The AMI project has a number of meeting rooms equipped with multimodal sensors and computers where meetings are recorded. The collection of recorded meetings is transcribed and annotated on a wide range of properties, ranging from higher level features (e.g. dialogue acts, gestures, emotions, focus of attention) to lower level features (e.g. words, hand and arm movements, facial display elements). Many coding schemes for annotating various aspects of behavior already exist (and have been tested). In the AMI project existing coding schemes are used, improved and extended and new coding schemes for phenomena which have not been investigated before (e.g. emotion) are developed.

AMI builds on previous European projects in the field of multimodal meeting modeling and annotation. Almost the same partners have collaborated in the M4 (Multimodal Meeting Manager) project [6]. Part of the AMI partners participated in the NITE (Natural Interactivity Tools Engineering) project [7] and ICSI (the ICSI Meeting Project [1]) participates in AMI as well. Other projects that work on the computational modeling of meetings and the development of tools for meeting support are for example the Meeting Room project at Carnegie Mellon University [5] and the NIST Meeting Room Project [8].

3 Access to meeting data
The annotated meetings can be used in the meeting browser and remote meeting assistant to enable efficient access to the full multimodal content. Navigation may for instance use identification of meeting participants, identification of focus of attention, or degree of involvement of participants at a particular time. If the assistance can be offered real-time during a meeting, a remote participant that simultaneously is engaged in other activities might be warned when a topic of interest is discussed or the chairman can get support in summarizing the decisions and action points.

Though in theory the possibilities of the use of technologies to facilitate meeting browsing and access seem promising, the actual success depends on the (potential) users. Therefore user requirements elicitation and evaluation of meeting browser concepts and demonstrators of (remote) meeting assistants are increasingly emphasized in the project. One of the main problems in deriving requirements for new technology is that potential users are often unaware of the possibilities. Consequently, when asked directly what they need, users will hardly ever express the need for new technology. In AMI the requirements gathering methods will try to deal with this well-known observation.

4 Outline of the symposium
This symposium presents a selection of methods and techniques used in the AMI project. Though far from complete this selection gives an impression of the wide range of subjects, disciplines and methods AMI deals with.

McCowan et al. give an overview of AMI research themes and discuss corpus design, data collection, annotation and distribution of the AMI meeting corpus.

The creation of large annotated multimodal corpora of meetings is very time consuming. Reidsma et al. stress the need for annotation tools that reduce the amount of work. They discuss the requirements for such tools and propose to design tools that are focused on specific tasks. An overview of properties of annotation problems is presented.

The methodology used to gather user requirements is discussed by Tucker et al. They propose a hybrid approach, combining the examination of current practices with user evaluation of new technology.

Finally, Rienks et al. present the measurement of the behavior of meeting participants. They describe the process from automatic observation of behavioral aspects through interpretations resulting in recognized behavior. They also discuss the use of a virtual meeting room as a research tool. A virtual meeting room can be used for visualization and evaluation of annotations and their interpretations, and for simulation to test models of behavior.

References

1. Janin, A.; Ang, J.; Bhagat, S.; Dhillon, R.; Edwards, J.; Macias-Guarasa, J.; Morgan, N.; Peskin, B.; Shriberg, E.; Stolcke, A.; Wooters, C.; Wrede, B. (2004). The ICSI Meeting Project: Resources and Research. *NIST ICASSP Meeting Recognition Workshop*, Montreal, Canada, May 2004
2. McCowan, I.; Gatica-Perez, D.; Bengio, S.; Lathoud, G.; Barnard, M.; Zhang, D. (2005). Automatic analysis of multimodal group actions in meetings. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **27**, 305-317.
3. McCowan, I., Gatica-Perez, D., Bengio, S., Moore, D. and Bourlard, H. Towards Computer Understanding of Human Interactions. In: Ambient Intelligence, E. Aarts, R. Collier, E. van Loenen & B. de Ruyter (eds.), *Lecture Notes in Computer Science*, **2875**, Springer-Verlag Heidelberg, 235-251.
4. McGrath, J.E. (1984). *Groups: Interaction and Performance,* Prentice Hall, Englewood Cliffs, NJ.
5. Meeting room project Carnegy Mellon University, http://www.is.cs.cmu.edu/meeting room/
6. Multimodal Meeting Manager project, 2002-2005, www.m4project.org
7. Natural Interactivity Tools Engineering project, 2001-2003, http://nite.nis.sdu.dk
8. NIST meeting room project http://www.nist.gov/speech/test beds/mr proj/

The AMI Meeting Corpus

I. McCowan, J. Carletta, W. Kraaij, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, and P. Wellner.
The AMI Project Consortium, www.amiproject.org, J.Carletta@edinburgh.ac.uk

Abstract

To support multi-disciplinary research in the AMI (Augmented Multi-party Interaction) project, a 100 hour corpus of meetings is being collected. This corpus is being recorded in several instrumented rooms equipped with a variety of microphones, video cameras, electronic pens, presentation slide capture and white-board capture devices. As well as real meetings, the corpus contains a significant proportion of scenario-driven meetings, which have been designed to elicit a rich range of realistic behaviors. To facilitate research, the raw data are being annotated at a number of levels including speech transcriptions, dialogue acts and summaries. The corpus is being distributed using a web server designed to allow convenient browsing and download of multimedia content and associated annotations. This article first overviews AMI research themes, then discusses corpus design, as well as data collection, annotation and distribution.

Keywords

Meetings, multimodal, corpus, annotation.

1 Introduction

The AMI (Augmented Multi-party Interaction) project is concerned with the development of technology to support human interaction in meetings, and to provide better structure to the way meetings are run and documented. The project has a number of instrumented meeting rooms that enable the collection of multimodal meeting recordings. For each meeting, audio (coming from multiple microphones, including microphone arrays), video (coming from multiple cameras), slides (captured from the data projector), and textual information (coming from associated papers, captured handwritten notes and the whiteboard) are recorded and time-synchronised. All of these streams are then available to be structured, browsed and queried within an easily accessible archive.

AMI is particularly concerned with the development of meeting browsers and remote meeting assistants, and the component technologies needed to achieve these demonstrators. A meeting browser is a system that enables a user to navigate around an archive of meetings, efficiently viewing and accessing the full multimodal content, based on automatic annotation, structuring and indexing of those information streams. For example, navigation may be enabled using automatic annotations such as speech transcription, identification of participants, and identification of their focus of attention at a particular time. The concept of the meeting browser may also be extended to a remote meeting assistant which will perform such operations in real time during a meeting, and enable remote participants to have a much richer interaction with the meeting.

The development of such meeting browsers depends on a number of technological advances. AMI is extending the state-of-the-art in several areas, including models of group dynamics, audio and visual processing and recognition, models to combine multiple modalities, the abstraction of content from multiparty meetings, and issues relating to human-computer interaction.

As part of the development process, the project is collecting a corpus of 100 hours of meetings using instrumentation that yields high quality, synchronized multi-modal recording, with, for technical reasons, a focus on groups of four people. All meetings are in English, but a large proportion of the speakers are non-native English speakers, providing a higher degree of variability in speech patterns than in many corpora. The corpus aims to benefit a range of research communities, including those working on speech, language, gesture, information retrieval, and tracking, as well as organizational psychologists interested in how groups of individuals work together as a team. This article lists AMI research themes, then describes the design, collection and annotation of the AMI Meeting Corpus.

2 AMI Research Themes

AMI research is structured according to the following themes:

1. *Definition and analysis of meeting scenarios*: To study the type of group, the nature of their interactions, and the means by which their members communicate.
2. *Infrastructure design, data collection and annotation*: To design and install infrastructure to collect data suitable for research of AMI technologies within the context of the defined scenarios.
3. *Processing and analysis of raw multi-modal data*: To research and develop techniques for the processing and analysis of audio, visual, and multimodal data streams from meetings. Specifically, development addresses the following core problems: 1) recognising what is said by participants, 2) recognising what is done by participants (physical actions), 3) recognising where each participant is, at each time, 4) recognising participants' emotional states, 5) tracking what (person, object, or region) each participant is focusing on, and 6) recognising the identity of each participant.
4. *Processing and analysis of derived data*: To research and develop techniques for segmentation, structuring, information retrieval, and summarization of meetings.
5. *Multimedia presentation*: To develop flexible frameworks to access and present streams of multimodal data and metadata.

Progress in these AMI research themes requires a large data set on which empirical observations may be made, and on which technologies may be developed. For example, automatic speech recognition systems require many hours of speech on which statistical models may be trained. A key effort early in the AMI project has thus been the production of the AMI Meeting Corpus, consisting of 100 hours of meetings data suitable for research. The remainder of this article details the design, collection, annotation and distribution of the corpus.

3 Corpus Design

Any study of naturally-occurring behaviour such as meetings immediately encounters a well-known methodological problem: if one simply observes behaviour "in the wild", one's results will be difficult to generalize, since not enough will be known about what is causing the individual(s) to produce the behaviour. [1] identifies seven factors that affect how work groups behave, ranging from the means they have at their disposal, to aspects of organisational culture and pressures coming from the external environment. The type of task the group is trying to perform, and the particular roles and skills the group members bring to it, play a large part in determining what the group does; for instance, if the group members have different roles or skills that bear on the task in different ways, that can increase the importance of some contributions, and be a deciding factor in whether the group actually needs to communicate or can simply leave one person to do the work. Variations to any of the above-mentioned factors will cause the data to change in character, but using observational techniques, it is difficult to get enough of a group history to tease out these effects.

One response to this dilemma is not to make completely natural observations, but to elicit data in a manner which controls as many of these factors as possible. Experimental control allows the researcher to find effects with greater clarity and confidence than in observational work. This approach, well-established in psychology and familiar from some existing corpora (e.g. [2]), comes with its own danger: results obtained in the laboratory will not necessarily generalise to the outside world, since people may simply behave differently when performing an artificial task.

Our response to this methodological difficulty is to collect our data-set in two parts: elicited scenario-driven data, and natural data. The first part (approximately 65 hours) consists of material elicited using a design task in which an effort is made to control the factors from [1]. Since this is the larger part of the data-set, the details of how it was elicited are important, and so we describe it below. The second part (approximately 35 hours) contains naturally occurring meetings in a variety of types, the purpose of which is to help us validate our findings from the elicitation and test their generality. We note that, in fact, a third type of data is also collected, consisting of data elicited using less controlled scenarios than the one which we will describe in this article.

3.1 Meeting elicitation scenario

In our meeting elicitation scenario [3], the participants play the roles of employees in an electronics company that decides to develop a new type of television remote control because the ones found in the market are not user friendly, as well as being unattractive and old-fashioned. The participants are told they are joining a design team whose task, over a day of individual work and group meetings, is to develop a prototype of the new remote control. We chose design teams for this study for several reasons. First, they have functional meetings with clear goals, making it easier to measure effectiveness and efficiency. Second, design is highly relevant for society, since it is a common task in many industrial companies and has clear economic value. Finally, for all teams, meetings are not isolated events but just one part of the overall work cycle, but in design teams, the participants rely more heavily on information from previous meetings than in other types of teams, and so they produce richer possibilities for the browsing technology we are developing.

3.2 Participants and Roles

Within this context, each participant in the elicitation is given a different role to play. The project manager (PM) coordinates the project and has overall responsibility. His job is to guarantee that the project is carried out within time and budget limits. He runs the meetings, produces and distributes minutes, and produces a report at the end of the trial. The marketing expert (ME) is responsible for determining user requirements, watching market trends, and evaluating the prototype. The user interface designer (UI) is responsible for the technical functions the remote control provides and the user interface. Finally, the industrial designer (ID) is responsible for designing how the remote control works including the componentry. The user interface designer and industrial designer jointly have responsibility for the look-and-feel of the design.

For this elicitation, we use participants who are neither professionally trained for design work nor experienced in their role. It is well-known that expert designers behave differently from novices. However, using professional designers for our collection would present both economic and logistical difficulties. Moreover, since participants will be affected by their past experience, all those playing the same role should have the same starting point if we are to produce replicable behaviour. To enable the participants to carry out their work while lacking knowledge and experience, they are given training for their roles at the beginning of the task, and are each assigned a (simulated) personal coach who gives sufficient hints by e-mail on how to do their job. Our past experience with elicitations for similar non-trivial team tasks, such as for crisis management teams, suggests that this approach will yield results that generalise well to real groups. We intend to validate the approach for this data collection both by the comparisons to other data already described and by having parts of the data assessed by design professionals.

3.3 The structure of the elicited data

[4] distinguishes four phases in the design process:

- *Project kick-off*, consisting of building a project team and getting acquainted with each other and the task.
- *Functional design*, in which the team sets the user requirements, the technical functionality, and the working design.
- *Conceptual design*, in which the team determines the conceptual specification for the components, properties, and materials to be used in the apparatus, as well as the user interface.
- *Detailed design*, which finalizes the look-and-feel and user interface, and in which the result is evaluated.

We use these phases to structure our elicitation, with one meeting per design phase. In real groups, meetings occur in a cycle where each meeting is typically followed by production and distribution of minutes, the execution of actions that have been agreed on, and preparation of the next meeting. Our groups are the same, except that for practical reasons, each design project was carried out in one day rather than over the usual more extended period, and we included questionnaires that will allow us to measure process and outcomes throughout the day. In future data collections we intend to collect further data in which the groups have access to meeting browsing technology, and these measures will allow us to evaluate how the technology affects what they do and their overall effectiveness and efficiency. An overview of the group activities and the measurements used is presented in fig. 1.

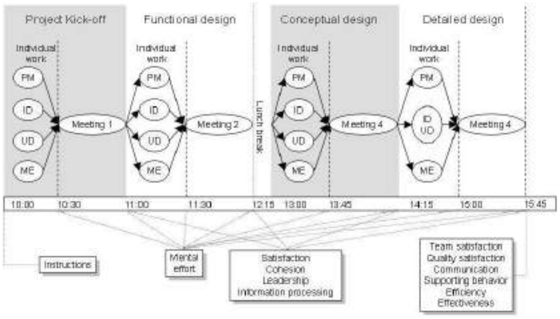


Figure 1: The meeting paradigm: time schedule with activities of participants on top and the variables measured below. PM: Project Manager; ID: industrial designer; UI: user interface designer; ME: marketing expert.

3.4 The working environment

Our collection simulates an office environment in which the participants share a meeting room and have their own private offices and laptops that allow them to send e-mail to each other, which we collect; a web browser with access to a simulated web containing pages useful for the task; and PowerPoint for information presentation. During the trials, individual participants receive simulated e-mail from other individuals in the wider organization, such as the account manager or their head of department, that are intended to affect the course of the task. These emails are the same for every group.

4 Data collection

The data is being captured in three different instrumented meeting rooms that have been built at different project sites. The rooms are broadly similar but differ in shape and construction and therefore in their acoustic properties. In addition, some recording details vary, such as microphone and camera placement and the presence of extra instrumentation. All signals are synchronized by generating a central timecode which is used to replace the timecodes produced locally on each recording device; this ensures, for instance, that videos acquire frames at exactly the same time and that we can find those times on the audio. An example layout, taken from the IDIAP room, is shown in figure 2.

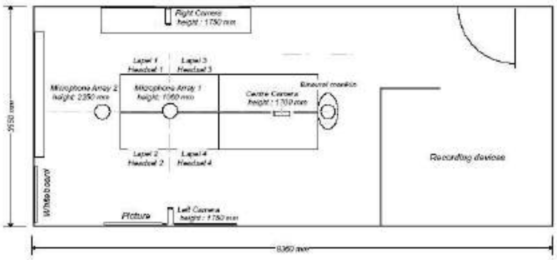


Figure 2: Overhead schematic view of the IDIAP Instrumented Meeting Room.

4.1 Audio

The rooms are set up to record both close-talking and far-field audio. All microphone channels go through separate pre-amplification and analogue to digital conversion before being captured on a PC using Cakewalk Sonar recording software. For close-talking audio, we use omni-directional lapel microphones and headset condenser microphones. Both of these are radio-based so that the participants can move freely. For far-field audio, we use arrays of four or eight miniature omni-directional electret microphones. The individual microphones in the arrays are equivalent to the lapel microphones, but wired. All of the rooms have a circular array mounted on the table in the middle of the participants, plus one other array that is mounted on either the table or the ceiling and is circular in two of the rooms and linear in the other. One room also contains a binaural manikin providing two further audio channels.

4.2 Video

The rooms include capture of both videos that show individuals in detail and ones that show what happens in the room more generally. There is one close-up camera for each of four participants, plus for each room, either two or three room view cameras. The room view cameras can be either mounted to capture the entire room, with locations in corners or on the ceiling, or to capture one side of the meeting table. All cameras are static, with the close-up cameras trained on the participants' usual seating positions. In two of the rooms, output was recorded on Mini-DV tape and then transferred to computer, but in the other output was recorded directly. Figure 3 shows sample output from cameras in the Edinburgh room.

Figure 3: Sample camera views in Edinburgh room

4.3 Auxiliary data sources

In addition to audio and video capture, the rooms are instrumented to allow capture of what is presented during meetings, both any slides projected using a beamer and what is written on an electronic whiteboard. Beamer output is recorded as a time-stamped series of static images, and whiteboard activity as time-stamped x-y co-ordinates of the pen during pen strokes. In addition, individual note-taking uses Logitech I/O digital pens, where the output is similar to what the whiteboard produces. The latter is the one exception for our general approach to synchronization; the recording uses timecodes produced locally on the pen, requiring us to synchronize with the central timecode after the fact as best we can. We intend to subject all of these data sources to further processing in order to extract a more meaningful, character-based data representation automatically [5, 6].

5 Annotation

The data set is being annotated for a range of properties:

- *Speech transcription*, including speaker turn boundaries and word timings,
- *Named entities*, focusing on references to people, artefacts, times, and numbers;

- *Dialogue acts*, using an act typology tailored for group decision-making and including some limited types of relations between acts;
- *Topic segmentation* that allows a shallow hierarchical decomposition into subtopics and includes labels describing the topic of the segment;
- A segmentation of the meetings by the current *group activity* in terms of what they are doing to meet the task in which they are engaged;
- *Extractive summaries* that show which dialogue acts support material in either the project manager's report summarizing the remote control scenario meetings or in third party textual summaries;
- *Emotion* in the style of FeelTrace [11] rated against different dimensions to reflect the range that occurs in the meeting;
- *Head and hand gestures*, in the case of hands focusing on those used for deixis;
- *Location of the individual* in the room and *posture* whilst seated;
- *Location of participant faces and hands* within video frames; and
- *Focus of attention*, i.e. at which other people or artefacts the participants are looking.

For each of these annotations, reliability, or how well different annotators agree on how to apply the schemes, is being assessed. Creating annotations that can be used together for such a wide range of phenomena requires careful thought about data formats, especially since the annotations combine temporal properties with quite complex structural ones, such as trees and referential links, and since they may contain alternate readings for the same phenomenon created by different coders. We use the NITE XML Toolkit for this purpose [12]. Many of the annotations are being created natively in NXT's data storage format using GUIs based on NXT libraries — figure 4 shows one such tool — and others require up-translation, which in most cases is simple to perform. One advantage for our choice of storage format is that it makes the data amenable to integrated analysis using an existing query language.

Figure 4: Example annotation GUI in NXT

6 Distribution

Although at the time of submission, the data set has not yet been released, it will become publicly accessible via http://mmm.idiap.ch. The existing Media File Server found there allows users to browse available recorded sessions, download and upload data in a variety of formats, and play media (through streaming servers and players), as well as providing web hosting and streaming servers for the Ferret meeting browser [13].

Supported by *EU 6th FWP IST Integrated Project AMI (FP6-506811, publication AMI-108)*

References:

1. McGrath, J.E., Hollingshead, A.: Interacting with Technology: Ideas, Evidence, Issues and an Agenda. Sage Publications, Thousand Oaks (1994)
2. Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H., Weinert, R.: The HCRC Map Task Corpus. Language and Speech 34 (1991) 351–366
3. Post, W.M., Cremers, A.H., Henkemans, O.B.: A research environment for meeting behavior. In Nijholt, A., Nishida, T., Fruchter, R., Rosenberg, D., eds.: Social Intelligence Design, University of Twente, Enschede, the Netherlands (2004)
4. Pahl, G., Beitz, W.: Engineering design: a systematic approach. Springer, London (1996)
5. Chen, D., Odobez, J.M., Bourlard, H.: Text detection and recognition in images and video frames. Pattern Recognition 37 (2004) 595–608
6. Liwicki, M., Bunke, H.: Handwriting recognition of whiteboard notes. In Marcelli, A., ed.: 12th Conference of the International Graphonomics Society, Salerno (2005)
7. Lathoud, G., McCowan, I.A., Odobez, J.M.: Unsupervised location-based segmentation of multi-party speech. In: ICASSP-NIST Meeting Recognition Workshop,Montreal (2004) http://www.idiap.ch/publications/lathoud04a.bib.
8. Hain, T., Dines, J., Garau, G., Moore, D., Karafiat, M., Wan, V., Oerdelman, R., Renals, S.: Transcription of conference room meetings: an investigation. In: InterSpeech 2005, Lisbon (to appear)
9. Fitt, S.: Documentation and user guide to UNISYN lexicon and post-lexical rules. Technical report, Centre for Speech Technology Research, University of Edinburgh (2000)
10. Greenberg, S.: Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. In: ESCA Workshop on modelling pronunciation variation for automatic speech recognition, Kerkrade, Netherlands (1998) 47–56
11. Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schroder, M.: 'FEELTRACE': An instrument for recording perceived emotion in real time. In Douglas-Cowie, E., Cowie, R., Schroder, M., eds.: ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research, Belfast (2000) 19–24
12. Carletta, J., Evert, S., Heid, U., Kilgour, J., Reidsma, D., Robertson, J.: The NITE XML Toolkit. (submitted)
13. Wellner, P., Flynn, M., Guillemot, M.: Browsing recorded meetings with Ferret. In Bengio, S., Bourlard, H., eds.: Machine Learning for Multimodal Interaction: First International Workshop, MLMI 2004, Martigny, Switzerland, June 21-23, 2004, Revised Selected Papers. Lecture Notes in Computer Science 3361. Springer- Verlag, Berlin (2005) 12–21

Multi-party Interaction in a Virtual Meeting Room

Rutger Rienks, Ronald Poppe, Anton Nijholt, Dirk Heylen and Natasa Jovanovic
Human Media Interaction Group, University of Twente, Enschede, The Netherlands

Abstract
This paper presents an overview of the work carried out at the HMI group of the University of Twente in the domain of multi-party interaction. The process from automatic observations of behavioral aspects through interpretations resulting in recognized behavior is discussed for various modalities and levels. We show how a virtual meeting room can be used for visualization and evaluation of behavioral models as well as a research tool for studying the effect of modified stimuli on the perception of behavior.

Keywords
Multi-party interaction, virtual environments, behavioral recording, automatic data acquisition techniques

1 Introduction
Meetings play an important part in daily life, they are everywhere. A meeting is when antropomorph entities interact [1] and can be seen as a gathering of thoughts where the exchange and generation of information leads to an enhanced level of knowledge improving the performance of the individuals as well as the group [2]. Generally, a meeting is held in order to move group actions forward through decision making by information presentation and collaboration. Ideally such a meeting proceeds efficiently and effectively, is manageable and accessible afterwards. A meeting can be seen as a series of related interactions amongst participants. The behavior of these participants during the interactions is crucial for the resulting outcome.

The HMI group of the University of Twente has a tradition in research in multimodal interaction with embodied conversational agents, research in computer graphics for virtual environments and machine learning techniques for recognition of higher level features (such as dialogue acts, gestures, emotions) from lower level features (such as words, hand arm movements, facial features).

This paper gives an overview of the current research carried out at the HMI group on automatic observations of behavioral aspects in meetings. The remainder of this paper is organized as follows: Section 2 deals with the typically behavioral meeting aspects. Direct and indirect observable meeting behavior is discussed in Section 3 and 4 respectively. In Section 5 we elaborate on fusion and visualization of several recognized aspects of human behavior.

2 Behavior in meetings
We define behavior as the set of external characteristics that an object exhibits as a response to external or internal stimuli that might be observed, taught, learned and measured demonstrating a competency, skill, ability, or characteristic.

The behavior of meeting participants is generally evaluated relative to social norms and regulated by various means of social control. These norms are unstated and generally unwritten. Typical forms of social norms one might encounter in meetings are that one should not yell or scream, that one should let people finish talking, that one should not start private talks, that one should not whisper and that e.g. '*Ad Hominem*' arguments are not allowed. These social norms or conventions define the shared belief of what is normal and acceptable and hence restricts the people's actions. Operand conditioning plays an important role in their establishment and fulfillment. Violations of norms can be punished with sanctions and violators are considered eccentric or even deviant and are stigmatized.

Sometimes a chairman is appointed and given the authority to manage the meeting process. He or she should make sure that social norms are adhered to, follow a predefined agenda and/or maximize the output of the meeting. This chairman is authorized to perform a set of interventions such as selective turn giving and interrupting. These typical actions are triggered, dependent on the displayed behavior of the participants.

The occurrence of unwanted situations such as a rare event with a large disturbing impact, or the repetitive occurrences of events with a smaller disturbing impact are typical examples that could trigger an intervention. A chairman, or in general every meeting participant could for example intervene if someone is disturbing the meeting process by e.g. continuously repeating him or herself without listening to the other participants.

A problem with human chairmen is that they are usually biased towards certain positions on issues or towards certain persons. Instead of appointing chairmen, one could use systems that are able to both automatically observe the behavior of participants and to automatically regulate the meeting. These systems are potentially cheap in use and can, if the observations are stored, be queried for all kinds of user interests (such as number of turns per speaker, total meeting duration etc.).

Human behavior reveals itself through several modalities over time. Meeting participants exhibit characteristics from the first moment they encounter each other. To explore some of these behavioral characteristics, someone could start by observing the happenings. Simple (possibly automatic) frequency counting could suffice in order to get some first impressions. Ethograms are generally used for this task, they are created using labels based on a predefined behavioral dictionary. For meetings, we could for instance be interested in the turn frequency of the participants. This all seems plausible if one is just willing to observe and nothing else. However, as we want to observe automatically and even more, also want to respond appropriately (e.g. to restore order), we need to know what caused the behavior.

The intentions of the exhibited behavior are related to the individual agenda of the participants and the amount of effort they are willing to put into realizing this. A typical agenda could consist of a number of topics for debate and a possible set of constraints such as a limited amount of available time. This last constraint could cause people to respond more briefly and compliantly.

So what is measured can be attributed to internal stimuli, but what these internal stimuli precisely are, is yet mostly something to guess about. Another aspect occurs when two people are engaged in a conversation. In this case, a balance should be maintained between various levels of communication. Tracy [2] describes these levels of communication as task, or instrumental and face goals. An example of such a balance for a participant is the urge to immediately achieve one's agenda or objective (task goal) on one hand and to act in line with social norms and roles (face goal) on the other hand.

Figure 1 describes the process that we think should take place in a system able to act upon recognized behavior or input in a meeting environment [1].

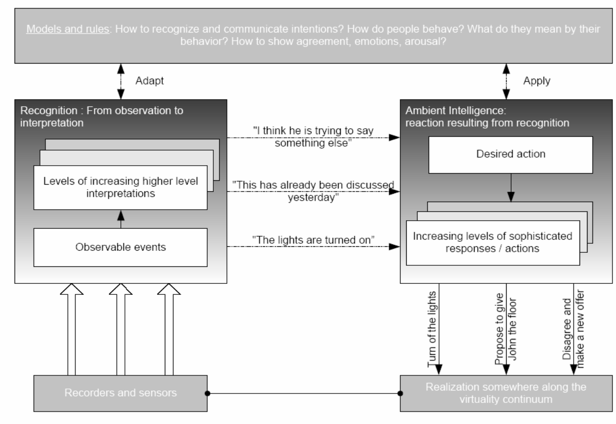


Figure 1. *Observed meeting behavior being recognized, interpreted and regenerated.*

On the lowest level, there are several sensors that observe the meeting. These observations are then to be recognized or labeled automatically using various models of behavior, or behavioral catalogs. The recognized behavior is then transferred to a module that should generate the appropriate actions. This way, various rules of 'accepted behavior' are monitored and in case of violations, the system should decide upon its action. This action selection can be done by either applying predefined actions for each violated threshold or deriving an action real time from which the expected impact is the most appropriate.

To let a system know what it actually observes we make a distinction between direct observable behavior and behavior that is derived or interpreted. We elaborate on this in the next two sections.

3 Directly observable behavior

Directly observable behavior can be immediately measured from the input media without taking the context into account. We measure behavior in each modality (poses, speech, gaze) separately. Speech recognition results in a transcript of what is said but does not yield a semantic interpretation. Body poses can be estimated but are not interpreted until the gesture recognition. Gaze is measured in terms of head orientations. We can obtain these observations through sensors. Within the AMI project, smart meetings rooms are used to collect this data. These rooms are equipped with sensors such as cameras, microphones and in certain scenarios also electronic pens and orientation sensors placed on the participants' heads are used. We will now discuss each of these modalities in turn.

3.1 Body pose estimation

Body pose is an important aspect of behavior. It can be an indicator of involvement (leaning backwards or forwards) and focus of attention (gazing at the speaker, looking at notes). In a multi-party setting, a body pose is estimated for each meeting participant individually. In general, the human body is modeled as segments that are connected with joints. Each joint can have a number of degrees of freedom. A pose is described by a value for each of these degrees of freedom.

Poses can be measured with motion capture equipment but state of the art in computer vision allows for cheap and relatively robust pose estimation without being obtrusive. An overview of recent work on vision-based human motion capture can be found in [4]. Our group has estimated poses from extracted silhouettes and tracked and labeled skin regions [5]

3.2 Gaze detection

Kendon [6] groups the determinants or functions of gaze behavior into five classes: providing visual feedback, regulating the flow of conversation, communicating emotions, communicating attitudes and interpersonal relationships, and improving concentration by restricting visual input. We carried out experiments where electromagnetic sensors were mounted on the heads of each participant in some four party meetings. This information was e.g. used to investigate whether the current speaker could be estimated [7] given the head orientations of all the participants at a specific time. We have shown that in a four person meeting in 79% of the cases a speaker can be predicted on the basis of head orientation of the participants only.

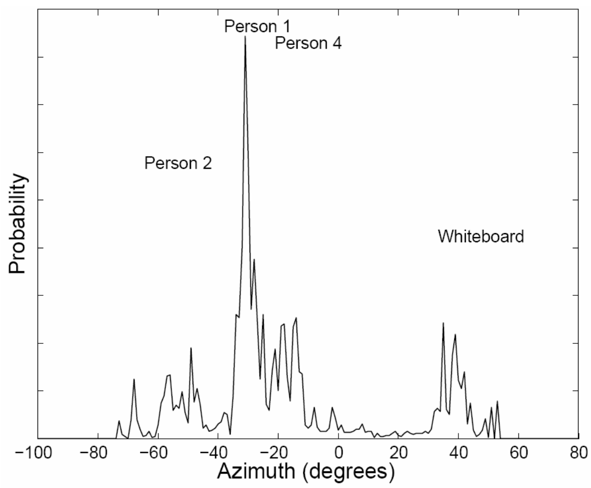


Figure 2. *Distribution of directly observed head orientations along the horizontal (azimuth) plane for person 3 in a four-person meeting setting.*

On the other hand we could derive the participants' attention by creating histograms along their lines of sight. Figure 2 shows the histogram of Person 3 in a particular meeting. A value of zero degrees along the x-axis corresponds with a straight head orientation. The location of the other participants is shown as well as the location of the whiteboard. It follows clearly that person 3 had his head much more often directed towards person 1 when compared to person 4, who was sitting almost opposite. This is information can be useful for addressee detection (See section 4.2).

3.3 Speech recognition

A transcript of what is said by each meeting participant can be obtained by automatic speech recognition (ASR). When issues with multi-speech and background noise can be solved, ASR can generate relatively accurate hypotheses about what is said. These issues are addressed within the AMI project.
Currently, more and more of our research focuses on the extraction of emotional state from speech. This information could be used to detect situations where a participant is enthusiastic or irritated. Other current research topics in this area are automatic topic segmentation and the creation of a meeting dictionary. In contrast to direct observations, specific language models are generally trained to further improve the recognition rates. These sorts of models are essential for the detection of interpreted behavior which is discussed next.

4 Derived or Interpreted behavior

Derived or interpreted behavior is behavior that is not directly observable for a system. Usually recognition systems do not receive more input than an audio and a video signal. To derive relevant aspects of meeting behavior for a chairman for example, one should have models describing how these aspects can be derived from one or more directly observable behaviors. These kinds of models and rules can be extracted by examining large data sets.

4.1 Dominance detection

People who are too dominant in meetings violate the process of collective decision making for which many meetings are intended. We were able to create a system that was able to reach an accuracy of 75% performance when classifying meeting participants as either 'Low dominant', 'Normal Dominant' and 'Highly Dominant' [8]. This classification appeared mainly dependent on the number of floor grabs by a participant and the number of turns someone took during a meeting.

4.2 Addressee detection

Addressing is another important aspect of every form of communication. In small group face-to-face meetings, a speaker can address his utterance to a single individual, to a subgroup of individuals, or to a whole audience. Due to the limitation of the available data for studying subgroup addressing, we limited our current research to the development of a system that automatically identifies whether an utterance of a participant is addressed to the whole group or to just one of the meeting participants. To train our system we used Bayesian Networks. These networks were supplied with a set of utterance, gaze and contextual features in order to automatically identify the participant(s) to whom the speaker is talking. The best performance (82.59 %) was achieved using a combination of all three types of features. [9]

4.3 Argumentation extraction

Once a meeting is over, all that is left are notes that typically contain decisions, action points and perhaps some issues that were left open. Current effort is put in revealing the decisions of a meeting as well as the lines of deliberated arguments in order to provide (automatic) access to representations of conveyed meeting information. The expected behavior of the participants during a discussion can be very interesting for management teams deciding about who to send to which meeting. The desired outcome of this system should eventually provide information about how decisions were made and who brought in which arguments.

An aspect of argumentation that is closely related to behavior is *rhetoric* (How something is told). The ancient Greeks already saw this as the logical counterpart of *dialectic* (What is told). Aristotle defined three main forms of rhetoric: *Ethos* (How the character of a person influences the audience to consider him to be believable), *Pathos (*How emotions affect the message) and *Logos* (How the use of language affects the message). Al these aspects relate to the way people behave or should behave in order to convince someone.

Our ultimate aim is to provide access to representations of conveyed meeting information showing decisions as well as the lines of the deliberated arguments in addition to ordinary meeting notes. Current effort is put into examining what sort of discourse structuring model should be applied to meetings in order to capture the discussion in a useful manner.

5 Regenerating Behavior

If a system has interpreted the observations and recognized one or several behavioral aspects we could use this information for various purposes. Section 5.1 will elaborate upon the visualization of the interpreted data for possible evaluation purposes for both observations and models of behavior. Section 5.2 discusses the possible influence of modified behavioral aspects on the perception of social behavior. For both visualization and evaluation a Virtual Meeting Room (VMR) is used.

5.1 Visualizing recordings

The behavioral aspects that are recognized in each modality can be visualized to evaluate the recognition. One way to do this is to reconstruct the media from which the aspects are recognized, thus to reconstruct the video, the audio and possible other media. This would ideally result in exactly the same video and audio as that we used as input. However, what we recognize is an abstraction of the media itself. For example, from audio we recognize what has been said but we abstract from intonation, pitch and even accent and speed. All these aspects have to be filled in to make our reconstructed media exact duplicates: an impossible task.

Instead, we can make a reconstruction of our recordings where only the semantics are preserved. For example, we could have a visualization where the gestures are performed differently but have the same meaning (e.g. pointing). We will now describe our VMR it enables us to create exactly these kinds of visualizations.

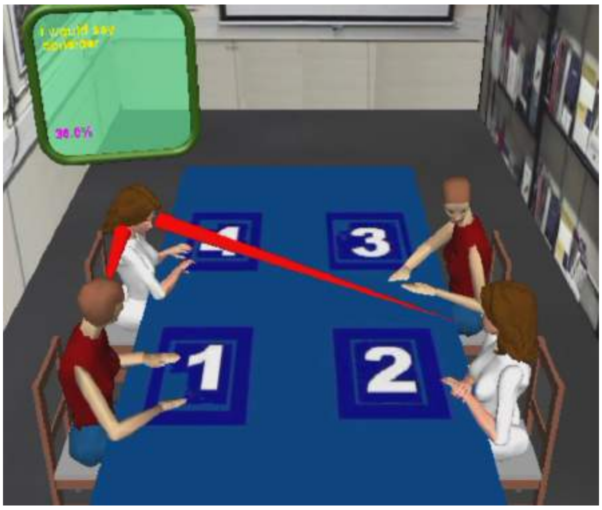


Figure 3. *The virtual meeting room showing gestures, head movements, the speech transcript, the addressee(s) of the speaker and the percentage a person has spoken the entire meeting.*

The VMR is a 3D virtual environment and a replica of a real meeting room, see Figure 3. A benefit of this virtual environment is that different modalities can be controlled. Each modality can be omitted or adapted individually. Another aspect is that the VMR allows visualization of recognized behavioral aspects that in turn can be evaluated. The meeting room can be viewed from all possible angles and it is even possible to be immersed in the VMR by means of a head mounted display (HMD).

5.2 A research environment

The ability of virtual environments to remove or manipulate modalities is very useful for research into social behavior. It allows investigation of which modalities are important for the perception of behavior and what the relation is between modalities. In the experiment we described in Section 3.3, we looked only at head orientations: speech and gestures were omitted.

Another research possibility is to transform modalities [10]. Instead of visualizing in an abstract form what behavior is recognized, the recognized behavior is transformed in order to evaluate the human perception of this behavior. For example, in a gaze experiment, the gaze behavior of the speaker can be mirrored and the perception of the addressee(s) can be measured.

6 Future Work

Meetings can nowadays be assisted with a huge variety of tools and technology, ranging from completely passive objects like a microphone to completely autonomous actors such as virtual meeting participants. Along this line it will not be long before meetings can be held where participants can participate remotely in an immersive virtual meeting space. Next to the virtual replicas of the actual persons, there could be all sorts of active software agents assisting the meeting [11]. Virtual participants (e.g. a virtual chairman) are examples of these possible meeting assisting agents.

When we talk about a virtual chairman, in the most ideal case we would like to have a software-driven virtual representation of a human that is indistinguishable from the representation of a real human. Therefore he or she should not just look like a real human, but also behave like a real human. In an ordinary meeting the chairman has to manage the meeting process in order to maintain the meeting atmosphere, follow a predefined agenda and/or maximize the output of the meeting. Typical mechanisms of the virtual chairman to steer the meeting process are e.g. selective turn giving, interrupting and summarizing.

All these mechanisms depend upon the underlying models of behavioral aspects describing how to interpret observations. These models, including the ones described in this paper will all be crucial for a successful realization. To further improve them in the near future more experiments are to be conducted aiming to find out which various modalities influence the perception and interpretation of social behavior and how they do it.

This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-101).

References:

1. Rienks, R.J., Nijholt, A., and Reidsma, D. (2005). Meetings and Meeting Support in Ambient Intelligence. *In Ambient Intelligence, Wireless Networking, Ubiquitous Computing*, Artech House, Norwood, MA, USA.
2. Moran, T.P., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., Van Melle, W., and Zellweger, P. (1997). I'll get that off the audio: a case study of salvaging multimedia meeting records. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 202–209. ACM Press.
3. Tracy, L.K., and Coupland, N. (1990). Multiple goals in discourse: An overview of issues. *Journal of Language and Social Psychology*, **9**,1–13.
4. Moeslund, T.B., and Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, **81**(3), 231–268.
5. Poppe, R.W., Heylen, D., Nijholt, A., and Poel, M. (2005) Towards real-time body pose estimation for presenters in meeting environments. *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen, Czech Republic
6. Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, **32**, 1–25.
7. Rienks, R.J., Poppe, R., and Poel, M. (2005) Speaker Prediction based on Head Orientations, *Proceedings of the Benelearn 2005*, Enschede, The Netherlands, pp. 73-79
8. Rienks, R.J., and Heylen, D. (submitted). Automatic Dominance Detection in Meetings Using Support Vector Machines.
9. Jovanovic, N., Op den Akker, R., and Nijholt, A. (submitted) Addressing in face-to-face meetings.
10. Bailenson, J.N., Beall, A.C., Loomis, J., Blascovich, J.J., and Turk., M. (2004) Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, **13**(4), 428–441.
11. Ellis, C.(S.), and Barthelmess, P. (2003) The Neem dream. *Proceedings of the 2003 conference on Diversity in computing*, 23–29. ACM Press.

Identifying User Requirements for Novel Interaction Capture

Simon Tucker, Steve Whittaker and Rachel Laban
Department Of Information Studies, University of Sheffield, Sheffield UK

Abstract
Prior to building software to review meeting recordings it is important that we understand both the needs and current practices of meeting participants. Traditionally, such user requirements are collected by either examining current practices or by performing a user evaluation of new technology. The approach taken in the AMI project is to combine the results of the two approaches in order to produce a rich set of user requirements. In this paper we describe the methods used in both approaches, highlighting the motivations behind both the combinatorial approach and the specific methodology applied in each case, proposing a hybrid approach that combines the benefits of each.

Keywords
Meetings Analysis, User Requirements, Practice-Centric, Technology-Centric Requirements Capture.

1 Introduction
The Augmented Multi-Party Interaction (AMI) project is a large multi-disciplinary project building technology to support and enhance the meeting process [1]. Multimedia recordings of meetings are made in special meeting rooms [9], where audio, video, presented slides, whiteboard markings and personal notes are recorded to form a rich record of the meeting. A variety of processing is performed on these recordings (e.g. emotion recognition, automatic text transcription, location analysis etc.) to enhance the record. Finally, a *meeting browser* is used to both gather information and review past meetings.

Focusing on the process of reviewing meeting records, a previous study [10] constructed a taxonomy of meeting browsers. Briefly, browsers were segregated into four categories depending on their *focus*. Browsers which focus on *audio* (e.g. [4]) typically present the user with a visualization of either speaker energy or an automatically produced (ASR) transcript. To overcome the problems of browsing speech recordings, such devices often allow users to mark portions of the recordings and then use this marks as an index for the recording. Furthermore, some devices allow the listener to skim the recording [1], by altering the playback rate and skipping pre-determined portions of the recording. *Video Browsers* present many different components to a user but the focus of such browsers is a video display. Such browsers can make use of video images purely for reviewing the meeting record; however, some [5] use video abstractions to present an overview of the meeting to a user via automatically determined keyframes.

The third category is *Artefact browsers*. Here the focus is placed on non-audiovisual raw data collected during the meeting. For example Lalanne [7] constructed a browser which links the current topic of conversation to the relevant portion in documents being discussed. There have also been several systems which made use of personal notes as an index into a meeting record (e.g. [13]). The final category is *discourse browsers*. The focus here is on derived data, typically related to properties of the discourse such as the speech transcript or participant involvement. An example in this category is the Ferret browser [11] where subject participation and an automatically produced transcript form an index into a rich multimedia meeting record.

While much prior research has explored technologies to facilitate meeting browsing and access, far less effort has been expended on examining user requirements and browser evaluation. In contrast, one focus of the AMI project is producing a methodology for independently assessing meeting browsers [12]. We are also investigating the requirements gathering process.

Whittaker *et al.* [13] produced a system which indexed personal notes against a multimedia meeting record. Two studies were used to collect requirements, the first examining use of audio records and the second assessing the current use of personal notes. In both cases the methodology used was a semi-structured interview with an average of 25 questions being used for each set of interviews.

A limitation of interviews in requirements gathering, especially in a relatively new field, is that the number of subjects that can participate is restricted. To address this problem, Jaimes [6] used an internet questionnaire to assess the current use of tools to review meetings, specifically focusing on the desire for using video recordings to review meetings. Following this study, interviews were used to determine what specific information (meeting participants, location, topics discussed etc.) could be recalled after participating in a meeting.

A third method of generating user requirements was described by Lisowska [8]. Here a number of use cases were generated from a literature survey of typical uses for meeting browsers. Participants were asked to select one of the use cases and then to generate questions they would like to ask a meeting browser system given their choice. An analysis of the responses revealed several properties that a meeting browser should possess but also highlighted a noticeable difference between the questions asked by project members and non-project members.

Cremers *et al.* [3] assessed a new user interface design concept using a task-based protocol. Eight subjects were asked to use a functionality-rich prototype browser to complete a number of tasks which had been determined by the authors. The tasks were chosen to evoke specific behaviors and a questionnaire was used to collect opinions and experiences.

Given the lack of consensus on methods for extracting user requirements for meeting browsers it is clear that there is not yet any agreed technique for performing this task. The approach taken by the AMI project is to combine the results from several studies, using a range of methodologies, to produce a single set of user requirements. The following section outlines the methodological approach, following which the methods used in two of the studies are described in detail.

2 Methodology

There are typically two contrasting approaches to collecting user requirements which could be employed when constructing new technology in the meeting review domain. Firstly, the browser could be designed to support current practices. An example of this *practice-centric approach* is the work of Whittaker [13] who constructed a system to support note-taking by meeting participants. A second *technology-centric* approach would be to create entirely new methods of accessing multimedia records. An example of this approach described above would be the personalized meeting browser [3]. Here, novel methods of accessing multimedia information were presented to users and the performance of these methods was then assessed.

There are strengths and weaknesses of the two approaches. The practice-centric approach analyzes what tools are currently used and discusses the use of these tools with a user community, allowing both the advantages and problems with these tools to be identified. Doing this can identify a suitable platform on which to build technology. A different method is required to support the *technology-centric* approach. In this case users can be shown the new means of accessing meeting records and their reactions to both the methods and the presentation of the methods are collected and analyzed. The advantage of the practice-centric approach is that it supports genuine user tasks with technology users can understand. However it is limited to supporting current activities and the status quo, rather than using technology to transform how things are done. In contrast, the technology-centric approach may generate radical new technologies but at the risk of failing to ground these in current practice, making them hard to use.

Typically one of these approaches is taken in order to collect user requirements. This has the problem that each study is liable to overlook possibilities presented by its counterpart. Thus, when examining current practices users will be largely unaware of the technological potential available and will therefore overlook solutions which new technology would enable. Furthermore, when exposing users to new technology there is the danger that too many novel approaches may be presented, many of which may be unnecessary.

To overcome these contradictions, both of these approaches are used in the AMI project. We have carried out an analysis of current meeting practices and are using the results of that study to construct prototype meeting browsers. These browsers will be supplemented with novel components and presented to users in order to collect reactions to both the browser design and the functionality of the new technology. By constructing the user requirements in this way we are able to capitalize on the advantages that each method brings.

3 Methodology

The following sections describe the methodology used in both studies. We start by describing the procedure used to determine current practices and then outline a technology-centric approach. Finally we describe a study to collect user reactions to functional and user interface designs.

3.1 Practice-Centric Requirements Capture

Information is exchanged and recorded in meetings in a wide variety of ways. Most meeting participants take personal notes and a public record (referred to as minutes), is often taken during the course of the meeting. In addition to this, information is often exchanged following a meeting via either email or informal post-meeting discussions conducted away from the meeting location.

There are also several analytic methods available for collecting and analyzing these records and exchanges. Personal notes and public minutes can be collected from meeting participants. Furthermore a recording can be made of the meeting and a transcript of the discussion can be made. Emails can be examined and interviews can be used to determine complaints and problems with current practices. It is difficult, however, to collect information about post-meeting discussions due to their ad-hoc nature. A problem with collecting this data is that meetings will often contain highly sensitive discussions, as will any of the associated data that the meeting generates and therefore procedures for dealing with privacy are required before data collection can begin. A secondary problem associated with the data is that it can be difficult to assess the quality and failings of personal and public records without being fully involved in the meeting.

Our study examined the meeting recording practices of two firms, one responsible for national and international mail deliveries and the other for supplying software services. In both cases we studied a core team through a series of meetings – thus allowing us to examine how information presented in earlier meetings was used and followed up on in later meetings. The core teams had 5 and 7 members and the meetings analyzed had between 3 and 16 participants. In the delivery firm, meetings were used to report information and co-ordinate the teams; in the software firm, meetings were largely between customers and suppliers. Both sets of meetings were task-oriented rather than being about idea generation – this being appropriate for the type of meetings that the AMI browser seeks to support. In both cases participants were familiar with each other, having worked together for over 6 months.

We collected a range of information from each meeting and from each meeting participant. Primarily we recorded each meeting onto cassette tape and produced a manual transcription of the meeting; we also collected meeting minutes and, where possible, copies of personal notes were also taken. To address any concerns participants had regarding their privacy we ensured that all subjects were aware that the meetings would be recorded and transcribed and were given a brief overview of the purpose of the study. Participants were able to request that the recorder be turned off at any stage and were also given the option of being able to edit and review the recording before the transcript was made. The names referred to in the meeting were converted into initials and the resulting analysis focused on high level aspects of the meeting rather than the fine details.

Further to the raw data we collected from the meeting, face-to-face interviews were conducted which examined the note-taking practices of meeting participants and their ability to remember information after the meeting had occurred. A semi-structured interview protocol was used as it allowed us to ensure that the interviewee covered all the topics we were interested in but it also allowed us to follow up any interesting answers. We also observed participants during the meeting – recording when they took notes, when they participated in the meeting etc.

Following the data collection six different analyses were used to process this raw data. Firstly, we examined the interviews and compared answers to specific questions. This allowed us to determine what parts of the meeting individuals thought were important to remember and why this was so. Secondly, we asked participants several questions based on the observations made during the meeting – combining the answers to these questions with the observations themselves allowed us to determine how individuals take notes and their behaviour relating to this.

Thirdly, we asked participants to read through the transcripts and to highlight points which they felt were important. We then compared the selected statements for each participant and analyzed whether their role in the meeting had an effect on the statements they chose as being most important. Fourthly, we examined the personal notes taken by each participant in order to establish what types of information individuals were recording and how they were making this record. We also used the notes to make an estimate of how much time the participant spent taking notes during the course of the meeting. We were then able to compare this estimate with our observations in order to determine the relationship between the amount of notes taken and the level of participation.

We also asked participants to select the 10 most important areas that had been discussed and then rank them in order of importance. This selection and the associated ranking were then compared to the personal notes made by that participant; we then formulated questions about any discrepancies between the two. These questions were asked via email since it allowed participants to answer in their own time. Finally, we made a further analysis of participation in the meeting by analyzing the time spent talking, the number of times an individual talked and the length of the notes they took. This, again, was used to address the issue of the effect that note taking has on being able to participate in the meeting.

The combination of the analyses allowed us to construct a good picture of not only how meeting records are currently constructed, but it also allowed us to determine any problems associated with such records. Furthermore, this study enabled us to determine how the various meeting records were used and what their strengths were. The second phase of user requirements capture for AMI will assess both the functional and user interface designs through exposure to new technology.

3.2 Technology-Centric Requirements Capture

The first stage of our study has shown how current meeting records are produced and used by both individuals and the group as a whole. The second phase of our study is intended to examine user reactions to new technology. We have some experience in this field, having previously carried out an field test of a browser evaluation test (BET) [12].

The BET is designed to allow browser designers to independently assess their browsers and, furthermore, to allow them to compare their browser to the field without having access to other browsers. Briefly, the BET consists of two stages. In the initial stage a set of judges watch a multimedia meeting recording and make a number of *observations of interest*. An observation of interest is a true-false pair of statements relating to a point in the meeting which the judge has determined that the participants found interesting. For example, the pair "The budget was £1m" / "The budget was £5m" may be an example observation of interest. In the second stage of the experiment subjects make use of novel browsers and are presented with a series of one half of the observations of interest. They must then determine whether they have been given the true or false part.

The BET is good at independently assessing and comparing novel meeting browsers. One possibility for building a user requirements study would be to make use of the BET protocol, however there are problems with this proposal - furthermore, these problems also face any study focusing on the assessment of novel technology.

A key problem of using the BET is that it only assesses the ability of the browser to identify specific facts from the meeting. The literature described in Section 1 and our own studies shows that users have other goals too; instead, they might wish to get an overview of the events of the meeting. Furthermore, since the observations are generated with a basic media player there is the problem that the generated observations may not account for novel technology. So for example, if judges are not aware that emotions can be determined from meeting recordings they are unlikely to make observations which relate to the emotional content of the meeting. Thus there is a limitation in the current version of the BET that reduces the range of questions which can be asked of meeting browsers.

3.3 Hybrid Requirements Capture

Although not ideal for the purposes of collecting user requirements, our experience with assisting the design of the BET allowed us to design a similar protocol for assessing user reactions to novel technology. We split the study into two stages – the first is a BET-like series of tasks which subjects undertake on real meetings data before being questioned on their experiences. In the second stage we present paper-based designs of new interface and technology concepts to users in order to collect their reactions.

In the first stage we make use of two distinct browsers. The first is a low-tech browser consisting of video, audio, an automatically generated transcript and presented slides. The purpose of using the low-tech browser is to determine problems that are apparent with the current state of browser components. This also enables us to assess any user interfaces issues arising from the use of the low-tech browser. The second browser is a high-tech browser; in addition to the components presented by the low-tech browser this will also have components for summaries, emotional content, gesture recognition and indexes of group decisions and individual actions.

To ensure subject engagement we ask them to imagine that they are joining a project in the middle of its development. They are to use to the browsers to familiarize themselves with the current progress of the project, where the project is going and what their role will be in the group. They are then given two tasks. The first is a general task – namely to acquaint themselves with the project. The purpose of this task is to assess how well the novel browser supports the general review of the meeting record. Since we propose to run several evaluations using this protocol a brief questionnaire and an analysis of the application logs will be used to assess the performance of the user in carrying out this task.

The second task is to answer four specific questions which we have manually determined for each meeting. Four questions are asked – the answers to which either require some understand of the meeting in its entirety or can be answered from specific points spread across the duration of the meeting. Each question focuses on a different aspect of the meeting – the chosen aspects are the same across meetings in order to ensure some level of commonality across the meeting set. We ask one question relating to decisions made during the meeting and we also ask one about actions which have been assigned to individuals. Furthermore, we ask a question about emotions – either a specific emotional response or to determine the emotional state of an individual or the group throughout the duration of the meeting. Finally, we ask one question about a visual event which occurred in the meeting – for example asking if someone left the meeting at any time. The classes of questions were determined by an analysis of the literature described above and also an awareness of the types of data being examined and produced by the AMI project.

Following the completion of both tasks we present the subject with a questionnaire asking them which browser components they found useful and, specifically which ones were useful for approaching each task. We then present the subject with a number of abstract paper-based browser designs and assess their reactions to both user interface designs and to proposed novel functionality. The advantage of this approach is that we are able to assess a number of different designs and functionality without the associated of cost of building the technology to perform the correct actions.

4 Conclusion

We have described two different approaches to collecting user requirements for the AMI meeting browser. The first practice-centric approach was an examination of current meeting recording practices, carried out in two companies. The second technology-centric approach evaluated novel technologies. We contrast these with a hybrid technique which examines user reactions to novel technologies using both prototypes and abstract paper based designs, based on tasks derived from practice-centric user observations.

Traditional user requirements studies have focused on either examining current practices or exposing users to new technologies. It can be argued that by carrying out both studies and combining their results we are able to produce a more fully-featured set of requirements. The first phase of our study allowed us to assess both the failings of current meeting records and also the desirable properties of these records. The second phase of our study will assess new technology for meeting browsers. Working prototypes are used to assess the functionality and interface requirements of the meeting browser, whilst paper-based design ideas can ascertain the best means of presenting novel data captured from the meeting record.

This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, Publication 106).

References:

1. AMI Project, http://www.amiproject.org
2. Arons, B. (1997). SpeechSkimmer: A system for interactively skimming recorded speech. *ACM Transactions on Computer-Human Interaction*, 3-38.
3. Cremers, A.H.M, Hilhorst, B. and Vermeeren, A.P.O.S. (2005). "What was discussed by whom, how, when and where?" personalized browsing of annotated multimedia meeting recordings, *HCI International*, Nevada, USA, 22-27th July.
4. Degen L., Mander, R., Salomon, G. (1992). Working with audio: integrating personal tape recorders and desktop computers. *Proceedings of CHI '92*, 413-418.
5. Girgensohn, A., Borczky, J., Wilcox, L. (2001). Keyframe-based user interfaces for digital video. *IEEE Computer*, 61-67.
6. Jaimes, A., Omura, K., Nagamine, T., Hirata, K. (2004). Memory cues for meeting video retrieval. *Proceedings of CARPE '04*.
7. Lalanne, D., Sire, S., Ingold, R., Behera, A., Mekhaldi, D., Rotz, D. (2003). A research agenda for assessing the utility of document annotations in multimedia databases of meeting recordings. *Proceedings of 3rd International Workshop on Multimedia Data and Document Engineering*.
8. Lisowska, A. (2003). Multimodal interface design for the multimodal meeting domain: preliminary indications from a query analysis study. *IM2.MDM Report 11,* Universite de Geneve.
9. Moore, D., Guillemot, M., Masson, O., Formaz, F., Flynn, M., Gatica-Perez, D., McCowan, I., Wellner, P., (2004). A technical overview of IDIAP infrastructure for the acquisition and distribution of multimodal meeting data. Proceedings of MLMI '04.
10. Tucker, S. and Whittaker, S. (2005). Accessing multimodal meeting data: systems, problems and possibilities. *Lecture Notes In Computer Science*, **3361**, 1-11.
11. Wellner, P., Flynn, M., Guillemot, M. (2005) Browsing recorded meetings with ferret *Lecture Notes In Computer Science*, **3361,** 12-21**.**
12. Wellner, P., Flynn, M., Tucker, S., Whittaker, S. (2005), A Meeting Browser Evaluation Test, *Proceedings of the Conference on Human Factors in Computing Systems.*
13. Whittaker, S., Hyland, P., Wiley, M. (1994). Filochat: Handwritten Notes Provide Access To Recorded Conversations. *Proceedings of CHI '94*.

Designing focused and efficient annotation tools

D. Reidsma, D. Hofs, N. Jovanović
Human Media Interaction research group, Centre for Telematics and Information Technology, Enschede, Netherlands

Abstract
The creation of large, richly annotated, multimodal corpora of human interactions is an expensive and time consuming task. Support from annotation tools that make the annotation process more efficient is required, especially if the annotation effort involves really large amounts of data. Therefore we investigated how different properties of specific annotation tasks can have an impact on the design of a tool focused on that general class of tasks. In this paper we present our view on the considerations that should drive the design of new tools geared to specific tasks. The main dimensions that we consider are: observation vs. interpretation, explicit and implicit input layers, segmentation, feedback, constraints, relations and the content of the annotation elements.

Keywords
large corpora, annotation, tools, reusable design

1 Introduction: Why design focused annotation tools?
Shriberg et al. [12] report an efficiency of 18xRT on annotation of dialog act boundaries, types and adjacency pairs on meeting recordings (i.e. annotation takes 18 times the duration of the video). Simple manual transcription of speech usually takes 10xRT. For more complicated speech transcription such as prosody 100-200xRT has been reported in Syrdal et al. [13]. The cost of syntactic annotation of text (PoS tagging and annotating syntactic structure and labels for nodes and edges) may run to an average of 50 seconds per sentence with an average sentence length of 17.5 tokens (cf. Brants et al. [1], which describes syntactic annotation of a German newspaper corpus). As a final example, Lin et al. [8] report an annotation efficiency of 6.8xRT for annotating MPEG-7 metadata on video using the VideoAnnEx tool (correction of shot boundaries, selecting salient regions in shots and assigning semantic labels from a controlled lexicon). It may be obvious that more complex annotation of video will further increase the cost.

Many large projects face the challenge of annotating a really large amount of data for many different modalities. Given the amount of work needed for each hour of recorded data it still seems useful to invest some effort in designing annotation tools that reduce the work. Tools which are highly efficient for one annotation task are not necessarily so for other tasks. Making a single tool for all conceivable annotations results in a monolithic, unwieldy tool. Nevertheless, designing annotation tools from scratch for each different annotation task is not efficient. This paper presents some ideas which may help reuse design and implementation considerations, which have been applied to the development of several new tools.

2 User types
The users of annotation tools may be divided into the groups described below [2].

- **Annotators:** Users who need a tool for their annotation task. They should not be bothered about data representations, internal design, or API design. A tool should help them work as quickly and efficiently as possible.
- **Corpus Consumers:** Users who want to use annotated data for all kinds of reasons, e.g. theory testing, evaluation and training of models, finding relations between phenomena. They have needs for querying and browsing annotated data.
- **Corpus developers**: Users responsible for corpus design and maintenance (e.g. design of new annotation schemas or altering existing ones, understanding of data representation supported by the tool and mapping of their data to the existing structures).

Since this paper is mostly about designing tools that help reduce annotation effort, we will focus only on the *annotators* in the rest of this paper.

3 Requirements for annotation tools
In the course of this work we also collected tool requirements from a few selected reviews [5,6,7,11]. These reviews together outline most of the criteria used in many papers to rate existing annotation tools or to design new tools. The chosen evaluations are performed from different perspectives reflecting different evaluation goals. The aim of the ISLE Natural Interactivity and Multimodality Working Group report [6] is to provide a survey of world-wide tools which support annotation of natural interactivity and multimodal data. As a result it outlines the most important overall user needs reflected in the tools and projects which created them. The aim of the evaluations presented in [5,7,11] is to select a tool or set of tools based on analysis of research project needs. The reviews follow the same evaluation procedure which consists of two steps. First, based on the analysis of the project needs, a list of requirements for annotation tools is defined (e.g. simplicity, quality assurance, compatibility with other tools, customization of the annotation scheme, etc.). Next, the 'evaluation criteria' are derived. Table 1 lists a reduced version of the collected criteria (the full version can be found in [10]).

The requirements for statistical data analysis and display are supported by software packages that a new tool would hardly displace. Furthermore, the requirements for input/output flexibility, flexibility in coding schemes and querying annotated data are covered by using a stand-off XML data format with a good API such as AGTK, NXT or ATLAS. In this paper we focus only on the annotators as target group and the requirements related to the efficiency of creating annotations such as an easy-to-use interface, marking, audio/ video interfaces, the annotation process and visualization.

4 Characterizing Annotation Problems
Different annotation problems, such as transcription, video labeling and text mark-up, each have their own properties.

The properties may have an impact on how the requirements from the previous section are to be interpreted and fulfilled. This section gives an overview of those properties and discusses how they can influence the design of efficient tools for annotation.

4.1 Observation vs interpretation

A specific layer of annotation in a corpus may pertain to *direct observations* of events in the physical world, such as certain movements, speech or gaze directions, or to *interpretations* of those observations, such as emotional states, dialog acts or complex semantic annotations. The interpretations involve deducing information about the internal mental state of the persons involved in the observation, about their beliefs, desires and attitudes [9].

Interpretation takes a lot more time than straightforward observations. Aiming for a real time coding process may be sensible while one is coding observations. When coding interpretations, this may be less feasible. If an annotation is part observation and part interpretation, it may be a good idea to split it up in two different tasks.

4.2 Input layers

Every annotation layer is based on certain sources of input. The most basic layers are based only on the audio and/or video (labeling of head nodding, transcription, hand tracking). More complex layers may also be based on other layers (e.g. dialog acts based on transcriptions, interpretation of gestures for their communicative function). Sometimes the reference from annotation elements to elements in input layers is made explicit, such as dialog acts referring to text fragments. Sometimes this relation is implicit, such as the relation between dialog acts and video or audio: though the explicit input is the speech of the participant, the video and audio offer valuable input for determining the exact dialog act (facial expression, intonation, etc).

Explicit and implicit input layers determine what should be displayed in the tool. An annotation tool should preferably display only the explicit and implicit input layers and the created annotations. Anything else would be a distraction. The explicit input layer should be displayed in a way that clearly shows its relation to the created annotation elements. The explicit input layers also influence the *selection mechanisms* of the tool.

4.3 Segmentation of the input layer

The segmentation properties of an annotation have a major impact on the design of the GUI. The segmentation determines which fragments of the explicit input layer(s) an annotation element can refer to. A list of possible characteristics of the segmentation is given below.

- Segments may or may not relate to overlapping parts of the explicit input layers.
- Segments may or may not interleave with each other.
- Segments may or may not be discontinuous.
- Each segment may be annotated with only one, or more than one, element.
- The segmentation may or may not fully cover the input layer.
- The size of segments differs from problem to problem: single words, sentences, arbitrary fragments, etc.

These properties determine how the selection mechanism should be designed, but also whether semi-automatic support is possible for segmentation and selection. If, for instance, a tool is being developed for manual coding of part-of-speech, the segmentation properties suggest that the tool might perform segmentation automatically and present the segments (i.e. words) one by one for labeling. For dialog acts, the segmentation is not obvious, so it should be done by the annotator.

4.4 Labeling vs complex information

Some annotation layers contain annotation elements that are just labels from a (possibly very complex) set or ontology. Other annotation layers have more complex structures as their constituent elements, such as the multiple labels in MRDA [4].

When the information per annotation element consists of a single label, one can for example decide to map labels to keystrokes or a set of GUI buttons. If the information is more complex, a separate edit panel for annotation elements is probably more suitable.

4.5 Relations between annotated elements

Some annotation elements may define relations between/to other annotation elements. As far as the annotator is concerned, there are two types of relations. One of the related elements may be considered an attribute of the other element, or their relation may be seen as an annotation element in its own right, stored in a separate layer. Depending on how complex the relational structures are, one may consider making the relational coding a separate task.

4.6 Feedback

There are several types of feedback: feedback about the contents of the annotation as the video of the observation is replayed, feedback about which elements and values are currently being annotated and feedback about the 'whole annotation up till now'. All three types of feedback should be present in an annotation tool, though they need not necessarily be given by the same components.

4.7 Constraints

There may be constraints on element contents and relations (e.g. an answer belongs to a question, certain combinations of tags are not allowed). The tool may help maintain integrity by enforcing those constraints, so limiting the choices of the annotator.

4.8 Default values

A special type of 'constraint' is a default value. If a default value for a certain attribute can be defined, the tool can support faster coding by pre-filling the attribute. Syrdal et al. show that, in some cases, default suggestions can speed up manual labeling without introducing too much bias in the annotation [13].

5 Designing annotation tools

Using the characteristics described in the previous section, several components and modules were developed that can be used to develop new annotation tools targeting specific annotation problems with only little extra effort. Two classes of annotation tasks were taken into consideration: discourse labeling and labeling of events in audio or video with a time and duration. Due to space constraints only one of these tasks is discussed partially. It can be seen as an illustration of how the dimensions presented above can help with the design of annotation tools.

The actual modules and tools were developed using the Nite XML Toolkit, an open source toolkit for heavily annotated corpora [3]. The actual annotation tools which have been developed in the course of this work are freely available as part of the Nite XML Toolkit, downloadable from http://www.sourceforge.net/projects/nite

5.1 Discourse labeling

Many annotation tasks involve the labeling of discourse. In the AMI project, in the context of which this work has been done, this means labeling the transcriptions of multi-party dialog. Examples of such tasks are named entity annotation and dialog act annotation. Annotations of this class share several properties along the dimensions described above, and may be different in other properties.

Input layers. For this class of problems, one explicit input layer is the transcription. Other explicit or implicit input layers may contain all kinds of codings which have already been defined on top of that transcription. Therefore it would be useful to have a customizable view that can show a multi-party discussion (dependent on the corpus structure), enhanced with mark-up from existing annotations. Since there are many ways in which the existing annotations can be related to the transcriptions, the mark-up should be highly configurable.

Feedback. Feedback about the annotation can be provided as soon as the transcription view allows visualization of existing mark-up.

Segmentation. With respect to segmentation, discourse labeling tasks may have widely diverging properties. Different types of discourse labeling involve single words or phrases, may or may not span multiple transcription segments, may or may not contain *partial* phrases, may or may not allow overlap between segments, etc. It is very useful if a transcription visualization component is able to support these different types of segmentation explicitly, by allowing or disallowing certain types of selection. The module that has been implemented allows a broad range of selection restrictions to reflect this.

Relations. Relational codings defined on top of a discourse labeling are very common. A module for relational annotation was therefore developed which does not depend on the exact structure of the discourse elements that are related but which nevertheless allows visualization which is integrated with the marked-up transcription.

Result. The result of such reflections is an annotation tool for discourse labeling which can be adapted to many different tasks, either through configuration settings, or, for more complex adaptations, through the extensive API of the modules and components. The tool is centered on a configurable transcription view and an audio/video viewer with time aligned highlighting of the transcription. Using a number of configuration settings, the tool can be used on any corpus defined in the NXT stand-off data format. The tool has already been used for annotation of RST information, dialogue acts and named entities.

6 Conclusions

In this paper we take the position that, to meet the annotation requirements for very large corpora, it may be necessary to develop annotation tools that are specialized to reduce the time effort for creating the annotations. We present our view on the considerations that should drive the design of new tools geared to specific annotation tasks. The main dimensions that we consider are: observation vs. interpretation, explicit and implicit input layers, segmentation, feedback, constraints, relations and the content of the annotation elements. Finally, we discuss one example class of annotation problem for which we have designed actual annotation tools, modules and components, using information about these dimensions.

This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-100).

References:

1. Brants, T; Skut, W; Uszkoreit, H (2003). Syntactic annotation of a German newspaper corpus. *Building and using Parsed Corpora*. Kluwer, Dordrecht (NL).
2. Carletta, J; Isard, A; Klein, M; Mengel, M; Moller, M.B (1999). The MATE Annotation Workbench: User Requirements. *Proc of ACL-99 Workshop 'Towards Standards and Tools for Discourse'*.
3. Carletta, J; Evert, S; Heid, U; Kilgour, J; Reidsma, D; Robertson, J; The NITE XML Toolkit. (submitted).
4. Dhillon, R; Bhagat, S; Carvey, H; Shriberg, E (2003). Meeting Recorder Project: Dialog Act Labeling Guide. *Report, ICSI Speech Group, Berkeley, USA*.
5. Dipper, S; Goetze, M; Stede, M (2004). Simple Annotation Tools for Complex Annotation Tasks: an Evaluation. *LREC2004 Workshop on XML-based Richly Annotated Corpora*, Lisboa, Portugal.
6. Dybkjaer, L; Berman, S; Kipp, M; Olsen, M.W; Pirelli, V; Reithinger, N; Soria, C (2001). Survey of existing tools, standards and user needs for annotation of natural interaction and multimodal data. *ISLE NIMM Working Group Deliverable D11.1*
7. Garg, S; Martinovski, B; Robinson, S; Stephan, J; Tetreault, J; Traum, D.R (2004). Evaluation of Transcription and Annotation tools for a Multi-modal, Multi-party dialog corpus. *LREC2004*, Portugal.
8. Lin, C.Y; Tseng, B.L; Smith, J.R (2003). Video Collaborative Annotation Forum: Establishing Ground-Truth Labels on Large Multimedia Datasets. *Proc of the TRECVID2003*, Gaithersburg, Md., USA.
9. Reidsma, D; Rienks, R.J; Jovanović, N (2004). Meeting modeling in the context of multimodal research. *MLMI'04*, Martigny, Switzerland.
10. Reidsma, D; Jovanović, N; Hofs, D (2004). Designing Annotation Tools based on the Properties of Annotation Problems. *Report, Centre for Telematics and Information Technology.*
11. Rydeman, B (2003). Using Multimodal Annotation Tools in the Study of Multimodal Communication Involving Non speaking Persons. *Report, Goteborg University.*
12. Shriberg, E; Dhillon, R; Bhagat, S; et al (2004). The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. *HLT-NAACL SIGDIAL Workshop*. Boston, USA.
13. Syrdal, A.K; Hirschberg, J; McGory, J; Beckman, M (2001). Automatic ToBi prediction and Alignment to speed manual labeling of Prosody. *Speech communication,* 33, 135-151.

CRITERIA	1	2	3	4	Annotator?
Portability		X	X	X	YES
Can the tool be used on different platforms? Does it require any additional packages? [2] Is it easy to install? [4]					
Flexible architecture	X				
Allows extension of the tool by adding new components.					
I/O flexibility	X	X		X	
What are the tool's input formats? Does the input data need any preprocessing? Is the output format compatible with other tools? Are there converters from/to other formats provided? Can annotation scheme be imported/exported and in which format?[4]					
Robustness and stability	X		X		YES
Is the tool robust, stable and fast?					
Audio/video interface		X			YES
Does the tool offer an easy-to-use method for playing audio and/or video and for segmenting it? Does the tool support handling large media files? Does the tool support playing back the media file aligned with an annotation element?					
Flexibility in coding scheme	X	X	X	X	
Does the tool support easy addition of a new coding scheme or altering of the existing one? [1,2,3] Does the tool allow user to restrict format and/or the content of annotation data? [4] Can annotation levels be defined as obligatory or optional? [4] Can tag sets be specified? Can tag sets be structured? [4] Are annotation levels and tag sets defined within the tool or by external files? [4]					
Easy to use interface	X	X	X	X	YES
The interface should support users as much as possible, be intuitive and based on standard interfaces conventions					
Learnability				X	YES
Is the tool easy to learn?					
Attractiveness				X	YES
Does the user enjoy working with tool?					
Transcription support	X			X	YES
Can the tool be used for speech transcription?					
Marking	X	X	X	X	YES
Does the tool support annotations at different levels, of different modalities and annotations across levels and modalities? How much can the tool mark (e.g. just words or group of words; entire sentences or segments of sentences)? Does it allow the marking of discontinuous fragments? [2] Does the tool support simultaneous annotation for several persons? [3]					
Meta-data		X		X	YES
Does the tool support 'meta-data' such as annotators' comments and notes referring to annotations or relating to the entire document?					
Annotation process				X	YES
Does the tool support some kind of (semi) automatic annotation? Does the tool support selection-based annotation where only appropriate the tags are presented to the user?					
Visualization	X	X	X	X	YES
Scope: Is the annotated information visible for all annotation elements or only the currently active element? Style: How are the annotated elements presented? [4] Can the user change visualization dynamically? Can the user define visualization? [1,4] Does the tool support synchronized view of different annotation layers and of different modalities? [1] Does the tool have a large display to show the current works and corresponding data in a clear manner? [2]					
Documentation		X		X	YES
Availability and quality of user manual; on-line help					
Querying, extraction	X			X	
Does the tool support (simple or powerful) search mechanisms and an interface to the search tool? Are the results presented in an intuitive and easy-to use way?					
Data analysis	X				
Does the tool support (statistical) analysis of annotated data?					

1: Dybkjaer et al. [6]
2: Garg et al. [7]
3: Rydeman [11]
4: Dipper et al. [5]

'Annotator?' marks whether the requirement is relevant for the annotator.

Table 1. Collected requirements for annotation tools.

A tool for the analysis of ordinal scale data: measuring consensus, agreement, and dissent

W. J. Tastle[1], M. J. Wierman[2]
[1] *Department of Business Administration, Ithaca College, Ithaca, New York, USA*
[2] *Department of Computer Science, Creighton University, Omaha, Nebraska, USA*

Abstract
This paper describes a new measure of dispersion used as an indication of consensus and dissention. Building on the generally accepted Shannon entropy, this measure utilizes a probability distribution and the hierarchically ranked categories of the distribution to yield a value confined to the unit interval. Unlike other measures that need to be divided by some maximum value, i.e., the maximum entropy, this measure is always a maximum value at 1 and a minimum value at 0. The measure is typically applied to the Likert scale to determine degrees of agreement among ordinal-ranked categories when dealing with data collection and analysis, though other scales are possible. Using this measure, investigators can easily determine the proximity of ordinal data to consensus (agreement) or dissention. The authors identify a set of criteria that a measure must satisfy in order to be an acceptable indicator of consensus and show how the consensus measure presented here satisfies all the criteria.

Keywords
Consensus, dissention, agreement, Ordinal scale, Likert scale

1 Introduction
Ordinal scales of measurement typically consist of categories of hierarchy such as strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree. The instrument typically used to collect this kind of data is called the Likert scale. Researchers utilize this kind of instrument to collect data that cannot be ascertained using traditional devices of measure since the data being collected are feelings, perceptions, sensations, emotions, impressions, sentiments, opinions, passions, or the like. Unfortunately, the application of standard statistics to these data is illegitimate [1,3]. This paper looks at the different kinds of scales and presents a new measure for analyzing ordinal scale data.

1.1 The Use of Mathematics in Measures
Collections of categories used to accumulate data that are without any sense of order are called *nominal* scales. An example of a nominal scale is a listing of continents, i.e., 1 = Africa, 2 = Antarctica, 3 = Asia, 4 = Australia. 5 = Europe, 6 = North America, and 7 = South America. It is not logical to say that any meaningful ranking order is possible among these categories, and an increasing or decreasing alphabetical order is not a ranking order. Hence, any attempt to make meaningful comparisons is inappropriate.

Ordinal scales are merely ordered categories, but the ordering makes comparisons implicitly possible. Testing the temperature of a cup of tea would permit someone to use the words cold, tepid, lukewarm, warm, moderately hot, hot, and very hot as their scale of comparative measure. There is no sense of interval scale in this measure and hence, equations such as cold + lukewarm = moderately hot, or the average of hot and very hot is hot and a half, is impractical and illogical. Likert scales fall into this category of measures. The number of choices that may be selected as categories of a Likert scale are virtually without limit, although the five or seven category scale are the most prominent. With an increase in the number of categories, it may be argued that the accuracy of the category selected becomes crisper. However, regardless of the granularity of the categories, ordinal scales are merely ordered. The distance between each category, sometimes referred to as the interval, is incorrectly assumed to be equal. There is no empirical evidence to suggest that the interval between neutral and agree is equal to the interval between strongly disagree and disagree. In friendly discussions we frequently can have our view on a particular matter change from agree to disagree and back, with little involvement in the neutral category, and to get someone to embrace an extreme view usually requires far more effort than to merely agree or disagree. This suggests that there is not necessarily symmetry in the ordinal scale. The interval between neutral and strongly agree can be quite different than the interval between neutral and strongly disagree.

Interval scales possess a definite and fixed interval about them such as is observed when viewing that the temperature is 70 degrees Fahrenheit. There does exist a scale created using the freezing and boiling points of water to define 32 degrees and 212 degrees, respectively. However, even though a fixed interval scale exists, there is no benchmark value that initiates an absolute zero value (the Kelvin scale, however, does possess such a value) so equations such as the product of 32 degrees and 2 = 64 degrees make no logical sense. Equations involving temperature must first be converted to Kelvin scale for computation purposes, and then converted back to the original scale.

Ratio scales have an absolute zero base, possess an interval, and implicitly possess order. The number line is such a scale, and all mathematical operations can be conducted on it (obviously, some operations are exceptions, i.e., division by zero). To perform any kind of meaningful assessment requires the utilization of a ratio scale. Hence, it is common to hear someone comment on the questionable suitability of using continuous (ratio) scale mathematics to form the measure of an ordinal or interval scale, as if one was mixing apples with oranges. In fact, there is no other generally acceptable method by which a computation can be made except using mathematics, and math requires the use of a ratio scale. The logical justification in using mathematics to determine a value for an ordinal measure rests in the underlying philosophy.

What is crucial to the calculation of a value by which an ordinal scale measure can be compared to another is the means by which the resulting value is interpreted. Thus, merely claiming that an average value of an ordinal scale based on the assignment of a set of integers to the values, i.e. SA = 1, A = 2, etc, is not sufficient to claim that a weighted average captures the respondents' intent, for there is an implicit interval in that interpretation and ordinal scales have no such interval associated with them. The dispersion of values about a central value, i.e., the weighted mean, permits an assessment of the strength of the collective respondents perceptions without placing a focus on an arbitrary numerical interval assignment. Thus, a collective set of ordinal scale values that yield a narrow dispersion can logically be viewed as possessing a greater agreement than one with a wide dispersion. The logic is identical to that of the standard deviation expect that the standard deviation, weighted or otherwise, is a number that lends little value to the understanding of the collective values. The consensus measure, on the other hand, informs the investigator of the sense of dispersion using the commonly understood concept of percentage.

The mean and standard deviation each requires a fixed interval, a zero value, and a continuous scale, none of which are available in an ordinal scale. One might expect, therefore, that any use of ratio measure mathematics on the analysis of ordinal scales is thus inappropriate. If that argument were true, then no mathematical measure can be used to make use of any ordinal scale, the Likert scale included. The alternative is to use a measure of dispersion that does not require a fixed interval or continuous scale, but does offer a minimum and maximum value. Hence, by using a relative measure, such as a percentage, there need not be any agreed to interval scale; all that needs to be in agreement are the extreme values. In the case presented here using a Likert scale, 0 represents complete dissent or an absence of agreement, and 1 represents complete agreement, or an absence of dissent. The consensus measure satisfies the above requirements and hence, is a suitable measure of agreement using ordinal scales. The original motivation for the measure was to resolve a problem dealing with group decision-making dynamics, and was thus named the consensus measure; it is equally proper to refer to it as a measure of agreement.

2 Rules For a Measure Of Consensus

We establish a set of rules that must be satisfied before any measure can be considered a viable solution to the Ordinal scale agreement/consensus problem.

For a given (even) number of n datum points used to collect information on some question of interest, if an equal number of data, n/2, separate themselves into two disjoint groupings each centered on the strongly disagree and strongly agree categories (the extrema), the data set is considered to have no agreement/consensus.

If all the data points are classified in the same category, regardless of the category, then the consensus of the data is considered to be complete at 100%.

If the mix of data points is such that $n/2 + i$, $1 \le i < n/2$, points are assigned to any one category, the degree of consensus must be greater than 0, for the balance among the data is no longer equal.

As the number of categories to which each data point is classified diminishes, the consensus must increase, eventually approaching 1 on the unit interval. Thus, when all data points are placed in a single category, consensus has been maximized and is considered to be perfect, and that is given a value of 1.

3 Brief Historical Overview

The motivation for this measure came out of a need to identify a means by which ordinal data, in the form of the Likert scale, could be used to provide a measure of consensus such that the progress, or lack thereof, of a group of individuals as they sought to attain consensus on some question, could be assigned a meaningful number. The number could then be used to compare against other numbers taken at previous times during the discussion to determine if the group was headed towards an eventual consensus. Thus, the history of its development centers on group decision-making, as this brief overview will demonstrate.

Szmidt and Kacprzyk [5] provide an interesting look at consensus from the perspective of individual intuitionistic fuzzy preference relations by which a distance from consensus can be determined. Using the unit interval, they define complete disagreement as equal to 0, and complete agreement as 1. This is superior to the method of merely using a mean and some measure of variance to determine proximity to consensus, for those values vary with the number of participants.

Herrera, et. al. [2] examines consensus from the perspective of linguistic labels: certain, extremely likely, most likely, meaning full chance, it may, small chance, very low chance, extremely unlikely, and impossible. After considerable matrix computation, they determine that a consensus measure is determined by degrees: level of preference, level of alternative, and the level of relation. A typical group might be challenged in determining how to utilize this measure.

Tcha, Ahammad and Qiang [7] wrote an interesting paper analyzing PhD student reflections, using the following equation to assign a value from the unit interval as a measure of consensus. Initially, this appears to be a nice application of the Shannon Entropy equation:

$$Consensus = 1 - \frac{\sum p_i \times \ln(p_i)}{3 \times .333 \times \ln(.333)}$$

However, the measure fails upon closer inspection. One of the attributes of the entropy equation is its ability to measure the amount of uncertainty associated with any

Table 1. Examples

	SA	A	N	D	SD	wMean	wStdDev	Cns	Cns%	Dis	Dis%
1	6	0	0	0	6	3.0	8.0	0	0%	1	100%
2	6	0	0	1	5	2.9167	5.6146	0.0492	5%	0.9508	95%
3	6	0	0	2	4	2.8333	5.2083	0.0996	10%	0.9004	90%
4	5	1	0	2	4	2.9167	4.3241	0.1468	15%	0.8532	85%
5	5	1	2	4	0	2.4167	2.3241	0.4437	44%	0.5563	56%
6	1	5	2	4	0	2.7500	1.3611	0.6103	61%	0.3897	39%
7	1	5	4	2	0	2.5833	0.9907	0.6866	69%	0.3134	31%
8	0	6	5	1	0	2.5833	0.6146	0.7676	77%	0.2324	23%
9	0	9	3	0	0	2.2500	0.3750	0.8553	86%	0.1447	14%
10	0	12	0	0	0	2.0	0	1	100%	0	0%

probability distribution. Hence, given a probability distribution on **n** categories, there are **n!** ways in which the categories can be ordered. Each separate ordering of the distribution will have the same entropy value associated with it, for it is a principle of entropy that the ordering of the categories within a distribution not effect the entropy value. Entropy is constant regardless of the order of the categories within the distribution. This is exactly in opposition to the requirements essential to a consensus measure, for as the shape of the distribution changes, so must the measure of consensus reflected by that change.

The Shannon entropy equation is

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

where X is the set of n categories under investigation, and p_i is the probability of each x_i. A comprehensive description of this measure and its properties can be found in [4]. Since the Shannon entropy already possessed some of the properties desired in the consensus measure, it was the focus of our research endeavor, and the results yielded the consensus measure.

4 Properties of the Consensus Measure

The consensus measure is defined [6] as:

$$Cns(X) = 1 + \sum_{i=1}^{n} p_i \log_2\left(1 - \frac{|X_i - \mu_X|}{d_X}\right)$$

where X is represented as the Likert scale, X_i is the particular Likert attribute, p_i is the probability of the frequency associated with each X_i, d_X is the width of X, and $E(X) = \sum_{i=1}^{n} p_i X_i = \mu_X$ is the mean of X. This measure adequately fulfills the above rules as evidenced by the following *illustrations*.

Let us assume that we have a five-attribute Likert scale: Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), and Strongly Disagree (SD). Let us further assign a numerical scale of SA = 1, A = 2, N = 3, D = 4, and SD = 5. Then X = {1, 2, 3, 4, 5}, X_1 = 1, etc., d_x = 5 - 1 = 4. Using an arbitrary number of random integer values to populate the scale, the following table denotes the required properties.

Table 1 contains data on six aspects: the first column is simply an index of the rows, columns SA through SD denote the values assigned to the Likert scale attributes (for comparison purposes all category values sum to 12), wMean is the weighted mean for the category values, wStdDev is the weighted standard deviation for the category values, Cns and Cns% are the consensus values in decimal and rounded percent, and Dis and Dis% are the dissention values in decimal and rounded percent (Cns = 1 - Dis).

Row 1 shows a maximum amount of dissent in agreement since n/2 values are reflected in the extreme attributes. As a point of interest, had the n/2 values been associated with Agree and Disagree, the consensus would have been 0.5850 and the weighted standard deviation 2.0, since these attributes are closer to each other. Rows 2 through 9 show a convergence of dispersion moving towards Agree. An examination of wMean shows a modest fluctuation of the values but, in general, a movement of value from 3 = Neutral to 2 = Agree. This is supported by wStdDev as the values continue to converge towards 0 as the values surrounding the attributes merge. The consensus shows continuous movement towards 1; it is arguably easier to associate the consensus as a percent in order to easily visualize the movement towards an agreement. Conversely, one can monitor the dissent from total (row 1) to total absence (row 10). Finally, row 10 shows the attribute values firmly in one category. The wMean is expectedly at 2, the wStdDev is now zero, consensus is complete at 100%, and dissent does not exist.

4.1 Weighted Mean and Weighted Standard Deviation

The weighted mean of a set of ordinal category values $\{x_1, x_2, \ldots, x_n\}$ with weights $\{w_1, w_2, \ldots, w_n\}$ is given by $\sum_{i=1}^{n} w_i x_i$ where each weight w_i is a nonnegative real number and the sum of all w_i = 1.

Given the weight for the i^{th} category w_i, the number of non-zero categories N', the total number of categories N, and the weighted mean of the categories $\bar{X}_w$, the weighted standard deviation is

$$\sqrt{\frac{\sum_{i=1}^{N} w_i\left(x_i - \bar{X}_w\right)}{\frac{(N'-1)\sum_{i=1}^{N} w_i}{N'}}}$$

Figure 1 shows that the relationship between the consensus measure and the weighted standard deviation is almost perfect.

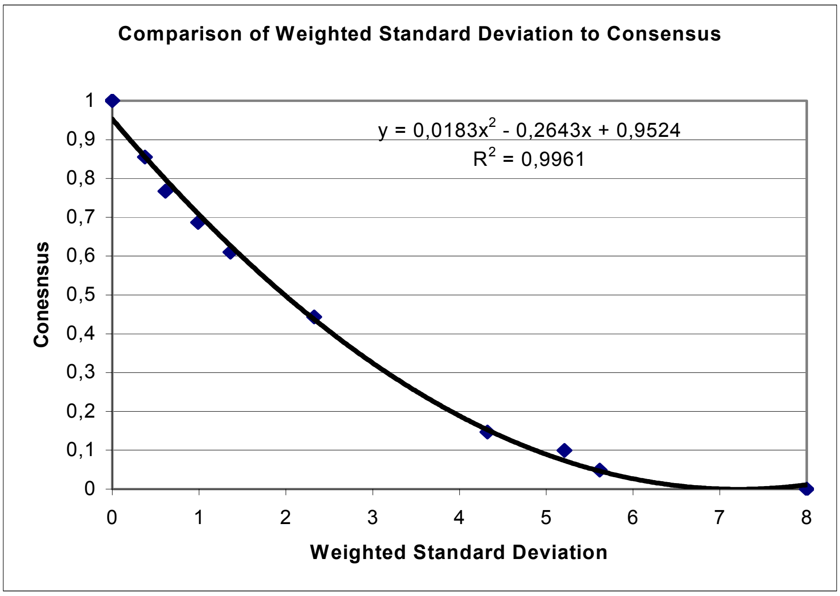


Figure 1. *A comparison of the weighted standard deviation and consensus values from table 1.*

The correlation is 0.996 for this very small set of values, but as the number of rows of data increases, the granularity softens and the correlation approaches 1. Thus, both are measures of dispersion and both capture the variance in the data, but the consensus measure is interpreted in terms of a percentage of the entire range of possible values without regard to any implicit interval that might or might not be present in the ordinal/interval measure. This property makes it easier for the investigator to understand both the dispersion of the data, and the proximity or distance a current investigation might be from its goal. Whatever the maximum weighted standard deviation of the category values, the consensus measure will always represent it as 0 and the minimal weighted standard deviation is always 1.

5 Conclusion

The consensus measure offers statisticians, psychologists, organizational behaviorists, social scientists, and a plethora of other disciplines a tool by which ordinal scales can be compared and contrasted. This new measure is equally applicable to interval and ratio scales; research into its usage is currently underway in various areas of interest. An interesting development is the softening of the requirement that the category values be integers. The measure works equally well under situations of ambiguity when the investigator cannot commit to one single action: perhaps the instance being measured is mostly of one category but part of another. Using real numbers in the category values permits the measure to be applied to another class of problems, with the single provision that the fractional parts sum to 1.

The new measure of consensus serves as a measure of agreement as well as the inverse of dissention. Using a measure devised from the discipline of information theory it is easy to develop an intuition for dispersion of data, be it ordinal, interval or ratio, in a manner that is impossible to ascertain with conventional statistics. Equally possible is the ability to compare measures of a data set over time to determine if the investigation being undertaken at hand, be it group dynamics, some activity in nature, or a laboratory experiment, is approaching consensus or agreement or dissent when the data being measured is qualitative in nature and the scale is ordinal, interval or ratio.

References:

1. Cohen, L.; L. Manion; K. Morrison, (2000). *Research Methods in Education*, 5th edition, Routledge Falmer, London.
2. Herrera, F.; E. Herrera-Viedma; J. L. Verdegay, (1996), “A model of consensus in group decision making under linguistic assessments,” *Fuzzy Sets and Systems*, **78**, 73-87.
3. Jamieson, S, (2004). Likert scales: how to (ab)use them. *Medical Education*, **38**, 1217-1218.
4. Klir, G.J.; M. J. Wierman (1998), *Uncertainty-Based Information*. Physica-Verlag, New York.
5. Szmidt, E.; J. Kacprzyk, (2003), “A Consensus-Reaching Process Under Intuitionistic Fuzzy Preference Relations,” *Int’l J of Intelligent Systems*, **18**, 837-852.
6. Tastle, W. J.; M. J. Wierman, (2005). Consensus and Dissention: A New Measure, *Intl J. of Intelligent Systems*, under review.
7. Tcha, M.; H. Ahammad; Y. Qiang, (ca. 2002) "What Future Australian Professors In Economics And Business Think: Results From Twin Surveys Of Phd Students." www.econs.ecel.uwa.edu.au/ (downloaded June 2004).

Automatic detection of abnormal human behavior using video processing

S.A. Velastin and A.C. Davies
Kingston University, Kingston upon Thames, United Kingdom

Closed-circuit television (CCTV) systems are regarded as a major tool for surveillance within urban environments, capable of monitoring wide areas to identify and track individuals, detect potentially criminal or terrorist behavior, locate intruders in secure areas, and provide various kinds of situation awareness. The socio-economic context is that in a country like the United Kingdom, for example, the annual cost associated with anti-social behavior alone has been estimated at around € 5000 million.

The level of deployment of CCTV in urban environments continues to grow, and there are indicators that they can be perceived by the public at large as having a positive benefit. Such confidence in their operation can only be maintained if the vast amounts of data generated by large numbers of cameras can be effectively and efficiently analyzed. Manual observation of many cameras is highly labor-intensive, and imposes unrealistic demands on the attention capacity of operators and as a consequence the majority of live images are un-observed, functioning mainly to support post-event inspection of recorded video (another labor-intensive task).

What is therefore desirable to support efforts to maintain an acceptable balance between the rights to personal security and privacy, is the availability of computer-based systems that can be capable of identifying potentially dangerous or suspicious behavior so as to timely inform those who might be in a position to prevent or control such situations. A particular challenge is to demonstrate robust systems that can operate continuously with little human intervention in fairly complex scenarios (e.g. public transport, retailing facilities, streets, etc.) and that also scale up so as to have the large geographical coverage required in practical applications.

This symposium gives attendees the opportunity to understand what is currently possible with computer vision technologies, what the main challenges are and discuss both the technical and social implications of the deployment of these systems.

Understanding and improving perceived personal security in public transport

M. A. Vicencio-Silva

This presentation sets the scene for this Symposium, from the perspective of those who use a public space, using public transport as its primary example. Perceived personal security of travellers and staff is important to public transport operators in terms of their responsibility to their customers and employees and its effects on patronage. More generally, improvements in social inclusion can be achieved through better access to public transport use.

The first steps to achieve an improved perception of personal security risks are,

- Gain an understanding of the difference and relationship between reported personal security incidents and people's perception of risks to their security.
- Understand people's concerns when using (or choosing not to use) public transport and thus inform improvements in this area.
- Gather a 'wish-list' from people regarding their personal security in public transport to help defining longer-term aims in the area.

The information gathered through these steps was collected from a diversity of stakeholders: staff, travelling and non-travelling public. Special attention was given to socially excluded sections of the population. Methods included extensive one-to-one and small group interviews. The insights obtained into how security is perceived have identified a range or practical ways of reducing risks to personal security in public transport systems.

The many efforts already being made by public transport operators were reviewed to consider how they can be complemented with wider initiatives in the community. The type of information and communications technologies that can contribute to these efforts is outlined.

The concept of public transport stations, vehicles, structures and stops as "places" which all kinds of people should feel that they have equal rights to share, emerges strongly – as does the need to deny these rights to the minority who exclude themselves by criminal or other deliberately antisocial behaviour. Possible means for detecting such behaviour effectively and thus to empower those who need to be in control of these public spaces, is then addressed by the rest of the speakers.

Automatic visual recognition for metro surveillance

F. Cupillard, M. Thonnat and F. Bremond

"The ADVISOR behavior recognition system"

We present in this work an approach for recognizing the behavior of either isolated individuals, groups of people or crowds in the context of visual surveillance of metro scenes using multiple cameras. In this context, a behavior recognition module relies on a vision module composed of three tasks: (a) motion detection and frame to frame tracking, (b) multiple cameras combination and (c) long term tracking of individuals, groups of people and crowd evolving in the scene. For each tracked actor, the behavior recognition module performs three levels of reasoning: states, events and scenarios. We have also defined a general framework to easily combine and tune various recognition methods (e.g. automaton, Bayesian network or AND/OR tree) dedicated to the analysis of specific situations (e.g. mono/multi actors activities, numerical/symbolic actions or temporal scenarios). Validation results on different methods used to recognize specific behaviors are described.

Making the home safer and more secure through visual surveillance

R. Cucchiara, A. Prati and R. Vezzani

Video surveillance has a direct application in intelligent home automation or domotics (from the Latin word domus, that means "home", and informatics). In particular, in-house video surveillance can provide good support for people with some difficulties (e.g. elderly or disabled people) living alone and with limited autonomy. New hardware technologies for surveillance are now affordable and provide high reliability. Problems related to reliable software solutions are not completely solved, especially concerning the application of general-purpose computer vision techniques in indoor environments. Indeed, assuming the objective is to detect the presence of people, track them, and recognize dangerous behaviors by means of abrupt changes in their posture, robust techniques need to cope with non-trivial difficulties.

In particular, luminance changes, shadows and frequent posture changes must be taken into account. Long-lasting occlusions are common due to the proximity of the cameras and the presence of furniture and doors that can often hide parts of a person's body. For these reasons, we developed computer vision techniques based on probabilistic and appearance-based tracking, particularly conceivable for people tracking and posture classification.

Despite its effectiveness for long-lasting and large occlusions, this approach tends to fail whenever the person is monitored with multiple cameras and he appears in one of them already occluded. Different views provided by multiple cameras can be exploited to solve occlusions by warping known object appearance into the occluded view. To this aim, this paper describes an approach to posture classification based on projection histograms, reinforced by HMM for assuring temporal coherence of the posture. The single camera posture classification is then exploited in the multi-camera system to solve the cases in which the occlusions make the classification impossible.

These above-mentioned problems are analyzed and solutions based on background suppression, appearance-based probabilistic tracking, and probabilistic reasoning for posture recognition are described. Experimental results with different amounts and types of occlusions are reported and demonstrate that the proposed approach is capable of correctly classifying the posture even in the case of large occlusions.

Design of vision technology for automatic monitoring of unexpected events

X. Desurmont, J. Meessen, A. Bastide, Ch. Parisot and J-F Delaigle

Due to the emergence of digital standards and systems it is now possible to deploy easily and rapidly CCTV in site for permanent or temporary uses. Examples of challenging surveillance applications are monitoring metro stations, detection of loitering or abandoned objects, etc. This talk describes a practical implementation of a distributed surveillance system with emphasis on video transmission issues (acquisition, visualization) and image processing necessary for useful event detection. The requirements for these systems are to be easy to use, robust and flexible. Our goals are to obtain efficiently implemented systems that can meet these strong industrial requirements. A computer cluster based approach with network connections is the innovative solution proposed. The main advantage of this approach is its flexibility. Since mobile objects are important in video surveillance, these systems will include image analysis tools such as segmentation and object tracking. First we present the typical requirements of such a system. We consider issues like the facility to deploy and administer network-connected real-time multi-cameras, with reusable modular and generic technologies. Then we analyze how to cope with the needs to integrate a solution with state-of-the-art technologies. As an answer we then propose a global system architecture and we describe its main features to explain each underlying module. To illustrate the applicability of the proposed system architecture in real case studies, we develop some scenarios of deployment for indoors or outdoors applications.

A multi-sensor surveillance system for the detection of anomalous behaviors

G.L. Foresti, C. Micheloni, C. Piciarelli and L. Snidaro

In this talk we present a hierarchical framework that uses networks of static and active cameras. Static cameras are organized into subnets for the local monitoring of sub-areas of the whole environment. Each subnet is organized in such a way that data acquired are first locally fused. As data are collected, information about position, velocity and shapes of the moving objects are extracted and used to identify and classify occurred events. A communication protocol allows the sharing of information between different networks and especially between the hierarchy levels. This establishes a cooperation that, driven by an operative control unit, allows the identification of complex events occurred inside the environment and not only in restricted local areas. Finally, as result of the event analysis, active cameras can be tasked to track and acquire suspicious targets at higher resolution for surveillance purposes as identification, recognition and behavior analysis.

A multi-sensor surveillance system for public transport environments

J-L Bruyelle, L. Khoudour and S.A. Velastin

To improve safety and security, public transport camera networks get bigger and bigger. However, operators are only able to monitor a few cameras at any given time. It is therefore highly desirable to automate the monitoring tasks through computer vision. Although considerable work has been carried out in the field of image processing to detect events and measure flows, effective implementation of these tasks in a real-world network has not been fully realized. We describe an approach, consisting of small, localized computers dedicated to local image processing of one or two cameras only. These processors are interconnected through a computer network. This talk describes the implementation of this system, as well as four computer vision applications that have been implemented on it. Two of them are linked to security purposes and deal with intrusions and abnormal stationarities. The two others are related to passenger flow measurements and perform counting of passengers and queue length measurement. They have all been intensively tested in real life conditions within several types of network such as an airport and underground railways. Results from trial results show the efficiency of the tested solutions.

Intelligent CCTV Surveillance: advances and limitations

S.A. Velastin, A.C. Davies
Digital Imaging Research Centre, School of Computing and Information Systems
Kingston University, Kingston upon Thames, United Kingdom

Abstract
This paper outlines the motivation and requirements for CCTV surveillance particularly in public spaces and on how practical computer vision systems are beginning to be proposed to support the human monitoring task. The explosion of CCTV puts increasing demands on the sophistication required of such systems. We highlight here what has been achieved and the current limitations of such systems.

Keywords
Closed-circuit television, surveillance, image processing, security, people tracking, crowd-monitoring.

1 Introduction
The installation of closed circuit television (CCTV) cameras in urban environments is now commonplace and well-known. Public attitudes to these systems reflect the balance needed between two conflicting requirements:

a) concerns over invasion of privacy and fears of authoritarian control of the population
b) welcoming the increased safety in public spaces and reductions in crime and antisocial behavior.

Fears in category (a) are not surprising, since within living memory, there have been countries governed by regimes with a tendency to the oppressive monitoring and control of their citizens, and the continuous tracking of actual or imagined 'dissidents'.

Support for category (b) arises from public concerns over both real and imagined risks of urban crime and terrorism. Controlling anti-social behavior and protection against terrorist threats is generally perceived to have a high priority, making intrusive monitoring relatively acceptable and encouraging the installation of advanced surveillance systems. Recent uses of the phrase 'homeland security', and plans for technologically-advanced personal ID cards have made this a topical issue. There is an assumption, founded on practice, that CCTV systems make a contribution to public security and safety, but the monitoring task is resource intensive and will benefit from some degree of automation.

2 The evolution of CCTV surveillance
CCTV based surveillance has developed from simple systems comprising a camera connected directly to a viewing screen with an observer in a control room, watching for incidents of crime or vandalism or searching for targeted individuals, to complex multi-camera systems spread over large geographical areas [15]. Advanced digital recording and playback techniques can be provided, with searching capabilities, and suitable for audit and to present observed results as evidence in legal proceedings. It is common to find manual camera control (pan, tilt and zoom) as an aid to track events or objects of particular types. As the number of cameras in each system increased, they first exceeded the number of monitoring screens and then exceeded the capability of the observing teams to watch events effectively, as their attention span is inevitably limited [17]. For example, even if we assume that at best a person could monitor, say, 10 video sources simultaneously, major cities like London and Paris have or plan to have more than 10,000 cameras in their underground train systems, thus requiring at least 1,000 people per working shift. A British newspaper reported in 2004 that over four million CCTV cameras were deployed in the UK [6]. There are significant pressures on managers to provide cover for such CCTV monitoring facilities at the possible expense of having staff on the ground that can interact directly with people and improve the feeling of personal security.

Although the word surveillance sometimes carries negative undertones, we use it in the context of what is called a "capable guardian" as an approach to dissuade people from criminal or antisocial activities and to re-assure citizens that they are protected and that public spaces are properly managed. In this context, what is most important is to pick up early indications of *potentially dangerous* situations that might affect safety or security so as to alert a human being who can then take control, assess the situation and deploy resources as necessary.

Human action is rarely decided on the basis of visual information alone, but to be able to cope with the vast and largely uneventful amounts of visual data it is highly desirable to automate the monitoring task through computer vision systems. Detection thresholds usually need to be biased in favor of false positives, since these can be quickly recognized and disregarded by a human observer, whereas missing real incidents is a serious deficiency. This is best done on-line to prevent and manage but also off-line to find evidence (searching large amounts of recorded video is never a trivial task). These are what we call Intelligent CCTV Surveillance systems.

In this paper we outline some representative work in the field of computer-assisted visual surveillance of human activity especially in public places which are subject to significant volumes and complexity of activity. Space limitations prevent us from mentioning all key work or from going into much detail (please see [4], on which this paper is partly based). This is better addressed by the accompanying papers that form part of a special symposium on "Automatic detection of abnormal human behavior using video processing" in this conference.

It is key to learn from what humans do in monitoring environments such as Control Rooms [12] where attention is focused on special (unusual) situations which are known (to an experienced operator) to be associated with the need to intervene. These situations are likely to be the *result* of unusual *behavior*. This subtle difference between behavior and situation is important to understand how reliable computer systems could be designed to guide human monitoring but not necessarily to replace it. A typical scenario is that of pick-pocketing where in most cases the behavior itself might be very difficult to see, but its situational effects (e.g. blockage in an escalator) is known to the trained eye to be correlated to a behavioral pattern.

Fairly successful means of detecting this type of situation even in cluttered conditions operate at the pixel (or groups of pixels) level without attempting to *understand* behavior itself and we show some examples of this in section 3. In section 4 we outline work aimed at tracking people as a means of identifying anomalous situations.

3 Crowd Monitoring

3.1 Pixel-level processing

Early uses of visual surveillance included the estimation of global properties of urban crowds such as density and flow [3]. Applications have been found in densely-populated urban spaces such as city rail-stations, shopping malls and airport departure lounges. Data can also be gathered to assist architects in the planning of urban environments. The techniques do not attempt to identify individual pedestrians in a crowd – rather, the crowd is monitored as a generalized entity, and 'average' properties sought. Some analogies with fluid behavior and with the behavior of charged particles in an electric field may be observed.

For example, if the area of an image occupied by pedestrians can be identified, then the ratio of 'crowd area' to 'background area' provides a rough estimate of crowd density (***Figure 1***). Compensation is needed for the different apparent size of objects at various distances. Recent work has shown has this can be measured automatically [11]. Variation of lighting levels and directions presents many problems. In traditional surveillance with sparse spatial and temporal activity (e.g. sterile zones), this can be overcome using methods based on statistical models of such variation, provided that most of the time what is seen by the camera is background. Foremost in this category are the Gaussian-mixture models proposed in [13].

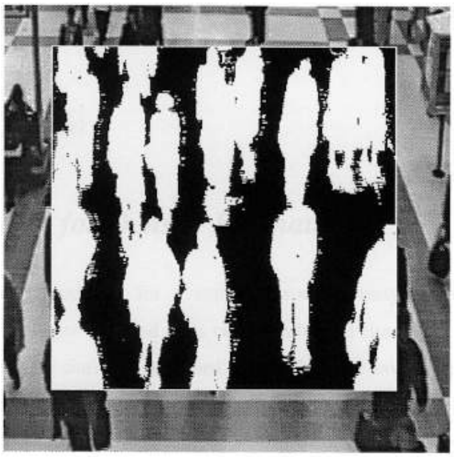


Figure 1. *Pedestrian extraction from crowd-scene [18].*

Difficulties still exist for crowded conditions, especially when people and objects might persist in the field of view for very long periods of time. Good results have been shown by methods that combine luminance statistics with image motion estimation (similar to the encoding of video material), an example of which is given in ***Figure 2*** [16]. In addition to improving background estimation, motion (or its absence) is an important clue for the detection of unusual situations.

Figure 2. *Extracted motion-vector field – arrows show direction of motion.*

3.2 Detecting unusual situations

Once a robust way has been established to measure motion properties, to separate foreground from background, and relating them to the geometry of the scene, then the detection of spatial and/or temporal variation of these image properties can be correlated to situations that might merit the attention of an operator. For example, by monitoring density, it is possible to set thresholds above which safety might be endangered due to overcrowding or congestion (***Figure 3***). This is important because entry to city-centre rail stations may have to be periodically closed during rush-hours in order to keep crowd density on platforms at a safe level.

Figure 3. *Automatic identification of congestion (dense non-moving area), denoted by array of white dots.*

Detection of individuals who linger in one place for long periods while surrounding crowds are moving, or of individuals moving in a different direction from the majority, may be a potential indication of planned or actual criminal behavior (***Figure 4***). This detection might be appropriate to spot people considering committing suicide on the railways, known to linger at the end of the platforms for a long time before jumping in front of a train (the cost of which, financially and emotionally, is very high). A similar technical challenge exists for the detection of abandoned packages which are responsible for daily public transport disruption (***Figure 5***).

***Figure 4**. Maintaining detection of lingering people during occlusion (person on left hand side by the wall poster).*

***Figure 5**. Detection of stationary object (a suitcase to the right of the central pillar).*

As far as public safety is concerned, a typical situation is that of people who willingly or not walk too close to the edge of a platform (***Figure 6***). This is an instance of a general class of detection of *intrusion* (trespassing into a forbidden area). Note that the computer vision system does not need to *know* that what is being detected is people, but that just in a given context the presence of an area of image foreground of a given minimum size is an indication of a situation of interest. We hope that through this example the reader gets an appreciation of the power but also of the limitation of current computer vision systems (in this example, it is not possible to infer the detailed behavior of the intruder, let alone their intention).

***Figure 6**. Detection of a person too close to the edge.*

4 Individual Tracking

Tracking is generally understood as the process of first localizing each person in an image and then following the progress of each person from one image to another and (ideally) from one camera to another. Note that we use the term *localization* to make a distinction with *identification* that implies obtaining the actual identity of a person through some biometric process (such as face recognition or gait). This process of tracking is currently only robust enough for relatively sparse human traffic (and mainly restricted to people walking). It is an important field of work because many researchers base leading edge attempts to interpret complex or subtle behavior on assumptions of accurate tracking.

Having obtained a classification of image pixels into background, foreground and moving, the foreground parts are segmented (grouped) into distinct objects, normally represented by an enclosing rectangle or "blob". The tracking process then consists on using kinematic constraints and sometimes *appearance* measurements (size, color) to match corresponding blobs from one image to the next. This is referred to as the *correspondence problem*. We then end up with blobs that have been labeled with identifiers that uniquely associate an object (or person) in a given video sequence. A typical example [7] is shown in ***Figure 7***.

***Figure 7**. Pedestrians marked by blobs (approaches marked by □ and departures by ├).*

The tracking data is then used to correlate unusual positional and velocity temporal patterns ("dynamics") with unusual behavior [10]. Typical examples include detection of lingering, movement from one car to another in a car park, taking an uncommon path. Difficulties naturally arise from occlusion. If the object disappears behind an obstacle, it is possible to use its velocity to estimate the place and time of its emergence assuming no change in its velocity (more sophisticated systems might use acceleration data too). Commonly, a Kalman filter is used to provide better estimates of future trajectories of objects which move into an occluded zone [8, 14]. However, it is possible to imagine many situations where this is unreliable – for example two people being tracked may disappear behind an obstacle, and while there may meet, hold a conversation, then split up and depart in opposite directions. To automatically determine which one is which from the image sequences, following their re-emergence, is still not at all easy [9].

Until recently, detection of these situations was effectively wired-in (pre-programmed). In environments that can be constrained, this is a reasonable approach that has shown to be appropriate in other fields (such as industrial inspection) and that could lead to successful implementations e.g. for home or office surveillance. However, for large systems deployed in open public places it is recognized that it is crucial to have systems that adapt and learn. For example, based on tracked positions a system can self-measure typical entry and exit points and common paths between them so that an observation that is outside what has been thus learnt, can be flagged (***Figure 8***). A different approach to behavior-related detection is described in [5], using goal-directed models to have a system separate between interesting and non-interesting behavior.

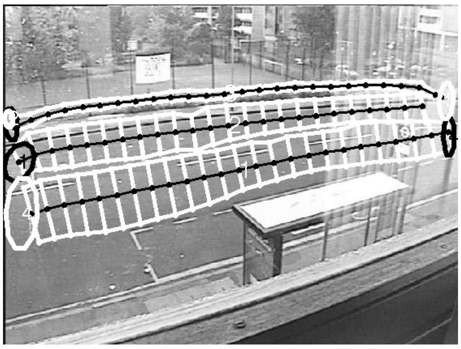


Figure 8. Example of learnt entries, exits and gates [1].

5 Conclusions

It is hoped that we have given here an indication of the current state of Intelligent CCTV Surveillance. The progress in computing and telematics (wireless LANs, mobile phones) all contribute to the continued deployment of more complex and advanced CCTV surveillance systems, which is likely to become increasingly unobtrusive as cameras decrease in size, and likely to be integrated with other sensors (audio, thermal, etc.).

The key current limitation of these surveillance systems is that they have to rely on human observation. Automatic recognition of behavior patterns is beginning to be demonstrated outside research laboratories and it will improve, so that it will become easier to detect and predict both legitimate and illegal activity. However, there is still a long way to go before these systems have capabilities approaching those of humans. Automatic analysis of gestures and posture [2] is key to identifying rich behavior in person-to-person interaction and requires significant advances for applicability in public places.

As a final related observation, clearly there is no assurance that these systems will always be used responsibly and only by those with the public interest and safety in mind. Misuse by official agencies and adoption by criminal elements in society may happen if there are insufficient safeguards. Just as the invention of the word processor has not yet resulted in the paperless office, the development of improved CCTV surveillance systems is not likely to lead to the crime-free city centre.

References:

1. Black J., Ellis T., Makris D. (2005). A Distributed Database for Effective Management and Evaluation of CCTV Systems. *Intelligent Distributed Surveillance Systems* (S.A. Velastin, P. Remagnino Eds.), Institution of Electrical Engineers, London (in press).
2. Boulay B., Bremond F., Thonnat M. (2005). Posture Recognition with a 3D Human Model. *Int. Symp on Imaging for Crime Detection and Prevention*, IEE, London, June 2005.
3. Davies A.C, Yin J.H., Velastin S.A. (1995). Crowd monitoring using image processing. *IEE Electronics and Communication Eng. J.*, 7, Feb 1995, pp 37-47
4. Davies A.C., Velastin S.A. (2005). A Progress Review of Intelligent CCTV Surveillance Systems. IEEE 3rd Int. Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2005), Sept 5-7, Sofia, Bulgaria.
5. Dee H, Hogg D. (2004). Detecting Inexplicable Behaviour. *British Machine Vision Conference*, Kingston, UK, Sept. 2004
6. Frith M. (2004). Big Brother Britain, 2004. *The Independent*, 12 Jan 2004
7. Fuentes L.M., Velastin S.A. (2003). From tracking to advanced surveillance. *Proc. Int. Conf. on Image Processing (ICIP 2003)*, Barcelona, Spain, Sept 2003, III-121-4
8. Haritaoglu I., Harwood D., Davis L.S. (2000). Real-time Surveillance of people and their Activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, August 2000, 22, 8, pp809-830
9. Makris D., T.J. Ellis T.J. (2005). Learning Semantic Scene Models from Observing Activity in Visual Surveillance. *IEEE Transactions on Systems Man and Cybernetics - Part B*, 35 (3), pp. 397-408
10. Piciarelli C., Foresti G.L. (2005). Event Recognition by Dynamic Trajectory Analysis and Prediction. *Int. Symp. on Imaging for Crime Detection and Prevention*, IEE, London, June 2005.
11. Renno J.P., Greenhill D., Velastin S.A., Jones G.A. (2005). Learning the Depth Structure of a Monitored Scene. *6th International Workshop on Image Analysis for Multimedia Interactive Services*. April 13-15, Montreux, Switzerland
12. Sanchez-Svensson M., Heath C, Luff P. (2005). Monitoring practice: Event detection and system design. *Intelligent Distributed Surveillance Systems* (S.A. Velastin, P. Remagnino Eds.), Institution of Electrical Engineers, London (in press).
13. Stauffer C., Grimson W.E.L. (2000). Learning Patterns of Activity using Real-Time Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, August 2000, pp 747-757
14. Tao H., Sawhney H.S., Kumar R. (2000). Dynamic Layer Representation and its Applications to Tracking. *IEEE Conference on Computer Vision and Pattern Recognition*, 2, June 2000, pp 134-141
15. Valera M., Velastin S.A. (2005). Intelligent distributed surveillance systems: A Review. *IEE Proceedings - Vision, Image and Signal Processing*. 152 (2) IEE, April 2005, pp. 192-204.
16. Velastin S.A., Boghossian B., Lo B., Sun J., Vicencio-Silva M.A. (2005). PRISMATICA: Toward Ambient Intelligence In Public Transport Environments. *IEEE Transactions on Systems Man and Cybernetics*, 35, Jan 2005, pp 164-182
17. Wallace E., et al (1977). Good Practice for the Management and Operation of Town Centre CCTV. *European Conf. on Security and Detection*, 28-30 Apr. 1997, pp 36-41 (IEE Conf. Pub. 437)
18. Yin J.H (2005). Automation of crowd data-acquisition and monitoring in confined areas using image processing. Ph.D. Thesis, King's College London, University of London, September 1996.

Understanding and improving perceived personal security in public transport

M. A. Vicencio-Silva
Centre for Transport Studies, University College London, London, United Kingdom

Abstract
Perception of personal security risks was researched from the perspective of those that use a public space, with emphasis on the socially or physically excluded, taking the important modern activity of the use of public transport as its primary example. Valuable insights were gained into how security is perceived through extensive one-to-one interviews that identified a range of practical ways of improving personal security in public transport systems. The concept of public transport stations, vehicles, structures and stops as "places" which all kinds of people should feel that they have equal rights to share emerges strongly - as does the need to deny these rights to the minority who exclude themselves by criminal or other deliberately antisocial behavior. Possible means for detecting such behavior effectively and thus to empower those who need to be in control of these public spaces are outlined as are the type of technologies that can contribute to these efforts.

Keywords
Personal security, public transport, image processing, antisocial behavior.

1 Introduction
One important aim of environmental policies is to obtain significant shifts in traveling patterns from private to public modes. At the same time, governmental social policies aim to improve the quality of life and health of all citizens and make societies more socially inclusive. According to some studies, the attractiveness of Public Transport Systems (PTS) increases with a reduction of fear for personal security. For example, it has been estimated that a 10% increase in patronage may be obtained just by improving the perception of personal security [2] of key stakeholders. The term stakeholders is used here to emphasize that these are not only staff and existing passengers but also potential passengers that currently might not be using the system partly because of social or physical exclusion (e.g. women at night or disabled people, respectively). The work described here was mostly performed as part of projects funded by the Engineering and Physical Sciences Research Council (UK) and the European Union.

1.1 Background
Traditionally, the views of stakeholders are gathered through what are generally called "customer satisfaction surveys", targeted at a given statistical sample of the population. Even simple forms of such surveys point to apparent differences (sometimes contradictory) between people's perception of risk and the statistics derived from registered personal security incidents [7].

"Registered" or reported personal security incidents are those that have been documented, whether with the police or with the Public Transport Operator (PTO). They may include crime (i.e. incidents for which the offender can been convicted) and unpleasant incidents reported by the public or the staff that may not reach the judicial system. Other findings have shown that modifications of the public transport space, directed at improving the perceptions of personal security, also have resulted in a reduction in recorded incidents [16]. For further details see [18, 9, 4].

Conventional customer satisfaction surveys tend to leave out not only people who could well use the PTS but choose not to, but also people who have been excluded from the system, for social or physical reasons. These people can benefit from modifications improving personal security perceptions in general, but they have additional needs that must be addressed at the same time, if the objectives of increasing public transport patronage and social inclusion are to be met. In this context, these people are also key stakeholders.

1.2 Stakeholders
Key stakeholders can be grouped roughly in terms of their different concerns with respect to PTS. For example, passengers, potential passengers, ground staff and drivers have their own fears of personal security while on the system. Security and law enforcement personnel deal mainly with preventing, recording and processing incidents.

When considering passengers' and potential passengers' concerns, every part of a journey should be taken into account, from walking from home to bus stop, to stations, etc. and back. For example, lack of information about the time of arrival of the next vehicle will increase feelings of vulnerability in a passenger and the trip may not be made or repeated in the future. Personal security fears can stop people from using public transport altogether [2, 17] and this restricts mobility for people with no (or limited) access to a private car. This has also implications for authorities in charge of other urban spaces.

2 Methodology
Although PTOs own quantitative data were considered, together with publications from European countries, to determine the state of the art and examine trends, it was decided not to conduct conventional surveys. Instead, qualitative empirical methods were used to obtain current information about the perception of personal security risks of key stakeholders in relation to public transport. These methods were based on in depth one-to-one and small group interviews. They provided a mechanism where people could speak freely about their experiences and concerns. Specifically, the qualitative approach provided a direct, deep and personal way to understand security concerns and wishes as perceived by interviewees.

Particular care was taken to understand that information gathering performed in this way could be affected by the subjective views of everyone involved (interviewers and interviewees), even about issues apparently outside the main subject matter. For example, awareness of the dynamics of race, social class, and the intersection of the agendas of interviewers and interviewees was necessary. Although it is hardly possible to eliminate subjectivities, every effort has been made to detect and diffuse situations that can give rise to misunderstandings, such as unfamiliar cultural conventions and language barriers. An important way in which interviewees were encouraged to interact frankly and freely was to assure them that their views would not be attributed when reporting the findings.

Details of the resulting framework and its variants for the interviews and categories of interviewees (community groups, drivers and ground staff, security personnel, law enforcement) can be found in [18,19].

3 Collected information

One of the most important findings from our studies is the identification of several personal and environmental factors that influence the perception of risks to personal security. What a person experiences as their own personal security is the result of the interplay of all these factors.

It was observed that efforts to improve perceived personal security usually seek to reduce the impact of personal factors by altering environmental ones. However, these efforts do not usually address sectors of the population with more specific needs.

The personal factors we found that have the greatest impact on the perception of risk to the personal security of people are their previous experience of aggression, "second-hand" knowledge of incidents (including associated characteristics such as neighborhood and time of the day), their perception of their own vulnerability in relation to others [9, 14, 15] and the perception of the PTO's effectiveness in protecting their space.

The most influential environmental factors can be grouped into spatial and information issues and they change very little from one person to another, or from one mode of transport to another. However their impact can differ greatly when considering different sectors of the population.

Space issues are concerned with the physical characteristics of the locations where people move and the factors that make this space feel insecure. Therefore, streets, bus stops, stations and their surroundings, as well as the interior of the vehicles, are taken into account. Space issues involve the design and condition of the buildings and property. Clean, well-lit spaces, with as much line of sight as possible, are high in the priority list of requirements of people. Such spaces not only reduce the opportunity for crime (e.g. no hiding places), but also allow people to be safely aware of one another. In well maintained spaces, the public feel safe and perceive that the space is being taken care of and someone is in control. Similarly, misuse of PTS space by other people is seen as threatening behavior.

From the point of view of personal security, the objective of the PTOs' design of space in buildings and carriages is to make the public transport system welcoming and safe for legitimate users and hostile (or at least, inconvenient) to offenders. The ideal result is to keep these people out of the system. However, this may only shift criminal and anti-social activity from one place to another. Therefore, efforts are been made to prevent criminality in cooperation with the rest of society, when designing and building new facilities.

The combination of space and information issues can also be seen from the point of view of communications. Every time space and information are adequate and accurate, the public is receiving a message of confidence and control from whoever is in charge of that particular space and information.

4 Impact on particular kinds of people

Insight into the perceptions and wishes of people (details in [18]) with respect to PTSs uncovered an underlying but pervasive thread in the idea of "place".

A friendly, welcoming space is a statement to its users that it is "ok" to be there, that it is the "place" of the user to be in that space. This is the overall message that the public transport system and its premises (and indeed any public space) needs to send, not only to its current users but also to its potential users, with special regard to people from ethnic minorities, elderly and disabled people and other excluded groups. In other words, that it is a "public place".

Public space and public institutions play, or are seen to play, a dual role. Firstly, they are an extension of what people might call home, i.e. a familiar community environment to which they belong and which they own. Secondly, they represent an enforcement of a set of agreed rules to which people, as citizens, conform. The degree to which individuals, groups or institutions are alienated from these rules would clearly be reflected in their perception and behavior in these spaces and it is an indication of their detachment from a sense of belonging and ownership.

Conversely, any sense of belonging or ownership can be adversely affected by what can be regarded as an invasion of the same place by people or institutions that do not share the same community values. In other words, not all people can consider a PTO's premises as a place of their own. For many young people, ethnic minorities, women, people with young children, and others, the space is not user-friendly and, therefore, it does not feel like it is their place to be there. It does not match their needs and it does not fit them. It is important that approaches to improving personal security consider this underlying issue.

In terms of specific situations, there are scenarios that everyone (including staff) feels as threatening. These situations include being involved in or witnessing fights and coming into contact with potentially aggressive people like groups of noisy and rowdy youths, noisy men (e.g. football crowds), drunks. This feeling is worse at night.

Furthermore, there are certain situations that specific groups of people feel to be threatening and which should be addressed if social inclusion is sought. A summary is given below.

4.1 Staff

Some jobs involve the need to interact directly and frequently with the public. Such responsibility can make staff feel more vulnerable, for example when they need to enforce rules (e.g. no smoking, tickets) or when they are seen as the visible front-end of the PTO, and thus responsible for any problems in the network (e.g. delays). Aggression can also be directed at staff when they are perceived as representatives of the state, or any other, authority. Bus drivers operating in single-person mode are particularly vulnerable because boundaries in bus services are less clear than in railway systems (with walls, ticket barriers, fences). Also, buses have a more complex pattern of patronage in terms of social differences (e.g. income, gender, age).

4.2 Older and Disabled People

As passengers, these two groups of people feel less able than younger and more able-bodied people to defend themselves and to avoid or cope with threatening situations. Accessibility is a specific concern. Older people express worries about uneven surfaces and risk of falls on escalators. Under crowded conditions there is the fear of being pushed. For both groups there is the obvious problems of negotiating long distances, climbing stairs, access to carriages, etc. and this vulnerability affects their perceptions of risk to their personal security.

4.3 Ethnic groups

In this context, we use the term "ethnic origin" as used by the UK's Office of National Statistics and related to the UK's Race Relations Act of 1976. Legislation and terms (e.g. nationality, race) vary from country to country. The same is true of equal opportunity frameworks and legal constraints or practices on recording the ethnic origin of people.

There is an increasing recognition [13] that personal security issues are broadly determined by (and therefore cannot be separated from) factors of gender, age and race. Here again we detect the underlying nature of the feeling of place. Members of ethnic minorities tend to fear being victims of racially motivated attacks and, at the same time, not being able to trust the "system" for protection. The extent of the problem cannot always be assessed from the statistics because, in some countries, the ethnic origin of victims and attackers goes unrecorded, often for legal reasons. Moreover, differences with respect to ethnic origin are difficult to distinguish from differences with respect to socio-economic circumstances.

Although this is a problem that extends beyond the PTS, PTOs can be perceived as reinforcing stereotypes. Some interviews with staff revealed a worrying tendency to associate crime and anti-social behaviour with ethnic minorities *per se*. In some cases, the image that PTOs project (e.g. for recruitment or general advertising) can be regarded by ethnic minorities as being excluding, notwithstanding their compliance with legal requirements to operate non-discriminatory recruitment policies.

The feeling of isolation and disassociation from the mainstream seems therefore to be particularly applicable to this group.

4.4 Women

As a group, women have a specific set of concerns that has been researched and is well documented [5, 1, 7, 6, 13, 20]. For example, results of several regional studies made in the '80s in the UK showed that personal security fears made women hire taxis twice as often as men, choosing if possible female drivers [1]. This is consistent with reports that show that around 20% of reported incidents of sexual assault have happened while waiting for, or using, public transport [7].

4.5 Young people

Young people and children are major users of public transport systems [1, 3]. At the same time, they are regarded as causing a significant proportion of the offences in the network. Gangs of rowdy, noisy youths are feared by almost everybody else using the PTS. However, young people also feel vulnerable (e.g. fear of bullying). In fact, in PTS environments [3], most of the verbal abuse aimed at young people originates from adults (staff and older passengers). In contrast, when they are the victims, the aggressors tend to be other young people, usually older and known to them.

4.6 Families

In general terms, concerns related to family groups tend to be focused on the possible lack of facilities necessary to make their journeys pleasant or indeed possible. In some major networks, traditionally there has been a lack of places in stations to care for babies, and toilets to which to take young children if necessary. Moving up and down stairs or escalators and getting on vehicles with prams and bags is inconvenient and unsafe. Accessibility problems make people feel vulnerable. Depending on social conventions, staff and other passengers might even behave rudely towards families. Incidents such as these increase the feelings of isolation, reinforce the impression that the PTS is not the place for families to be in and use and increase the perception of personal security risks.

Another concern in this group is the decision by many parents not to let their young children travel to school on their own by public means for fear for their personal security.

Improving personal security through environmental factors is the most widespread instance of implementation, because it is normally under the full control of PTOs [18]. One of the common themes across people from every group was the perception that a safe environment includes some sort of surveillance system, intelligent enough to produce a fast response in case of emergency, so that an act of aggression can be stopped or avoided. The obvious answer of having more visible staff may be too costly to implement, given the size and range of public transport spaces. An on-line automatic system warning security staff of potentially dangerous situations in different places across that space (e.g. help points, waiting areas, vehicles) is the next best alternative, if the technology is supported by an appropriate management system (e.g. to respond to emergencies in a time acceptable to stakeholders). Such technology can be implemented by an image processing system, operating at real-time, that automatically analyses the video signal coming from multiple cameras installed in the PTS space and warns an operator in a Control Room when a potentially dangerous situations arises, so that it can be assessed and appropriate action taken accordingly.

However, there is always the issue of displacing crime. Any measure should take into account the whole community. One of the basic messages to be put across then is that this public space is not private, that it is intended to be shared with other people who have the right to be there as much as anybody else. At the same time, a strong opposite message should reach the potential offender, i.e. that the public transport space is not an offender's place. This message should be clearly directed to the offender's actions rather than the person. It is these actions what is causing the distress, not the mere presence of certain people in a public place.

5 CCTV

CCTV has been mentioned almost in most interviews and articles in a positive way. In general, passengers, potential passengers, staff and security people think CCTV systems are useful in deterring unsocial behavior and may help the identification of offenders in case of trouble. Also, CCTV has been mentioned often by people talking about measures to improve personal security, e.g. in waiting places, accesses to stations. In particular, people in the UK show a positive inclination, perhaps because of high profile criminal cases that have been solved using CCTV footage.

Additional expectations vary. Most people seem to be aware that the images cannot be watched at every minute. However, most people seem to believe tacitly that, in case of trouble, CCTV will enable help to arrive quickly to the particular location. For example, disabled people mentioned that CCTV may help the detection of people in need of assistance and could be used to link a disabled person with potential helpers. In practice, such a response is feasible only with an on-line (real-time) detection system.

Privacy issues need to be balanced with the apparent advantages of CCTV surveillance. The amount of cameras in public spaces is such that everybody can expect to have their image captured and perhaps stored without their knowledge. In general, the interviewees agreed that the use should be regulated and the public informed when recording is taking place. For legal considerations in a European context, see [12] and for a different perspective, see [11]. Technical aspects of CCTV are being continuously being updated and evaluated. Police have been investigating specifications for reliable use and the possible use of digital video as evidence in court [8, 10].

6 Conclusions

Perception of risks to personal security can affect the patronage of public transport systems, because people may prefer not to make a trip at all, depending on personal factors and the way in which environmental factors affect each person. There are some well-known lists of practical measures dealing with specific aspects of space and information issues to improve personal security in public transport spaces. This improvement will help almost everybody. However, for people with disadvantages of any type, additional considerations are necessary if social inclusion is to be achieved. These considerations can be actual physical changes (like ramps for access, better lighting) but also some others that involve the way in which the community, including the PTO and local authorities, sees itself.

The concept of public transport spaces as “places” in which all kinds of people should feel that they have equal rights to share emerges strongly - as does the need to deny these rights to the minority who exclude themselves by criminal or other deliberately antisocial behaviour.

Possible means for detecting such behaviour effectively, and thus to empower those who need to be in control of these public spaces, are then addressed elsewhere in this Symposium.

References

1. Beuret, K. (1991). Women and transport. Maclean, M. and Goves, D. (Eds.) *Women's Issues in Social Policy.* Routledge. London, UK. 1991.
2. Crime Concern and Transport and Travel Research (1997). Perceptions of safety from crime on public transport. Report for the UK Department of Transport. June 1997.
3. Crime Concern (1999). Young people and crime on public transport. *Report for the UK Department of the Environment, Transport and the Regions*. London, UK, May 1999.
4. Sanchez-Svensson, M. *et al* (2001). *D3: Common set of security statistics and indicators*. PRISMATICA project, 5th Framework Programme, CEC DG-TREN, Brussels, 2001.
5. Department of the Environment, Transport and the Regions, “Women and public transport: the checklist”, March 2000.
6. EPAG (2001). Projects’ Users Meeting. King’s College London, UK. 19th July 2001.
7. 11-6-7 Hall, R. (1985). *Ask Any Woman: A London Inquiry Into Rape and Sexual Assault.* Falling Wall Press, Bristol, UK. 1985.
8. Home Office Police Scientific Development Branch. UK. http://www.homeoffice.gov.uk/crimpol/police/scidev/news/items/digital_cctv_2005.htm
9. Hough, M. (1996). Anxiety about crime: findings from the 1994 British Crime Survey. Home Office Research Study 147. London, UK. 1996.
10. House of Lords (1998). Science and Technology – Fifth report. UK. http://www.parliament.the-stationery-office.co.uk/pa/ld199798/ldselect/ldsctech/064v/st0501.htm
11. Lyon, D. (2001). Surveillance Society: Monitoring everyday life. *Issues in Society*, Open University Press, UK. February 2001.
12. Niclause, C. *et al* (2002). Annex 5: Legal Aspects. *D5-6 Identification and development of innovative processes*. PRISMATICA Project, 5th Framework Programme, CEC DG-TREN, Brussels, 2002.
13. Pickup, L. (1988). Hard to get around: a study of women’s travel mobility. Little, J., Peake, L. and Richardson, P. (Eds.) *Women in Cities*, MacMillan Education, Hampshire, UK. 1988.
14. Rutherford, J. (Ed.) (1999). *Identity: Community, Culture, Difference*. Lawrence and Wishart, London, UK. 1990.
15. Rutherford, J. (1990). A place called home: Identity and the cultural politics of difference. *Identity: Community, Culture, Difference*. Lawrence and Wishart, London, UK. 1990.
16. Steer Davies Gleave (1992). Passenger security on the London Underground: ‘Post’ stage survey. London, UK. 1992.
17. Stafford, J. (1996). Travelling in fear: personal security worries deter greater public transport usage. *Local Transport Today*, London, UK. October 1996.
18. Vicencio-Silva, M.-A. *et al* (2001). Empirical studies of the perception of key stakeholders. *D4: Report on requirements for project tools and processes*, PRISMATICA Project, 5th Framework Programme, CEC DG-TREN; pp 57-83, 2001, Brussels.
19. Vicencio-Silva, M.-A. (2001). Annex 3: Interview Outline. *D4: Report on requirements for project tools and processes*, PRISMATICA Project, 5th Framework Programme, CEC DG-TREN, 2001, Brussels.
20. Women and Transport Forum(1988). Women in the move: How public is public transport? Kramarae (Ed.) *Technology and Women’s Voices*. Routledge. 1988.

Automatic visual recognition for metro surveillance

F. Cupillard, M. Thonnat, F. Brémond
Orion Research Group, INRIA, Sophia Antipolis, France

Abstract

We propose in this paper an approach for recognizing either isolated individual, group of people or crowd behaviors in the context of visual surveillance of metro scenes using multiple cameras. In this context, a behavior recognition module relies on a vision module composed of three tasks: (a) motion detection and frame to frame tracking, (b) multiple cameras combination and (c) long term tracking of individuals, groups of people and crowd evolving in the scene. For each tracked actor, the behavior recognition module performs three levels of reasoning: states, events and scenarios. We have also defined a general framework to easily combine and tune various recognition methods (e.g. automaton, Bayesian network or AND/OR tree) dedicated to the analysis of specific situations (e.g. mono/multi actors activities, numerical/symbolic actions or temporal scenarios). Validation results on different methods used to recognize specific behaviors are described.

Keywords

Visual surveillance, behavior recognition, real-time.

1 Introduction

In this article, we present a method for recognising specific people behaviors such as fighting or vandalism occurring in a cluttered scene (typically a metro scene) viewed by several cameras. The development of visual surveillance systems, as proposed by Hongeng [1], Pentland [2] and Xiang [3], presents several difficulties and one of the most challenging is behavior analysis, since it requires the inference of a semantic description of the features (moving regions, trajectories,...) extracted from the video stream. Our ambitious goal is to recognize in real time behaviors involving either isolated individuals, groups of people or crowd from real world video streams coming from metro stations. This work is performed in the framework of the European project ADVISOR (http://www-sop.inria.fr/orion/ADVISOR).To reach this goal, we developed a system which takes as input video streams coming from cameras and generates annotation about the activities recognized in the video streams. The paper is organised as follows: in the first part, we present briefly the global system and its vision module. Then, we detail the behavior recognition process illustrated through three behavior recognition examples: "Fighting", "Blocking" and "Fraud" behaviors.

2 Overall System Overview

The video interpretation system is based on the co-operation of a vision and a behavior recognition module as shown on Figure 1.

The vision module is composed of three tasks. First a motion detector and a frame to frame tracker generates a graph of *mobile objects* for each calibrated camera. Second, a combination mechanism is performed to combine the graphs computed for each camera into a global one. Third, this global graph is used for long term tracking of individuals, groups of people and crowd evolving in the scene (typically on hundreds of frames). For each tracked actor, the behavior recognition module performs three levels of reasoning: states, events and scenarios. On top of that, we use 3D scene models, one for each camera, as a priori contextual knowledge of the observed scene. We define in a scene model the 3D positions and dimensions of the static scene objects (e.g. a bench, a ticket vending machine) and the zones of interest (e.g. an entrance zone). Semantic attributes (e.g. fragile) can be associated to the objects or zones of interest to be used in the behavior recognition process.

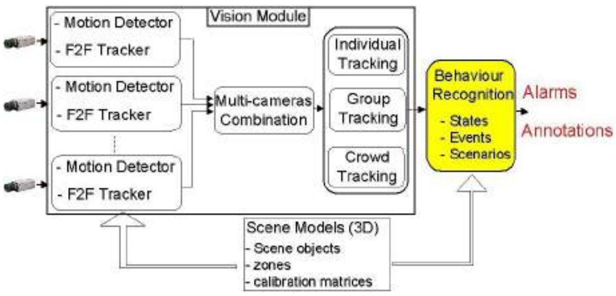


Figure 1. *Video interpretation system.*

2.1 Motion Detector and Frame to Frame Tracking

The goal of the Motion Detector is to detect for each frame the moving regions in the scene and classify them into a list of *mobile objects* with labels corresponding to their type based on their 3D size, such as *PERSON*. This task can be divided into three sub-tasks: detection of *mobile objects*, extraction of features, classification of *mobile objects*. A list of *mobile objects* is obtained at each frame. Each *mobile object* is described by 3D numerical parameters (center of gravity, position ,height, width,...) and by a semantic class (*PERSON, OCCLUDED PERSON, GROUP, CROWD, METRO TRAIN, SCENE OBJECT, NOISE or UNKNOWN*).

The goal of the frame to frame tracker (F2F Tracker) is to link from frame to frame the list of *mobile objects* computed by the motion detector. The output of the frame to frame tracker is a graph of *mobile objects*. This graph provides all the possible trajectories that a *mobile object* may have. The link between a new *mobile object* and an old one is computed depending on three criteria: the similitude between their semantic classes, their 2D (in the image) and their 3D (in the real world) distance.

2.2 Multiple cameras Combination

In order to take advantage of all the calibrated cameras viewing the same scene (cameras with overlapping field of views), we combine all the graphs of *mobile objects* computed by the F2F Tracker for each camera into a global one that we called the **Combined** Graph (see [7] for more details). As a result, the features (the 3D positions and the dimensions) of the *mobile objects* computed in the Combined Graph give a better estimation of the positions and the dimensions of the real persons evolving in the scene.

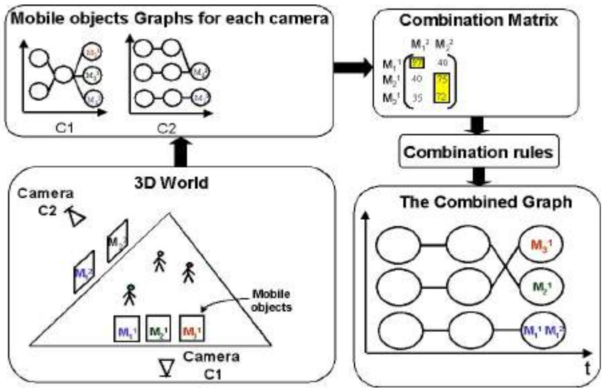


Figure 2. This figure illustrates the multiple cameras combination process. Three persons are evolving in the scene. Camera C1 detects three mobile objects whereas camera C2 detects only two mobile objects. The combination matrix enables to determine (a) a high correspondence between the mobile object M_1^1 of C1 and the mobile object M_1^2 of C2; these two mobile objects are fused together in the combined graph, and (b) an ambiguous correspondence between the two mobile objects M_2^1 and M_3^1 of C1 and the mobile object M_2^2 of C2; the two mobile objects M_2^1 and M_3^1 detected by C1 are selected in the combined graph.

To compute the global graph, we combine at each frame the new *mobile objects* detected for 2 cameras using a combination matrix and a set of rules (see illustration on Figure 2).
The combination matrix gives the correspondences between the *mobile objects* detected for two cameras by using a 3D position and a 3D size criteria. In the case of none ambiguities between the *mobile objects* detected by the two cameras, we fuse the *mobile objects* by making an average on their 3D features. In case of ambiguities, a set of rules is used to either select or eliminate the *mobile object* detected by one of the two cameras.

2.3 Individual, Group of people and Crowd Long Term Tracking

The goal here is to follow on a long period of time either Individuals, Groups of people or Crowd to allow the scenarios involving these three different types of actors to be recognized. For example, when we want to detect a group of people (at least two persons) which is blocking an exit zone, we prefer reasoning with the Group Tracker because it provides a more accurate 3D location of the group of people in the scene.
The Individual Tracker tracks each person individually whereas the Group Tracker tracks globally all the persons belonging to the same group. Both trackers perform a temporal analysis of the Combined Graph. The Individual Tracker computes and selects the trajectories of *mobile objects* which can correspond to a real person thanks to an explicit model of person trajectory. In a similar way, the Group Tracker computes and selects the trajectories of *mobile objects* which can correspond to the persons inside a real group thanks to an explicit model of the trajectories of people inside a group.
Individual and Group Trackers are running in parallel. When the density (computed over a temporal window) of detected *mobile objects* becomes too high (typically if the *mobile objects* overlap more than 2/3 of the image), we stop these two trackers because in such a situation, they cannot give reliable results. At this point, we trigger the Crowd Tracker which is in fact the Group Tracker with an extended model of the trajectories of people inside a group allowing a large density of detected people belonging to the same group that by this way defines a crowd.

3 Behavior Recognition Process

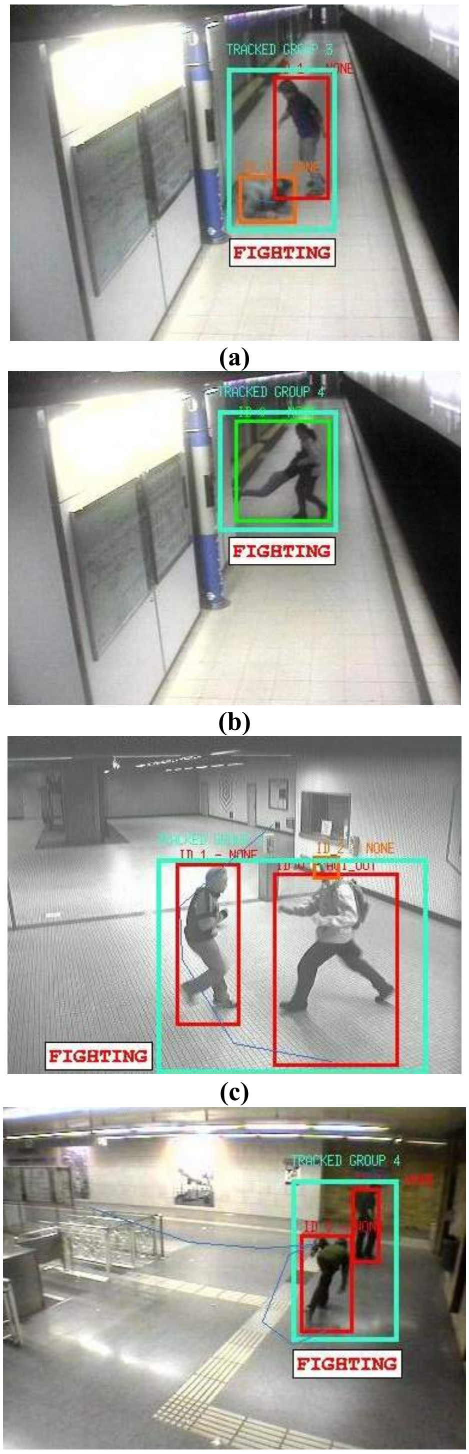
The goal of this task is to recognize specific behaviors occurring in a metro scene. A main problem in behavior recognition is the ability to define and reuse methods to recognize specific behaviors, knowing that the perception of behaviors is strongly dependent on the site, the camera view point and the individuals involved in the behaviors. Our approach consists in defining a formalism allowing us to write and easily reuse all methods needed for the recognition of behaviors. This formalism is based on three main ideas. First the formalism should be flexible enough to allow various types of operators to be defined (e.g. a temporal filter or an automaton). Second, all the needed knowledge for an operator should be explained within the operator so that it can be easily reused. Finally, the description of the operators should be declarative in order to build an extensible library of operators.

3.1 Behavior representation

We call an actor of a behavior any scene object involved in the behavior, including static objects (equipment, zones of interest…), individuals, groups of people or crowd. The entities needed to recognize behaviors correspond to different types of concepts which are:

1. **The basic properties**: a characteristic of an actor such as its trajectory or its speed.
2. **The states**: a state describes a situation characterising one or several actors defined at time *t* (e.g. a group is agitated) or a stable situation defined over a time interval. For the state: "an individual stays close to the ticket vending machine", two actors are involved: an individual and a piece of equipment.
3. **The events**: an event is a change of states at two consecutive times (e.g. a group enters a zone of interest).
4. **The scenarios**: a scenario is a combination of states, events or sub scenarios. Behaviors are specific scenarios (dependent on the application) defined by the users. For example, to monitor metro stations, end-users have defined 5 targeted behaviors: "Fraud", "Fighting" "Blocking", "Vandalism" and "Overcrowding".

To compute all the needed entities for the recognition of behaviors, we use a generic framework based on the definition of **Operators** which are composed of four attributes:
Operator name: indicates the entity to be computed such as the state "an Individual is walking" or "the trajectory is straight".
Operator input: gives a description of input data. There are two types of input data: basic properties characterising an actor and sub entities computed by other Operators.
Operator body: contains a set of competitive methods to compute the entity. All these methods are able to compute this entity but they are specialised depending on different configurations. For example, to compute the scenario "fighting", there are 4 methods (as shown on Figure 3). For example, one method computes the evolution of the lateral distance between people inside a group. A second one detects if someone, surrounded by people, has fallen on the floor.
Operator output: contains the result of the entity computation accessible by all the other Operators. This result corresponds to the value of the entity at the current time.



(a) (b) (c) (d)

Figure 3. This figure illustrates four methods combined by an AND/OR tree to recognize the behavior "Fighting". Each image illustrates a configuration where one method is more appropriate to recognize the behavior: (a) lying person on the floor surrounded by people, (b) significant variation of the group width, (c) quick separation of people inside a group and (d) significant variation of the group trajectory.

This generic framework based on the definition of Operators gives two advantages: It first enables us to test a set of methods to compute an entity, independently of other entities. So we can locally modify the system (the methods to compute an entity) while keeping it globally consistent (without modifying the meaning of the entity). Second, the network of Operators to recognize one scenario is organised as a hierarchy. The bottom of the hierarchy is composed of states and the top corresponds to the scenario to be recognized. Several intermediate levels, composed of state(s) or event(s) can be defined.

3.2 Behavior recognition

We have defined four types of methods depending on the type of entities:

Basic properties methods: we use dedicated routines to compute properties characterising actors such as trajectory, speed and direction. For example, we use a polygonal approximation to compute the trajectory of an individual or a group of people.

State methods: we use numerical methods which include the computation of: (a) 3D distance for states dealing with spatial relations (e.g. "an individual is close to the ticket vending machine"), (b) the evolution of temporal features for states dealing with temporal relations (e.g. "the size of a group of people is constant") and (c) the speed for states dealing with spatio-temporal relations (e.g. "an individual is walking") and (d) the combination of sub states computed by other operators.

The output of these numerical methods is then classified to obtain a symbolic value.

Event methods: we compare the status of states at two consecutive instants. The output of an event method is boolean: the event is either detected or not detected. For example, the event "a group of people enters a zone of interest" is detected when the state "a group of people is inside a zone of interest" changes from false to true.

Scenario methods: for simple scenarios (composed of only 1 state), we verify that a state has been detected during a predefined time period using a temporal filter. For sequential scenarios (composed of a sequence of states), we use finite state automatons. An automaton state corresponds to a state and a transition to an event. An automaton state also corresponds to an intermediate stage before the complete recognition of the scenario. We have used an automaton to recognize the scenarios "Blocking" and "Fraud" as described on Figure 4 and 5.

For composed scenarios defining a single unit of movement composed of sub scenarios, we use Bayesian networks as proposed by Hongeng [4] or AND/OR trees of sub scenarios as illustrated on Figure 6. A description of Bayesian networks for scenario recognition can be found in [6]. We have defined one Bayesian network to recognize the "violence" behavior composed of 2 sub scenarios: "internal violence" (e.g. erratic motion of people inside a group) and "external violence" (e.g. quick evolution of the trajectory of the group).

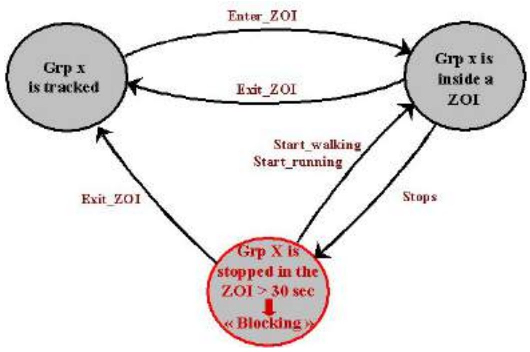


Figure 4. To check whether a group of people is blocking a zone of interest (ZOI), we have defined an automaton with three states: (a) a group is tracked, (b) the group is inside the ZOI and (c) the group has stopped inside the ZOI for at least 30 seconds.

Both of these methods need a learning stage to learn the parameters of the network using ground truth (videos annotated by operators). Bayesian networks are optimal given ground truth but the AND/OR trees are easier to tune and to adapt to new scenes.

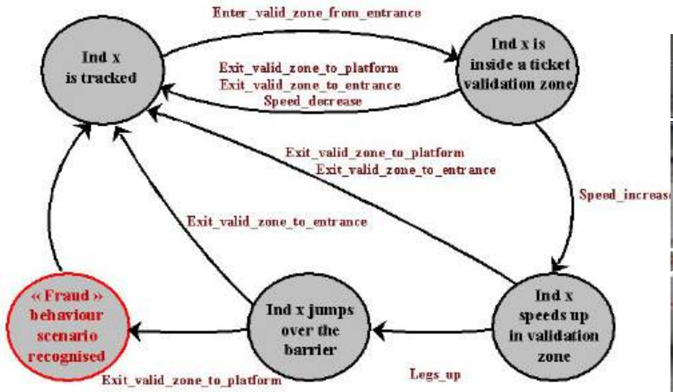


Figure 5. *To check whether an individual is jumping over the barrier without validating his ticket, we have defined an automaton with five states: (a) an individual is tracked, (b) the individual is at the beginning of the validation zone, (c) the individual has a high speed, (d) the individual is over the barrier with legs up and (e) the individual is at the end of the validation zone.*

For scenarios with multiple actors involved in complex temporal relationships, we use a network of temporal variables representing sub scenarios and we backtrack temporal constraints among the already recognized sub scenarios as proposed by Van Vu [5].

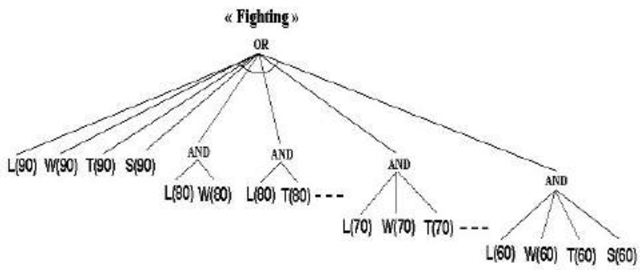
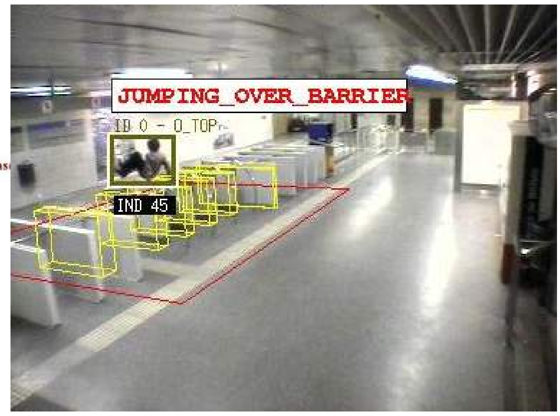
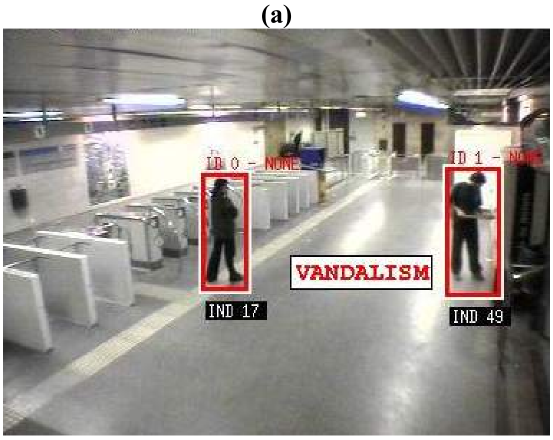


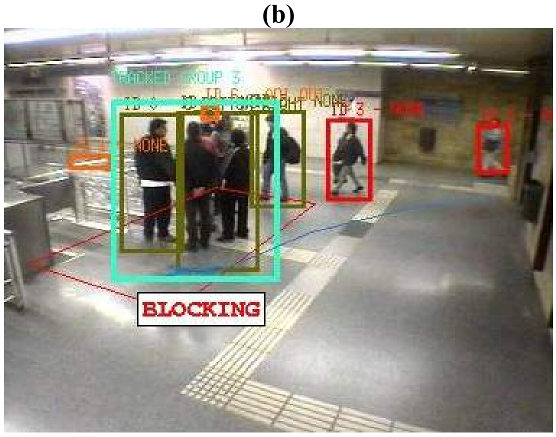
Figure 6. *To recognize whether a group of people is fighting, we have defined an AND/OR tree composed of four basic scenarios: (L) lying person on the floor surrounded by people, (W) significant variation of the group width, (S) quick separation of people inside the group and (T) significant variation of the group trajectory. Given these four basic scenarios we were able to build an OR node with all combinations (corresponding to 15 sub scenarios) of the basic scenarios. These combinations correspond to AND nodes with one up to four basic scenarios. The more basic scenarios there are in AND nodes, the less strict is the recognition threshold of each basic scenario. For example, when there is only one basic scenario (e.g. L(90)), the threshold is 90 and when there are four basic scenarios, the threshold is 60. To parameterise these thresholds, we have performed a learning stage consisting in a statistical analysis of the recognition of each basic scenario.*



(a)



(b)



(c)

(d)

Figure 7. *This figure illustrates four behaviors selected by end users and recognized by the video interpretation system: (a) "Fraud" recognized by an automaton, (b) "Vandalism" recognized by a temporal constraint network, (c) "Blocking" recognized by an automaton and (d) "Overcrowding" recognized by an AND/OR tree.*

3.3 Behavior recognition results

The behavior recognition module is running on a PC Linux and is processing four tracking outputs corresponding to four cameras with a rate of 5 images per second. We have tested the whole video interpretation system (including motion detection, tracking and behavior recognition) on videos coming from ten cameras of Barcelona and Brussels metros. We correctly recognized the scenario "Fraud" 6/6 (6 times out of 6) (Figure 7.a), the scenario "Vandalism" 4/4 (Figure 7.b), the scenario "Fighting" 20/24 (Figure 3), the scenario "blocking" 13/13 (Figure 7.c) and the scenario "overcrowding" 2/2 (Figure 7.d). We also tested the system over long sequences (10 hours) to check the robustness over false alarms. For each behavior, the rate of false alarm is: 2 for "Fraud", 0 for "vandalism", 4 for "fighting", 1 for "blocking" and 0 for "overcrowding".

Moreover, in the framework of the European project ADVISOR, the video interpretation system has been ported on Windows and installed at Barcelona metro in March 2003 to be evaluated and validated. This evaluation has been done by Barcelona and Brussels videosurveillance metro operators during one week at the Sagrada Familia metro station. Together with this evaluation, a demonstration has been performed to various guests, including the European Commission, project Reviewers and representative of Brussels and Barcelona Metro to validate the system. The evaluation and the demonstration were conducted using both live and recorded videos: four channel in parallel composed of three recorded sequences and one live input stream from the main hall of the station. The recorded sequences enabled to test the system with rare scenarios of interest (e.g. *fighting*, *jumping over the barrier*, *vandalism)* whereas the live camera allowed to evaluate the system against scenarios which often happen (e.g. *overcrowding*) and which occurred during the demonstration and also to evaluate the system against false alarms. In total, out of 21 *fighting* incidents in all the recorded sequences, 20 alarms were correctly generated, giving a very good detection rate of 95%. Out of nine *blocking* incidents, seven alarms were generated, giving a detection rate of 78%. Out of 42 instances of *jumping over the barrier,* including repeated incidents, the behavior was detected 37 times, giving a success rate of 88%. The two sequences of *vandalism* were always detected over the six instances of *vandalism*, giving a perfect detection rate of 100%. Finally, the *overcrowding* incidents were also consistently detected in the live camera, with some 28 separate events being well detected.

In conclusion, the ADVISOR demonstration has been evaluated very positively by end-users and European Committee. The algorithm responded very successfully to the input data, with high detection rates, less than 5% of false alarms and with all the reports being above approximately 70% accurate.

4 Conclusion and Future Work

In this article, we have described a video interpretation system able to automatically recognize high level of human behaviors involving individuals, groups of people and crowd.

Different methods have been defined to compute specific types of behaviors under different configurations. All these methods have been integrated in a coherent framework enabling to modify locally and easily a given method. The system has been fully tested off-line and has been evaluated, demonstrated and successfully validated in live condition during one week at the Barcelona metro in March 2003. The next step consists in designing the video interpretation system to be operational (able to cope with any unpredicted real world event) and working on a large scale. For that, we need to design a platform able to be configured dynamically and automatically.

References:

1. Hongeng, S.; Brémond, F.; Nevatia, R. (2000). Bayesian framework for video surveillance application. *Proc. of the 15th International Conference on Pattern Recognition*, Barcelona, Spain, Sept. 2000.
2. Pentland, A.; Liu, A. (1999). Modeling and prediction of human behavior. In : *Neural Computation*, pp. 229-242.
3. Xiang, T.; Gong, S.; Parkinson, D. (2003). On the structure of dynamic Bayesian networks for complex scene modelling. *Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS).* Nice, France, Oct. 2003.
4. Hongeng, S.; Nevatia, R. (2001). Multi-agent event recognition. *International Conference on Computer Vision (ICCV2001)*, Vancouver, B.C., Canada, 2001/07/12.
5. Van Vu, T.; Brémond, F.; Thonnat, M. (2003). Automatic video interpretation: A recognition algorithm for temporal scenarios based on pre-compiled scenario models. *ICVS2003*.
6. Moenne-Loccoz, N.; Brémond, F.; Thonnat, M. (2003). Recurrent Bayesian network for the recognition of human behaviors video, *ICVS2003*.
7. Cupillard, F.; Brémond, F.; Thonnat, M. (2002). Group behavior recognition with multiple cameras. *IEEE Workshop on Applications of Computer Vision*, Orlando, Dec. 2002.

Making the home safer and more secure through visual surveillance

R. Cucchiara[1], A. Prati[2], R. Vezzani[1]
[1] *Dipartimento di Ingegneria dell'Informazione, University of Modena and Reggio Emilia, Italy*
[2] *Dipartimento di Scienze e Metodi dell'Ingegneria, University of Modena and Reggio Emilia, Italy*

Abstract
Video surveillance has a direct application in intelligent home automation or domotics (from the Latin word domus, that means "home", and informatics). In particular, in-house video surveillance can provide good support for people with some difficulties (e.g. elderly or disabled people) living alone and with limited autonomy. A key aspect in video surveillance systems for domotics is that of analyzing behaviors of the monitored people. To accomplish this task, people must be detected and tracked, and their posture must be analyzed in order to model behaviors recognizing abrupt changes in it.
Problems related to reliable software solutions are not completely solved, in particular luminance changes, shadows and frequent posture changes must be taken into account. Long-lasting occlusions are common due to the proximity of the cameras and the presence of furniture and doors that can often hide parts of a person's body. For these reasons, a probabilistic and appearance-based tracking, particularly conceivable for people tracking and posture classification, has been developed. However, despite its effectiveness for long-lasting and large occlusions, this approach tends to fail whenever the person is monitored with multiple cameras and he appears in one of them already occluded. Different views provided by multiple cameras can be exploited to solve occlusions by warping known object appearance into the occluded view. To this aim, this paper describes an approach to posture classification based on projection histograms, reinforced by HMM for assuring temporal coherence of the posture.

Keywords
People's behavior analysis, computer vision, video surveillance, posture analysis

1 Introduction and Related Work
A very relevant application for behavior analysis is that related to people's health care. This has been a crucial topic in many fields of research from almost the beginning. The increase each year in deaths and injuries for domestic incidents has shown up in-house safety as an emergent field of research. The widespread distribution of cameras in our world for video-surveillance purposes makes available a huge amount of visual information with which the people's safety can be improved.
With these premises, many computer vision techniques have been proposed to (semi-)automatically assure the people's safety. Among them, the detection of people's posture has recently gained credit, because of fundamental importance for people's behavior analysis or for detecting alarming situations (such as a fall). Many methods based on computer vision have been proposed in the literature to classify people's posture. They can be differentiated by the more or less extensive use of a 2D or 3D model of the human body [9]. In accordance with this, most of them can be classified into two basic approaches to the problem. From one side, some systems (such as that proposed in [4]) use a direct approach and base the analysis on a detailed human body model. In many of these cases, an incremental predict-update method is used, retrieving information from every body part. These approaches are often too constrained to the human body model, resulting in unreliable behaviors in the case of occlusions and perspective distortions that are very common in cluttered, relatively small, environments like a room. Moreover, for in-house surveillance systems, low-cost solutions are preferable, thus stereovision or 3D multi-camera systems should be discarded. Consequently, algorithms of people's posture classification should be designed to work with simple visual features from a single view, such as silhouettes, edges and so on [4, 9].
Since the posture is often related to the shape, occlusions are very critical. Even though some probabilistic approaches are able to maintain the shape in the case of occlusions (as in [4]) they are likely to fail if the occlusion occurs at the beginning, i.e. when the track is firstly created. This problem can be solved by the use of multiple cameras. Moreover, the need for multiple points of view and a distributed system is required in order to cover the entire environment (e.g. the house) and continuously track people in it.
Unfortunately, the possibility to have a full coverage of the environment and to solve occlusions does not come for free. The technical problems in multiple camera systems are several and they have been summarized in [5] into six classes: installation, calibration, object matching, switching, data fusion, and occlusion handling. Among these, object matching is the most addressed problem in the literature and provides the basic tools also for the occlusion handling.
Several works have been proposed to maintain the correspondence of the same tracked object during a camera handoff. Most of those require a partially overlapping field of view [1]; other ones use a feature based probabilistic framework to maintain a coherent labeling of the objects [7]. All these works aim at keeping correspondences between the same tracked object (in order to continuously analyze the behavior in a wide area), but none of them are capable of handling the occlusions during the camera handoff phase.
Approaches to multi-camera tracking can be generally classified into three categories: geometry-based, color-based, and hybrid approaches. The first class can be further subdivided into calibrated and uncalibrated approaches. A particularly interesting paper of calibrated approach is reported in [12] in which homography is exploited to solve occlusions. Single camera processing is based on particle filter and on probabilistic tracking based on appearance to detect occlusions. Once an occlusion is detected, homography is used to estimate the track position in the occluded view, by using the track's last

valid positions and the current position of the track in the other view (properly warped in the occluded one by means of the transformation matrix). A very relevant example of the uncalibrated approaches is the work of Khan and Shah [7]. Their approach is based on the computation of the so called "Edges of Field of View", i.e. the lines delimiting the field of view of each camera and, thus, defining the overlapped regions. Through a learning procedure in which a single track moves from one view to another, an automatic procedure computes these edges that are then exploited to keep consistent labels on the objects when they pass from one camera to the adjacent.

Color-based approaches base the matching essentially on the color of the tracks, as in [8] where a color space invariant to illumination changes is proposed and histogram-based information at low (texture) and mid (regions and blobs) level are exploited to solve occlusions and match tracks by means of a modified version of the mean shift algorithm.

Hybrid approaches mix information about the geometry and the calibration with those provided by the visual appearance. These methods use probabilistic information fusion [6] or Bayesian Belief Networks (BBN) [1].

Our approach is similar to that proposed in [12], but, differently from it, appearance models of the tracks are warped from one view to another not using the ground plane, but a vertical plane passing from the person's feet and triggered by an external or internal input. We will also report results of an experimentation that aims at analyzing the limits of the approach depending on the amount and type of the occurred occlusion.

2 Single Camera Behavior Analyzer

In our multi-camera system, moving objects are extracted from each camera by exploiting background suppression with selective and adaptive update in order to react quickly to the changes and to also take "ghosts" (i.e., aura left behind by an object that begins to move) into account [2]. After the object extraction, a sophisticated tracking algorithm is used to cope with occlusions and split/merge of objects. A probabilistic and appearance-based tracking, similar to that proposed in [11], is used to handle objects with non rigid motion, variable shape (like people), and frequent occlusions. This tracking algorithm maintains, in addition to the current blob *B*, the appearance image *AI* (or temporal template) and the probability mask *PM* of the track. *AI* is obtained with a temporal integration of the color images of the blobs, while the probability mask *PM* associates to each point of the map a probability value that indicates its reliability. Comparing the current blob with the appearance image of the tracks it is also possible to detect if the person is subject to an occlusion or not [3].

Finally, tracks that satisfy some geometrical and color constraints are classified as people and submitted to the posture classifier. Four main postures are considered: *Standing, Crawling, Sitting,* and *Lying*. To this aim, similarly to [4], we exploit a classifier based on the projection histograms computed over the blobs of the segmented people. The projection histograms $PH = (\vartheta(x);\pi(y))$ describe the way in which the silhouette's shape is projected on the *x* and *y* axes. Since the projection histograms depend on the blob size, and, consequentially, on the position of the person inside the room, we first scale them according with the distance of the person with respect to the camera. To compute this distance, a feet detection and tracking module together with a homography relation obtained through camera calibration are exploited [3].

Though projection histograms are very simple features, they have proven to be sufficiently detailed to discriminate between the postures we are interested in. However, this classifier is precise enough if the lower level segmentation module produces correct silhouettes. By exploiting knowledge embedded in the tracking phase, many possible classification errors due to the imprecision of the blob extraction can be corrected. In particular, to deal with occlusions and segmentation errors due to noise, the projection histograms are computed over the temporal probabilistic maps obtained by the tracker instead of the blobs extracted frame by frame.

Despite the improvements given by the use of appearance mask instead of blobs, a frame-by-frame classification is not reliable enough. However, the temporal coherence of the posture can be exploited to improve the performance: in fact, the person's posture changes slowly and through a transition phase during which its similarity with the stored templates decreases. To this aim, a Hidden Markov Model formulation has been adopted. Using the notation proposed by Rabiner in his tutorial [10], the followings sets are defined:

- the state set S, composed by N states: $S = \{S_1, \ldots, S_N\} = Main\_Postures$
- the initial state probabilities $\Pi = \{\pi_i\}$, set equal for each state ($\pi_i = 1/N$). The choice of the values assigned to the vector Π affects the classification of the first frames only, and then it is not relevant.
- the matrix **A** of the state transition probabilities, computed as a function of a reactivity parameter α (empirically determined; for example, we set α = 0.95 during our experiments). The probabilities to remain in the same state and to pass to another state are considered equal for each posture. Then, the matrix **A** has the following structure:

$$\mathrm{A} = \mathrm{A}(\alpha) = \{A_{ij}\},\ A_{ij} = \begin{cases} \alpha & i = j \\ \frac{1-\alpha}{N-1} & i \neq j \end{cases} \qquad (1)$$

The Observation Symbols and the Observation Symbol Probability distribution **B** have to be defined. To this aim we can use the set of possible projection histograms as observation symbols, since it is numerable. But that means the computation of a very large matrix, composed by *N* rows and w^h columns (where *w* and *h* are the sizes of the images). Thus, we prefer to directly compute the probability values b_j, that indicate the probability to have a particular observation (histograms) belonging to the state (posture) *j*, through the output of the frame-by-frame classifier:

$$b_j = P_j = P(\widehat{PH} \mid posture = S_j) \qquad (2)$$

The HMM presented does not require any additional training phase because it exploits directly the Probability Maps. Then, at each frame, the probability of being in each state is computed with the traditional forward algorithm. At last, the HMM input has been modified to keep into account the visibility status of the person. In fact, if the person is completely occluded, the reliability of the posture must decrease with the time. In such a situation, it is preferable to set $b_j = 1/N$ as the input of the HMM. Making that, the state probabilities tend to a uniform distribution (that models the increasing uncertainty) with a delay that depends on the previous states: the higher the probability to be in a state S_j, the higher the time required to lose this certainty. To manage simultaneously the two situations and to cope with the intermediate cases, (i.e., partial occlusions), a generalized formulation of the HMM input is defined:

$$b_j = P(\widehat{PH}|S_j) \cdot \frac{1}{1+n_{fo}} + \frac{1}{N} \cdot \frac{n_{fo}}{1+n_{fo}} \qquad (3)$$

where n_{fo} is the number of frames for which the person is occluded. If n_{fo} is zero (i.e., the person is visible), b_j is computed as in Eq. 2, otherwise the higher the value of n_{fo}, the more it tends to a uniform distribution.

In Figure 1, the benefits of the introduction of the HMM framework are evidenced. The results are related to a video in which a person passes behind a stack of boxes always in a standing position. During the occlusion (highlighted by the grey strips) the frame-by-frame classifier fails (it states that the person is laying). Instead, through the HMM framework, the temporal coherence of the posture is preserved, even if the classification reliability decreases during the occlusion.

Posture classification is the first, essential step for behavior analysis. For instance, dangerous situations such as falls, can be identified by detecting abrupt changes from a standing to a laying posture. More advanced behaviors can be detected and analyzed by including also other mid-level features, such as the people's gait, into the modeling of the behaviors.

3 Multi-camera Occlusion Manager

As stated above, the probabilistic tracking is able to handle occlusions and segmentation errors in the single camera module. However, the strong hypothesis to be robust to occlusions is that the track has been seen for some frames without occlusions in order for the appearance model to be correctly initialized. This hypothesis is erroneous in the case the track is occluded since its creation (as in Figure 2.b).

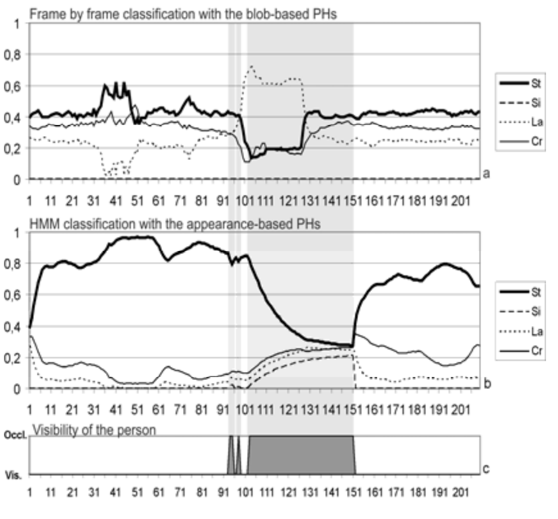
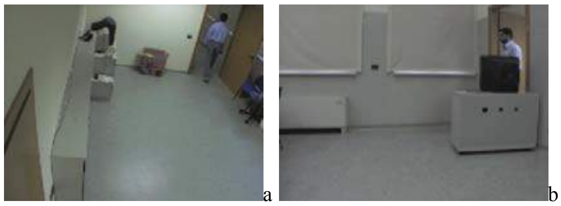


Figure 1. Frame by frame and HMM posture classification during a strong occlusion



a b

Figure 2. An example of track occlusion during its creation: the source (a) and destination view (b)

Our proposal is to exploit the appearance information from another camera (where the track is not occluded) to solve this problem. If a person passes between two monitored rooms, it is possible to keep the temporal information stored into its track extracted from the first room (Figure 2.a) and use them to initialize the corresponding track in the second room (Figure 2.b).

To this aim, with reference to Figure 3, we assume that the two cameras are calibrated with respect to the same coordinate system (X_W,Y_W,Z_W); the equation of the plane $G=f(X_W,Y_W,Z_W)$ containing the entrance is given; it is possible to obtain the exact instant when the person passes into the second room; all the track points lie on a plane P parallel to the entrance plane G and containing the feet.

The first three assumptions imply only an accurate installation and calibration of cameras and sensors, while the last one is a simplification needful to warp the track between the two points of view. Under this condition, in fact, the 3D position of each point belonging to the appearance image of the track can be computed and, then, its projection on a different image plane is obtained.

In particular, the above mentioned process is applied only to the four corners of the tracks, and, exploiting them, the homography matrix H that transforms each point between the two views can be computed. Through H it is possible to re-project both the appearance image AI and the probability mask PM of each track from the point of view of the leaving room to the point of view of the entering one. The re-projected track is used as initialization for the new view that can, in such a manner, solve the occlusion by continuing to detect the correct posture.

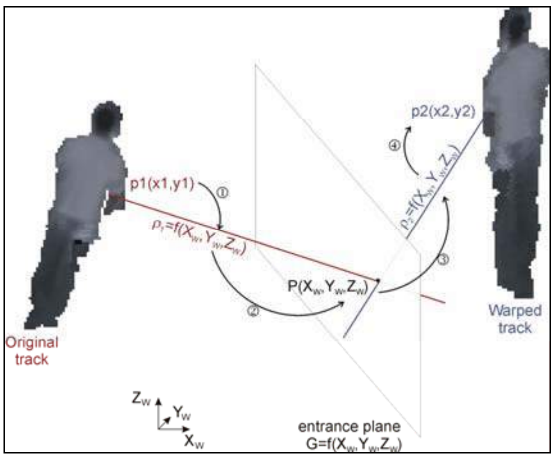


Figure 3. Track warping: 1) exploiting the calibration of camera1, 2) intersection with entrance plane, 3) calibration of camera2, 4) intersection with camera2 plane

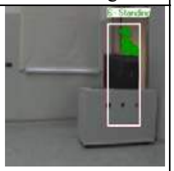
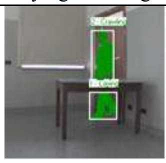
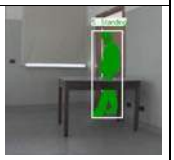
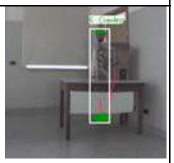
Input	W/o warping	With warping
	 Standing	 Standing
	 Laying/Standing	 Standing
	 Crawl. / Standing	 Standing
	 Missed	 Crawling

Table 1. *Classification results with and without the multi-camera track warping*

4 Experiments and Conclusions

As a test bed for our multi-camera system, a two-rooms setup has been created. The two rooms share a door equipped by an optical sensor used to trigger the passage of the people. We have taken several videos of transition between the first and the second room. Furthermore, in the second one we have placed various objects between the door and the camera to simulate different types and amounts of occlusions. In particular, occlusions starting from both the bottom part and the middle part of the body have been created. In the case of the videos of the first type (bottom occlusions), the single camera posture classifier tends to fail because the body shape is incomplete and the feet are not visible to be tracked. In the second case (middle occlusions) the feet of the person are visible, but two or more tracks are generated and both the tracking and the posture classifier are misled.

Table 1 reports the corresponding posture detection results, showing the classifications given by the system for all the frames subjected to the occlusion and a single snapshot as visual example. Whenever two postures are listed, this means that different postures are associated to either the same track in successive moments or two split tracks.

Nevertheless, when the occluded part is too large, the warped track could be very different with respect to the segmented blob and the tracking algorithm is not capable of taking advantage of the initialization provided by the multi camera module. As a consequence, after some frames the posture classifier fails (see last row of Table 1).

References:

1. Cai, Q.; Aggarwal, J.K. (1999). Tracking Human Motion in Structured Environments Using a Distributed-Camera System. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(**12**),1241-1247.
2. Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A. (2003). Detecting Moving Objects, Ghosts and Shadows in Video Streams". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(**10**):1337-1342.
3. Cucchiara, R.; Grana, C.; Prati, A.; Vezzani, R.. (2005). Probabilistic Posture Classification for Human Behaviour Analysis. *IEEE Trans. on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 35(**1**), 42-54.
4. Haritaoglu, I.; Harwood, D.; Davis, L.S.. (1998). Ghost: A Human Body Part Labeling System Using Silhouettes. P*roc. of Intl Conf. on Pattern Recognition.*
5. Hu, W.; Tan, T.; Wang, L.; Maybank, S. (2004). A Survey on Visual Surveillance of Object Motion and Behaviours. *IEEE Trans. on Systems, Man, and Cybernetics - Part C*, 34(**3**):334–352.
6. Kang, J.; Cohen, I.; Medioni, G. (2003). Continuous Tracking within and Across Camera Streams, *Proc. of IEEE Intl Conf. on Computer Vision and Pattern Recognition*, **1**, 267–272.
7. Khan, S.; Shah, M. (2003). Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(**10**):1355-1360.
8. Li, J.; Chua, C.S.; Ho, Y.K (2002). Colour Based Multiple People Tracking. *Proc. of IEEE Intl Conf. on Control, Automation, Robotics and Vision*, **1**, 309–314.
9. Moeslund, T.B.; Granum, E, (2001). A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, **81**, 231-268.
10. Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(**2**):257–286.
11. Senior, A. (2002). Tracking People with Probabilistic Appearance Models. *Proc. of Intl Workshop on Performance Evaluation of Tracking and Surveillance Systems*.
12. Yue, Z.; Zhou, S.K.; Chellappa, R. (2004). Robust Two-camera Tracking using Homography. *Proc. of IEEE Intl Conf. on Acoustics, Speech, and Signal Processing*, **3**, 1–4.

Design of vision technology for automatic monitoring of unexpected events

X. Desurmont, J. Meessen, A. Bastide, Ch. Parisot, J-F. Delaigle
Video Analysis Group, Image Department, Multitel, Belgium

Abstract

Due to the emergence of digital standards and systems it is now possible to deploy easily and rapidly vision technology on site for permanent or temporary uses for automatic monitoring of unexpected events. Examples of challenging applications [1,2] are surveillance, traffic monitoring, marketing, etc. These techniques could be also useful for instrumental recording of human or animal behavior for research purpose. This talk describes a practical implementation of a distributed video system with emphasis on video hardware issues like acquisition and image processing necessary for useful event detection. The requirements for these systems are to be easy to use, robust and flexible. Our goals are to obtain efficiently implemented systems that can meet these strong industrial requirements. A computer cluster based approach with network connections is the innovative solution proposed. The main advantage of this approach is its flexibility. Since mobile objects are important in video surveillance, these systems will include image analysis tools such as segmentation and object tracking. First we present the typical requirements of such a system. We consider issues like the facility to deploy network-connected real-time multi-cameras, with reusable modular and generic technologies. Then we analyze how to cope with the needs to integrate a solution with state-of-the-art technologies. As an answer we then propose global system architecture and we describe its main features to explain each underlying module. To illustrate the applicability of the proposed system architecture in real case studies, we show some scenarios of deployment for indoors or outdoors applications.

Keywords

Multi-camera, real-time, computer vision, tracking, behavior.

1 Introduction

Video surveillance is a large market as the number of installed cameras can attest. Nevertheless, there is still a need for complete and generic systems that can be inserted in an existing camera network (e.g. CCTV) to increase intelligence and handle automatic processing. These technologies could also be used in traffic monitoring, marketing and general behavior recognition. The requirements for these systems are to be network-connected, multi-cameras, modular, the display must be user-friendly, the vision modules should be plug-and-play and the overall system must be highly reliable and robust.

The work reported here has both research [3,4,5] and industrial motivations. In this article we present a generic, flexible and robust approach for an intelligent real-time video-analysis system.

The paper is organized as follows: section 2 describes the global system and its main characteristics; section 3 is devoted to the image analysis module. Section 4 gives some applications and section 5 concludes and indicates future work.

2 System overview

The video-analysis system presented in this paper [6] is based on digital network architecture. This kind of system can be deployed in a building, for instance or can be connected to an existing data network. Basically, the system is composed of computers connected together through a typical LAN. The various cameras are plugged either on an acquisition board on a PC or directly on the local network hub for IP cameras. A human computer interface and a storage space are also plugged on this system. The main advantage of such architecture is its flexibility. The logical architecture has been designed in a modular way to allow a fair resource allocation over the cluster. Future needs in computing power will be simply addressed by adding a PC in the cluster. Videos are compressed in MPEG4 video stream.

2.1 Hardware

Typical hardware could be smart network cameras (see figure 1 and 2) performing acquisition, processing (detection, video compression) and network communication. Usual network components like switches and hubs, storage units or other processing units and complete PC for user interaction and also illumination sub-systems for night vision are also part of the distributed system.

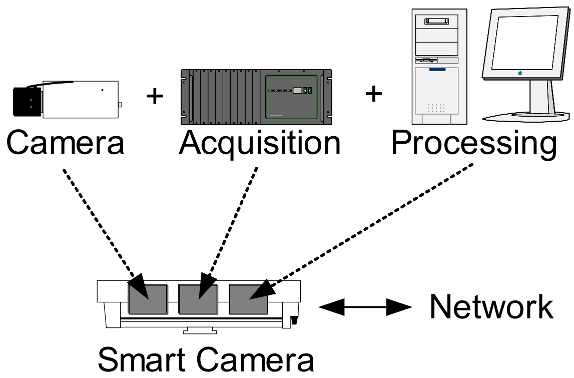


Figure 1. *The smart network camera principle.*

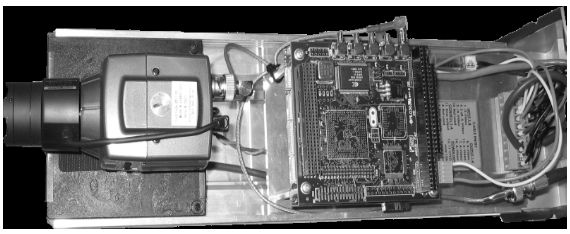


Figure 2*. The ACIC SA smart camera CMVision (Photo by courtesy of ACIC SA).*

3 Image analysis module

High-level interpretation of events within the scene requires low level vision computing of the image and of moving objects. It is usually needed to generate a representation for the appearance objects in the scene. For our system, the architecture of the vision part is divided in three main levels of computation that achieve the interpretation (figure 3):

- Image level (acquisition, image filtering, background evaluation and segmentation),
- Blob level (description, blobs filtering, matching, tracking description and filtering),
- Event level (tracking analysis, finite state machine, performance evaluations).

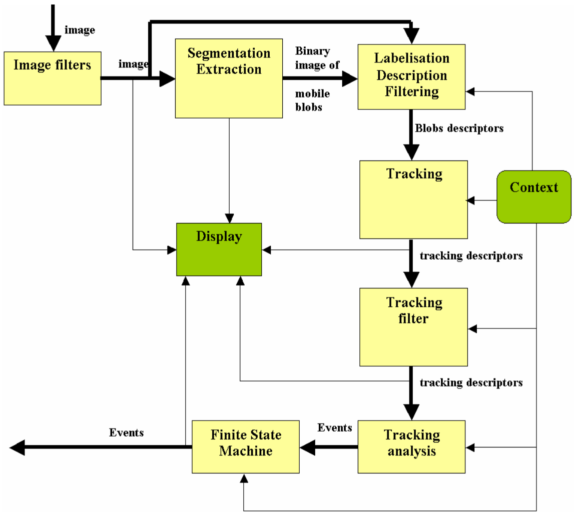


Figure 3. *Design of the vision system components.*

3.1 Segmentation

The common bottom-up approach for segmenting moving objects is the background estimation and foreground extraction (usually called background subtraction) [7]. Many reference image models could be used for representing the backgrounds as they are implemented in the system (low pass temporal recursive filter, median filter, unimodal Gaussian, mixture of Gaussians, vector quantization [8]). We are commonly using the mixture of Gaussians background, as it is quite robust to common noises such as monitor flattering or branches moving in trees.

3.2 Blobs description and filtering

The aim of blobs description and filtering is to make the interface between foreground extraction and tracking and to simplify the information. The description process translates video data into a symbolic representation (i.e. descriptors). The goal is to reduce the amount of information to what is necessary for the tracking module. The description process calculates, from the image and the segmentation results at time t, the k different observed features of a blob i: 2D position in image, 3D position in the scene, bounding box, mean RGB color, 2D visual surface, inertial axis of blob shape, extreme points of the shape, probability to be a phantom blob, etc. At this point, there is another filtering process to remove small blobs, blobs in an area of the image not considered, etc. Other model-based vision descriptors could be also integrated for specific application such as vehicle or human 3D model parameters.

3.3 Tracking algorithm

As the other modules, the tracking part of the system is flexible and fully parametrical on-line. The set-up should be done for a trade-off between computational resources, needs of robustness and segmentation behavior. It is divided in four steps that follow a straightforward approach: estimation, cost matrix computation, matching decisions, tracks updates. Note that there are multiple predictions and cost matrixes when the last matching decision is not unique, and there are only multiple matching decisions for some matching algorithms in MHT (multiple hypothesis tracking [9]). Figure 4 briefly explains the architecture.

The tracking filtering is processed at the tracking description output. It is just as necessary as the other filters of the vision system. As usual the filter is used to remove the noise. At this level of processing, it can use the temporal consistency. We described above some types of filters that can be used in chain. Because the tracking description is a construction built piece by piece during the progression of the video sequence, it can process on-line or off-line. One filter detects and removes tracks that last for less that a fixed duration. This kind of noise comes when the segmentation detects noise in the image as an object. Another filter simplifies tracks by removing samples of blobs that give poor information (e.g. If the blob moves slowly). It could be seen as a dynamic re-sampling algorithm.

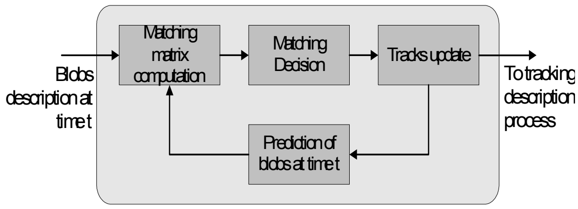


Figure 4. *Basic architecture of Tracking.*

3.4 Tracking description and filtering

The aim of tracking description and filtering is to make the interface between the tracking and the analysis processes and to simplify the information. The tracking description converts the internal tracking result to a graph (figure 5), and then it adds some useful information to the matching data. It computes the time of life of every blob of a track (i.e. the duration of the track from apparition to the specify blob), the time before death (i.e. the duration of the track to disappearance of the specify blob). It also describes a piece of track restricted in a small area as a stopped object. The grammar of the tracking description of blobs behavior includes apparition (new target), split, merge, disappearance, stopped, unattended object, entering a zone, exiting a zone.

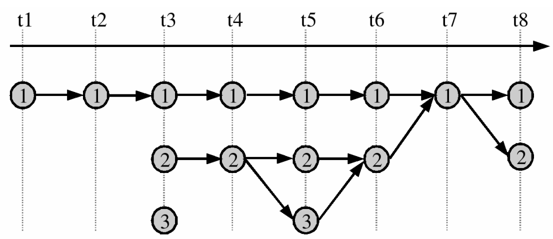


Figure 5. *Internal tracking result description. t1-t8 are the time-stamps. Circles show objects in the image at time t and arrows show matchings between objects in different frames.*

At the tracking description output, the tracking filtering is performed. It is as necessary as the other filters of the vision system. As usual the filter is used to remove the noise. At this level of processing, it can use temporal consistency. We describe below some types of filters that can be used in chain. Because the tracking description is a construction built piece by piece during the progression of the video sequence it can process on-line or off-line.

"*smalltrack*": It detects and removes tracks that last for less that a fixed duration, i.e. the delay between apparition and disappearance of the track. One can see on figure 6 the object labeled 3 at t3 has been erased by this filter.

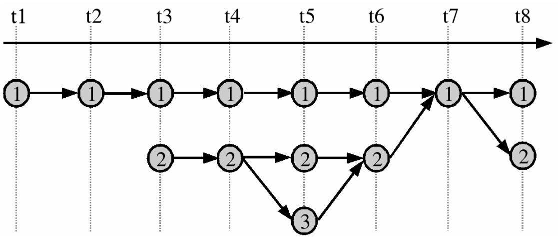


Figure 6. *Output of "smalltrack" filter.*

"*simplifycurvetrack*": Simplifies tracks by removing samples of blobs that give poor information (i.e. if we delete it, we can interpolate it from other parts of the track). Figure 7 shows graphically the difference with and without this filter. Figure 8 shows the output in tracking description.

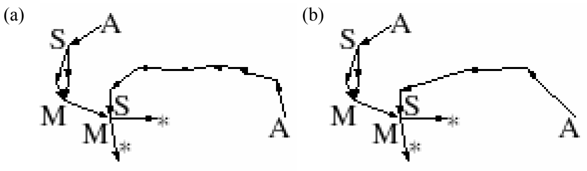


Figure 7. *(a) raw tracking description, (b) tracking description filtered by "simplifycurvetrack". Symbols A, D, S, M and * mean respectively apparition, disappearance, split, merge and "object in current frame".*

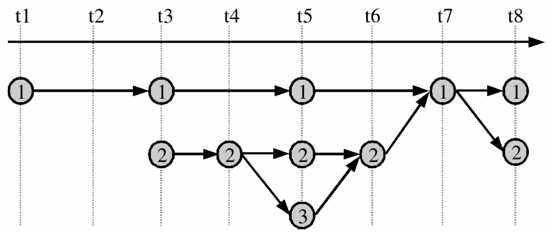


Figure 8. *After filter "simplifycurvetrack" a track has been simplified by removing some objects instances.*

"*simplifysplitmerge*": Removes one part of a double track stemming from a split and then a merge. This kind of noise comes when the segmentation detects two blobs when in all likelihood there is a unique object. Figure 9 and 10 shows the results.

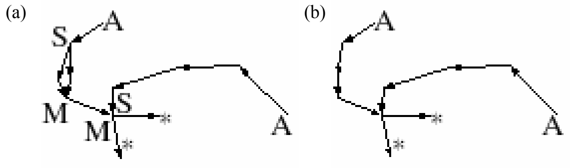


Figure 9. *(a) raw tracking description, (b) tracking description filtered by "simplifysplitmerge"*

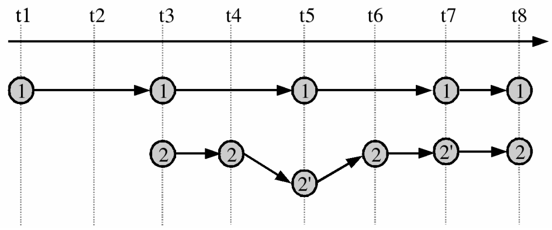


Figure 10. *Output tracking after "simplifysplitmerge".*

We do not describe here other implemented filters in detail. Another filter removes tracks that start or end in defined region of the image.

3.5 Tracking analysis and event generation

The tracking analysis is a process that receives the tracking description. It can find predefined patterns like objects entering in a defined zone of the image and exiting by another one, objects which have exceeded a certain speed limit, or immobile objects for a minimum time which stem from another mobile object. Figure 11 shows this particular pattern of tracking description. In this example we want to know how many people enter or not, with or without looking at the menu near the entrance door. The processing of this module is to look into the tracking description graph (e.g. figure 10) to find the predefined patterns.

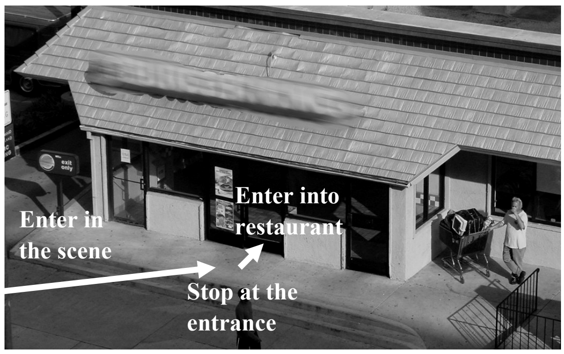


Figure 11. *Tracking description pattern for "read menu before entering".*

4 Applications

In this section, we give some applications of this kind of system. For vision systems, the representation of the scene is quite basic (positions of objects along the time). Thus it is difficult to automatically find people thoughts because, for example, of the difficulty to recover the facial expression. Therefore, to be automatically found, behavior should be easily visually understandable.

4.1 Traffic monitoring

These technologies are helpful to find automatically traffic jams, accidents, fires in tunnel, pedestrians near highways.

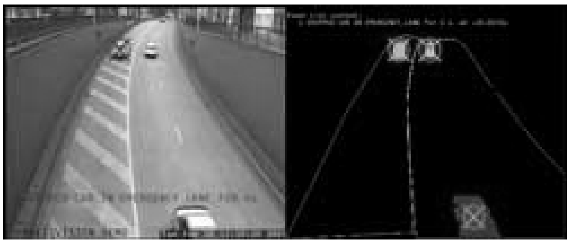


Figure 12. *Example of detection of stopped car.*

4.2 Surveillance:

There are many applications, like car park, public place (e.g. monitoring metro stations, detection of loitering, abandoned objects), private place, jails.

Figure 13. *Example of car park surveillance scene.*

4.3 Makerting

The purpose is to count people that enter and exit from each door of a shop. To know in which part and at what time people behave in the shop, according to some changes in goods layout.

Figure 14. *View of a scene of a shopping center.*

4.4 Behavior

The Relevance is to understand how humans or animals behave with automatic annotation of motion, movement and interactions. The interest of the system presented in this paper is that it is non intrusive, you don't need to put captors (like magnetic, inertial, GPS, etc.) or markers (like colors). After processing, intelligent content access is possible.

5 Conclusion

In this paper, we have proposed an approach for video-analysis platform [6] that can provide the flexibility needed by researchers and that can meet the strong efficiency requirements of industrial applications. We have described the whole image analysis module with segmentation, tracking and event analysis. Originally it was dedicated to surveillance, but it could be used for others applications. These techniques could be useful for instrumental recording of human behavior. The basic limitation is the type of behavior: behavior should be visually understandable.

Figure 15. *View of animals to find motion behavior.*

References

1. Cavallaro, A.; Douxchamps, D.; Ebrahimi, T.; Macq, B. (2001). Segmenting moving objects: the MODEST video object kernel, WIAMIS 2001, *Workshop on Image Analysis for Multimedia Interactive Services*, Tampere, Finland, May 16-17, 2001.
2. Cupillard, F.; Brémond, F.; Thonnat, M. Tracking groups of people for video surveillance, *2nd European Workshop on AVBS Systems.*
3. Shcoepflin, T.; Lau, C.; Garg, R.; Kim, D.; Kim, Y. (2001) A research environment for developing and testing object tracking algorithms, *Proceedings of the SPIE, Electronic Imaging 2001*, **4310**, 667-675.
4. Jaynes, C.; Webb, S.; Steele, R.; Xiong, Q. An open development environment for evaluation of video surveillance systems, *3rd Int. Workshop on PETS*, 1, 32-39.
5. Valera, M; Velastin, S.A. (2003). An approach for designing a real-time intelligent distributed surveillance system, *First Symposium on Intelligent Distributed Surveillance Systems (IDSS)*, IEE, 26 February 2003, London, pp.6/1-6/5.
6. Desurmont, X.; Bastide, A.; Chaudy, C.; Parisot, C.; Delaigle, J.F.; Macq, B. (2005). Image analysis architectures and techniques for intelligent surveillance ystems, *Special Issue on Intelligent Distributed Surveillance Systems on the IEE Proc.-Vis. Image & Signal Process*., **152**(2), April 2005.
7. Elgammal, A.; Duraiswami, R.; Harwood, D.; Davis, L.S. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance, *Proceedings of the IEEE*, **90**(7), July 2002.
8. Kim, K.; Horprasert, T.; Harwood, D.; Davis, L. *Codebook-based background subtraction and performance evaluation methodology.*
9. Cox, I.J.; Hingorani, S.L. *An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking*.

A multi-sensor surveillance system for the detection of anomalous behaviours

G.L. Foresti, C. Micheloni, C. Piciarelli, L. Snidaro
University of Udine,Udine, Italy

Abstract

In this paper, a hierarchical framework employing networks of static and active cameras is presented. Static cameras are organised into subnets for the local monitoring of sub-areas of the whole environment. Each subnet is organised in such a way that data acquired are first locally fused at the connection level (i.e., where different sensors are connected to an elaboration point), then fused at network level. As data are collected, information about position, velocity and shapes of the moving objects are extracted and used to identify and classify occurred events. Robust tracking allows the use of a trajectory analysis module, which clusters together common trajectories and performs a probabilistic analysis of the cluster structure. Defining an anomaly as an event with low probability, we can perform anomalous behaviours identification at the trajectory level using the computed probability of each track.

Keywords

video surveillance, multi-sensor tracking, trajectory analysis, behaviour analysis.

1 Introduction

Nowadays, surveillance of important areas such as shopping malls, stations, airports, seaports, etc. is becoming of great interest for security purposes. In the last years, we have witnessed the installation of a large number of cameras that now are almost covering every place in the major cities. Unfortunately, the majority of such sensors belong to the so-called second generation of CCTV surveillance systems where human operators are required to interpret the events occurring in the scene. Only recently modern and autonomous surveillance systems have been proposed to address the security problems of such areas [1]. Many systems have been proposed [1, 2, 3, 4] for a wide range of security purposes from traffic monitoring [5, 6] to human activity understanding [7]. Video surveillance applications often imply paying attention to a wide area, so different kinds of cameras are generally used, e.g. fixed cameras [6], omni-directional cameras [8] and pan tilt and zoom (PTZ) cameras [9, 10, 11].

The use of these kinds of cameras requires that their number and placement must be fixed in advance to ensure an adequate monitoring coverage of the area of interest. Moreover, developed systems usually present a logical architecture where the information acquired by the sensors is processed by a single operative unit.

In this paper, we propose a hierarchical framework representing a network of cameras organised in subnets in which static and moving (PTZ) sensors are employed. The logical architecture has been designed to monitor motion inside a local area by means of static networks. Furthermore, the PTZ sensors can be tasked, as a consequence of an event detection, to acquire selected targets with higher resolution. Such a property gives an augmented vision to the system that allows better understanding of the behaviours of the objects of interest.

Information about the trajectories of the moving objects, computed by different sensors, is sent, bottom-up, through the hierarchy to be analysed by higher modules. At the first level, trajectories are locally fused to detect suspect behaviours inside local areas. Such trajectories are sent to higher nodes that analyse the activities that have occurred through the different local areas. This allows the detection of events with higher complexity.

2 System Description

The proposed system is composed of a network of cooperating cameras that give a solution to the problem of wide area monitoring (e.g. parking lots, shopping malls, airport lounges, etc.). Specifically, the system could employ a generic number of both static and active sensors in a scalable and hierarchic architecture where two main types of subsystems have been considered: a) Static Camera System (SCS) and b) Active Camera System (ACS). The former is composed by a network of static cameras, while the latter is represented by a set of PTZ cameras.

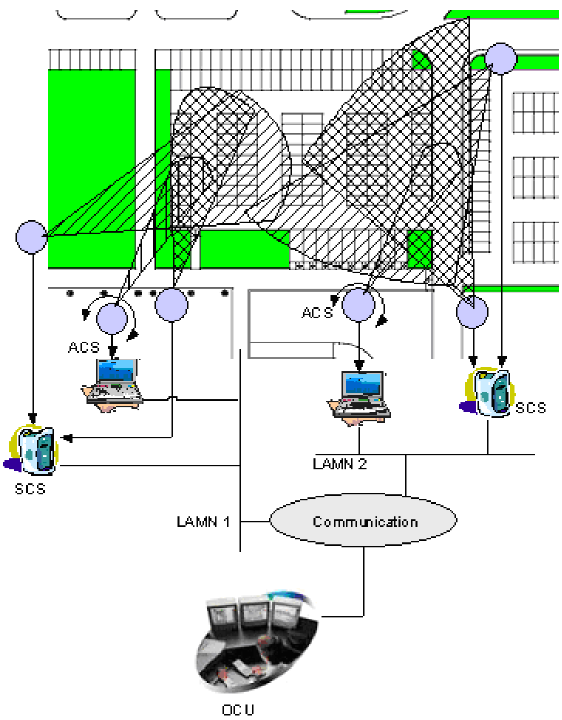


Figure 1 Logical architecture of the system

The hierarchy of the system is given by two main entities: the Local Area Monitoring Networks (LAMNs) and the Operative Control Unit (OCU).

LAMNs are organised in a hierarchical way to provide the monitoring of a local areas. Such an architecture employs the SCSs to detect and track all the objects moving within the assigned area, and the ACSs to actively track a particular object.

First level nodes compute trajectories and send important features to the OCU where all the trajectories computed by the first level nodes are analysed to detect possible dangerous actions. As consequence one or more ACS are tasked to acquire the object of interest with higher accuracy.

At the highest level of the hierarchy we developed a centralised control unit, the OCU, where all the information collected by the single networks is received and processed to understand the activities occurring inside the monitored area. One of the OCU modules is the trajectory analysis system described in this paper. This module processes the raw data on object trajectories and performs a high-level analysis in order to point out behavioural information about the tracked objects. In particular the trajectories are grouped into clusters and the clusters are used to analyse the frequency of the most common patterns. This kind of information can be used to perform statistic modelling of the behaviours, for example identifying those activities that are uncommon if compared to previously acquired data.

3 Camera system

The active camera system allows active tracking of objects of interest. Indeed, by means of active vision techniques it is possible to adapt the image acquisition process for security purposes. Moreover, the ability to control gaze as well as to acquire targets at high resolution gives an augmented perception of the monitored scene. Such a property turns out to be fundamental to support the behaviour understanding process, especially in the case of wide areas.

For such purposes, the system delivers a request for focusing the attention of the active sensor on a selected target inside the monitored area. Such a request is fulfilled by changing the pan and tilt parameters of the PTZ unit to head the camera toward the spot occupied by the target.

To maintain the gaze on the target a frame by frame motion detection system is adopted. To compute the image registration between two consecutive frames *i* and *i+1* the computation of a simple displacement vector $\mathbf{d}$ has been considered. A feature clustering technique [10] allows tracking of features on the background and to estimate which ones should be selected for the computation of the transform.

When a zoom operation is involved, since the registration techniques are not reliable to detect moving objects, we have adopted a technique that tracks a target by computing a fixation point from a set of features selected and tracked on the target. In such a scheme, after having deleted all the features that either belong to the background or to a cluster with cardinality lower than three, we can apply the technique proposed by Tordoff-Murray [11] over each cluster to determine the fixation point.

In order to task the active cameras the system needs to maintain tracks for all the objects that exist in the scene simultaneously and classify their behaviour.

Figure 2 Example of active tracking of an object inside the monitored area

To track all the objects a first processing may occur locally to each image plane. For each sensor, the system can perform an association algorithm to match the current detected blobs with those extracted in the previous frame. Generally, a 2D top view map of the monitored environment is taken as a common coordinate system. The pixel usually chosen to represent a blob and be transformed into map coordinates is the projection of the blob centroid on the lower side of the bounding box. Data obtained from the different sensors (extracted features) can be combined together to yield a better estimate. A typical feature to be fused is the target position on the map. A simple measurement fusion approach through a Kalman filter [12], was employed for this purpose. This fusion scheme involves the fusion of the positions of the target (according to the different sensors) obtained right out of the coordinate conversion function.

The fused state $\hat{\mathbf{x}}_{k|k}$ on the position and velocity of a target is therefore obtained fusing the local estimates $\hat{\mathbf{x}}^i_{k|k}$ and $\hat{\mathbf{x}}^j_{k|k}$ from sensors i and j respectively as follows:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}^i_{k|k} + [\mathbf{P}^i_{k|k} - \mathbf{P}^{ij}_{k|k}][\mathbf{P}^i_{k|k} + \mathbf{P}^j_{k|k} - \mathbf{P}^{ij}_{k|k} - \mathbf{P}^{ji}_{k|k}]^{-1}(\hat{\mathbf{x}}^j_{k|k} - \hat{\mathbf{x}}^i_{k|k}) \quad (1)$$

where $\mathbf{P}^i_{k|k}$ and $\mathbf{P}^j_{k|k}$ are the error covariance matrices for the local estimates and $\mathbf{P}^{ij}_{k|k} = (\mathbf{P}^{ji}_{k|k})^T$ is the cross covariance matrix, which is given by the following recursive equation:

$$\mathbf{P}^{ij}_{k|k} = [\mathbf{I} - \mathbf{K}^i_k \mathbf{H}^i_k][\mathbf{F}_{k-1}\mathbf{P}^{ij}_{k-1|k-1}\mathbf{F}^T_{k-1} + \Gamma_{k-1}\mathbf{Q}_{k-1}\Gamma^T_{k-1}][\mathbf{I} - \mathbf{K}^j_k \mathbf{H}^j_k] \quad (2)$$

where $\mathbf{K}^s_k$ is the Kalman filter gain matrix for sensor s at time k, Γ_k is the process noise matrix at time k, and $\mathbf{Q}$ is the process noise covariance matrix.

The process involves the following steps for each target: (1) collection of measurements available from the local sensors; (2) grouping and assignment of the measurements to each target known at the previous time instant; (3) updating each target's state by feeding the associated filtered estimates to the fusion algorithm.

4 Trajectory analysis

Trajectory analysis is a difficult task to perform if the data are acquired by a single camera. In this case the trajectories are typically affected by many errors, due to the intrinsic ambiguity of the single camera model and to the errors introduced by occlusions (objects hidden by other moving objects or by parts of the background). The hierarchical sensor network described in the previous sections can solve many of these problems using sensor fusion and can gather robust information about the trajectories of the moving objects. Once the robust trajectories are obtained, they can be further processed at a higher level in order to perform trajectory analysis and consequently behaviour analysis and recognition.

In order to work on meaningful data for high-level processing, trajectories need to be clustered together, grouping in the same cluster the trajectories with similar features; therefore a suitable way to represent trajectories and clusters is needed.

In the proposed method, trajectories are simply represented as a list of coordinates acquired at fixed time intervals and representing the position of the moving object projected on the site map. Clusters are represented in a similar way, but also include information on the local variance (the width of the cluster at a given point). With this representation it is easy to define a proper similarity measure in order to check if a trajectory matches a given cluster, based on the distance minima between the trajectory and cluster elements. It is also straightforward to define a cluster updating formula, so that a cluster can be dynamically adapted to the matching trajectories. The full process of cluster updating and the similarity measure are described in detail in [13].

Since our final aim is to study trajectories for behaviour modelling, we developed a way to represent trajectories that can highlight behavioural similarities between clusters of trajectories. In particular, we claim that, in many real-world environments, the clusters are not totally independent each other and often have pieces of trajectories in common, generally due to environmental constraints (for example, all the trajectories can share the same origin in proximity of a gate). A behaviour analysis system should be aware of these shared features. In particular we want to explicitly model common prefixes (where a *prefix* is the initial part of a cluster) because the identification of those prefixes will allow making predictions on the future movements of an object. We choose to model each common prefix with a cluster, in such a way that trajectories are no longer represented by simple clusters, but by groups of clusters organized in a tree-like structure, as shown in Figure 3.

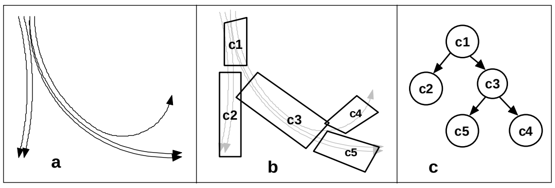


Figure 3. Representing trajectories as trees of clusters

The whole process of building trees of clusters is summarized in Figure 4. When a new trajectory is detected, its similarity with each of the root-node clusters is computed, in order to see if the trajectory falls inside a cluster. If a match is not found a new cluster is instantiated, otherwise the cluster is updated with the incoming data. The system continuously computes the similarity measure to check if the trajectory is leaving the cluster or not. If the trajectory is exiting from the cluster's end, a new match is found among the children of the just-left cluster (for example in Figure 3 if a trajectory is exiting from cluster c1 a new match is searched with c2 and c3). Otherwise, if the trajectory is exiting far from the end of the cluster, the cluster itself is split in two different parts, as shown in Figure 5.

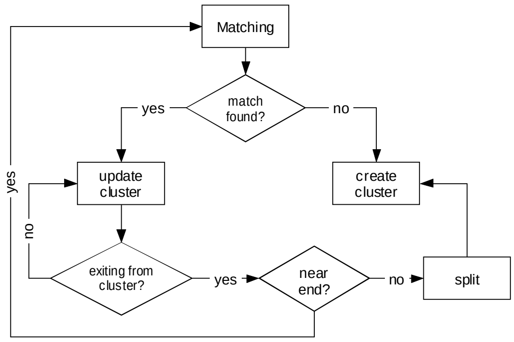


Figure 4. Tree building process

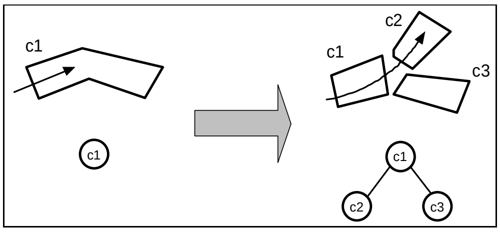


Figure 5. The split of a cluster

5 Behaviour Analysis

The clustering technique proposed in the previous section is a good way to organize the raw data acquired by the low-level tracking system in a format that can be useful for an high-level analysis module. Our final aim is in fact to give a high-level, semantic interpretation of the monitored scene, in which some particular behaviours should be detected. Since the proposed system is developed for video surveillance purposes, we are mostly interested in anomalous, dangerous behaviours.

There are two possible approaches in the definition of what an anomalous behaviour is:

- by explicit modelling
- by probabilistic definition

In the first case the system has a complete description of what an anomalous behaviour is, and recognize it through the comparison with the behaviour models stored in the system. This approach is difficult to implement, because the right features that best describe a behaviour must be chosen, and the matching procedure should be very robust to noise. Furthermore, the system must know *a priori* all the possible kinds of behaviours and must have their description. Note however that the models do not need to be explicitly built by a human operator. They can for example be learnt by an automatic, unsupervised machine learning approach (e.g. using unsupervised, auto-associative neural networks).

To overcome the problems of the explicit modelling approach, in this paper we investigate the probabilistic way of describing anomalies. This approach is based on the simple consideration that an anomaly is something that, by definition, happens rarely and differs from the normal patterns of activity. Of course not all the anomalous events are dangerous ones, but most probably a dangerous behaviour is an anomalous one, thus justifying the study of anomaly identification.

The framework proposed so far gives a natural way to analyse and identify trajectory anomalies, allowing to associate probabilities to each sequence of clusters. If we keep track of how many times each cluster is updated, each arc of a tree of clusters can be labelled with a probability defined as $n_c \Big/ \sum_k n_k$, where n_x is the number of times the cluster *x* was matched, *c* is the child node of the arc and the index *k* cycles over all the indices of the children of the parent node of the arc. Using this approach, trees like the one shown in Figure 6 can be obtained.

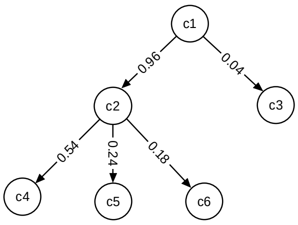


Figure 6. Tree of clusters with probabilities.

In this example the past observations led to a computed probability of 0.04 of going from cluster c1 to cluster c3, of 0.96 of going from c1 to c2 and so on. These probabilities can be used to calculate the global probability of a fully developed trajectory. For example, if a trajectory starts at cluster c1, then moves to c2 and finally to c5, its a-priori probability is the product of the probabilities associated to the arcs being traversed, in this case 0.96*0.24=0.2304. These probabilities can easily be used to identify anomalous trajectory behaviours, because such behaviours will simply be the ones associated to a small probability. Of course if a trajectory does not match any cluster its probability is not defined, but this would say that the trajectory is totally new to the system, and thus should anyway be considered as anomalous.

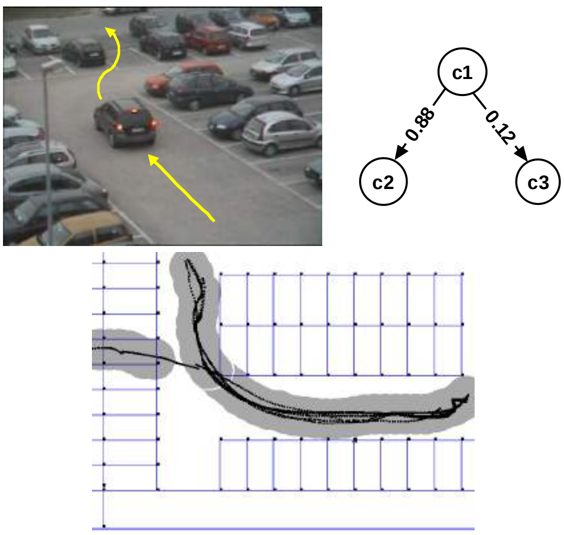


Figure 7. Trajectories and clusters in a parking lot

Figure 7 shows an example of the probabilistic approach. The camera is monitoring a parking lot, and in the map are drawn the last car trajectories detected. The car shown in the picture is following an anomalous trajectory, exiting from a non authorized point. The main cluster representing the previous trajectories is split and the new situation is represented by three clusters, organized in a tree structure as shown in Figure 7. Note how the new trajectory has a lower probability (0.12) respect to the others, thus it is possible to identify it as an anomalous and potentially dangerous behaviour.

6 Conclusions

In this paper we have proposed a multi-sensor, behaviour-oriented surveillance system based on trajectory analysis. The trajectories are first robustly computed thanks to a hierarchical network of both fixed and active cameras; once detected, the behaviour analysis module groups trajectories in clusters and perform a probabilistic identification of anomalous behaviours.

Future works will involve the extension of the probabilistic approach to other kind of data, since trajectories are just one of the many features that can be collected about a moving object. We are also planning to study a possible integration of the probabilistic approach with the explicit-model approach described in section 5.

References

1. C. Regazzoni, V. Ramesh, and G. Foresti, Special issue on video communications, processing, and understanding for third generation surveillance systems, *Proceeding of the IEEE*, vol. 89, no. 10, pp. 1355-1539, Oct. 2001.
2. L. Davis, R. Chellapa, Y. Yaccob, and Q. Zheng, Visual surveillance and monitoring of human and vehicular activity, *Proceedings of DARPA97 Image Understanding Workshop*, 1997, pp. 19–27.
3. R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, Algorithms for cooperative multisensor surveillance, *Proceedings of the IEEE*, vol. 89, Oct. 2001, pp.1456–1477.
4. G. Foresti, P. Mahonen, and C. Regazzoni, *Multimedia Video-Based Surveillance Systems: from User Requirements to Research Solutions*. Kluwer Academic Publisher, Sept. 2000.
5. D. Koller, K. Daniilidis, and H. H. Nagel, Model-based object tracking in monocular sequences of road traffic scenes, *Int J Computer Vision*, vol. 10, pp. 257–281, 1993.
6. Zhigang Zhu, Guangyou Xu, Bo Yang, Dingji Shi, and Xueyin Lin, VISATRAM: a real-time vision system for automatic traffic monitoring, *Image and Vision Computing*, Vol. 18, no. 10, 2000, pp.781-794.
7. S. Dockstader and M. Tekalp, Multiple camera tracking of interacting and occluded human motion, *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1441-1445, 2001
8. J. Gluckman and S. Nayar, Ego-motion and omnidirectional camera, *IEEE Int. Conf. on Computer Vision*, Bombay, India, Jan.3-5 1998, pp. 999–1005.
9. D. Murray and A. Basu, Motion tracking with an active camera, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 5, pp. 449–454, May 1994.
10. C.Micheloni, G. Foresti and F. Alberti, A new feature clustering method for object detection with an active camera, *IEEE Inter. Conf on Image Proc.*, Vol. 4, 24-27 Oct. 2004, , pp. 2587-2590.
11. B.J. Tordoff and D.W. Murray. Reactive control of zoomwhile tracking using perspective and affine cameras. *IEEE Trans. on Pattern Anal. Machine Intell.*, 26(1):98–112, 2004.
12. J. B. Gao and Chris J. Harris, Some remarks on Kalman filters for the multisensor fusion, *Information Fusion*, Vol. 3, n.3, pp.191-201, 2002.
13. C. Piciarelli and G. Foresti. Toward Event Recognition Using Dynamic Trajectory Analysis and Prediction, *IEE Int. Symp. on Imaging for Crime Detection and Prevention*, pp. 131-134, London, UK, 2005.

A multi-sensor surveillance system for public transport environments

J.-L. Bruyelle [1], L. Khoudour [1], S.A. Velastin [2]
[1] *INRETS-LEOST, Villeneuve d'Ascq, France*
[2] *Kingston University, Kingston upon Thames, UK*

Abstract
To improve safety and efficiency, camera networks in public transport get bigger and bigger. However, operators can only monitor a few cameras at a time, hence the need to automate the monitoring tasks through computer vision. We describe an approach consisting of small, localized computers each dedicated to the local image processing of only one or two cameras, interconnected through a standard computer network carrying only the small amount of useful data extracted from the image processing.
This short paper describes the implementation of this system, as well as the four tasks that have been implemented on it. These examples belong to both incident detection and traffic measurement tasks, and are better implemented in the decentralized architecture we propose. These processes have all been intensively tested in real life conditions within several types of transport network, such as an airport and underground railways. Trials show the efficiency of the tested solutions.

Keywords
Camera network, public transport, video monitoring, image processing, incident detection.

1 Introduction
Public transport operators are facing an increasing demand for improvements in efficiency and security, both from the public and the governments. As part of their effort to meet these demands, they are making an ever-increasing use of video cameras, in order to measure the performance of their networks, detect traffic incidents and criminal or antisocial offences. However, the efficiency of these large camera networks is impaired by the inability of human operators to watch a large number of video screens at the same time.

Over recent years, image processing has grown as an appealing solution to "watch the screens" more effectively, detect incidents more reliably, and make accurate measurement of traffic parameters that could only be grossly estimated by human operators. Another advantage of computers is that they do not suffer from visual fatigue, and are not distracted by external events, so they are, theoretically, 100% reliable.

However, although the solutions found are usually very effective on a single camera, the integration of many cameras remains a problem. Historically, the cost of processing each camera separately with a dedicated computer was high, so the preferred solution is to use a single, centralized computer to process all the cameras in a time-shared way (e.g., process one frame from camera 1, then one frame from camera 2, etc.). In addition to its implementation complexity and its lack of flexibility, this approach reduces drastically the capability to react to incidents, which in turns impairs the interest of such systems.

However, the ever-decreasing cost and size, as well as the ever-increasing processing power of small computers, allow foreseeing the capability to fit an image processor to each camera in the system. This allows keeping the capacity to respond in real time to fast-occurring incidents, even in large networks, while keeping the programming and operating loads to a minimum. Another interesting feature of this approach is that it allows transmitting over the network only the amount of data that is really necessary - for instance, the images of detected incidents, or the calculated statistical data extracted from the image shot by the camera. This allows implementing a complete camera network using standard computer networking technologies such as Ethernet, further reducing the installation and operation cost and hassles.

In the framework of the EU's PRISMATICA program [1, 2, 3, 4], the authors have devised and tested in real-life conditions a practical camera network using these techniques. Typical functionalities useful to public transport operators have been implemented on this network : they are presented in section 3 of this paper.

2 Architecture of the network
The general architecture of the camera network is illustrated on Figure 1. The prototype camera network we have implemented uses a small, shoebox-type industrial PC to process locally each camera in the network (Figure 2).

Each processor is linked to the global Ethernet network via the closest available socket. Thanks to the long runs of cable allowed by Ethernet, a processor and camera can be installed virtually anywhere, providing an Ethernet and a power socket are available. One of the advantages of this flexibility is the ease to reconfigure the camera network by just moving the cameras and processors to different places.

The center of the network is a supervising computer, normally located in, or close to, the control room. This computer gathers the data sent by the image processors, and runs the human-machine interface that displays the data to the operators in the control room. It is also in charge of choosing, among the incoming data, the most relevant ones to be displayed.

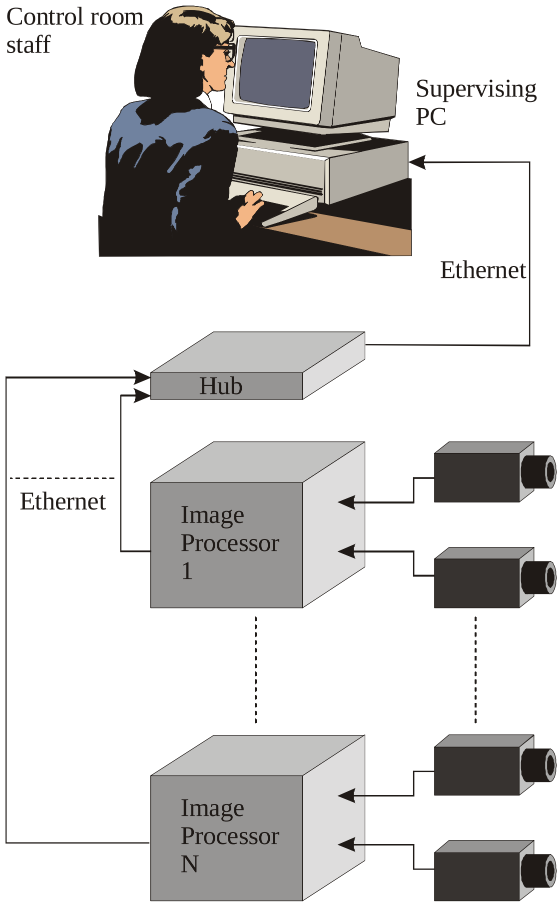


Figure 1. *General architecture of the system.*

Figure 2. *Local image processing computer.*

Configuration of the local processors is done from the control room, on the supervising computer, via a technical interface. The configuration data is then uploaded to the processors via the Ethernet network. This is probably the only part of the operation that is best done centrally, and the decentralized architecture easily allows it. Using multitasking and the concurrent nature of the Ethernet, it is not necessary to stop the normal operation of the camera network when configuring a processor.

In order to further reduce the cost and number of Ethernet addresses, it is possible to connect a second camera to the local processors, allowing them to process two nearby cameras. This capability is offered if the total required processing power does not exceed the capabilities of the little shoebox computer.

3 Main applications

Several functions have been implemented in the network, which can be divided in two classes: incident detection functions and passenger flow measurement functions.

3.1 Incident detection functions

These functions detect abnormal situations, such as intrusions in forbidden areas or left luggage, and then send a message to the control room, along with sufficient details for the staff to evaluate the situation, namely the image that has raised an alarm, and highlighting the part of the image where the incident lies. The functions that have been implemented are the following :

- **Intrusion in forbidden or dangerous areas**

  The aim of this application is to improve the safety by detecting, automatically and in real time, people entering areas which are not allowed or hazardous. This type of incident can imply malevolent behaviour, but also such events as a fire, causing people to flee by any available exit. For this application, detection speed is crucial, and the use of cameras, as opposed to conventional sensors (e.g. door switches or optical barriers), allows more flexibility (one camera can cover more surface than a door switch, and can be reconfigured by software much more easily than an optical barrier which require physical reinstallation) and can provide more information to the security staff, by providing them with images of the incident.
  For our purpose, intrusion is visually defined as an "object" (in the most general sense of the term), having the size of a human being, moving in the forbidden zone
  This definition requires to solve three separate problems :
    - Detect something that moves.
    - Define its size.
    - Decide whether or not it is in the forbidden area.

  All these being done in real time, with the available processing power (which is not much in the case of our miniature local processors), and with the objective of no missed alarms and as few false alarms as possible [4, 5].

- **Abnormal stationarity**
  This section will deal with automatic detection of motionless persons or objects using CCTV cameras (Figure 3). In the public transport network context, this type of detection is required for reacting quickly to events potentially impacting user safety (e.g. drug deals, a thief waiting for a potential victim, person requiring assistance, unattended objects) or decreasing the system's attractiveness (presence of vagrants, beggars, illicit vendors, graffitists).

  In our context "stationarity" has to be understood in the broadest sense. Indeed, in the application this word does not always mean "completely motionless". In fact, even a person standing still makes motion with his legs, arms, etc. Thus, in the following, we will mean by "stationarity" a person or an object located in a small image area during a given period of time.

The fact that a person may move adds complexity to the detection process. Other difficulties depend on the site and the network. For instance, the camera view angle may be low, increasing the occlusion phenomena. If the site is crowded, the "stationary" person or object will be partially or completely occluded from time to time, making it more difficult to detect and to estimate its stop duration. The lighting environment may affect the perception and the detection too. As a matter of fact, the method to create needs to be robust to contrast changes, to deal with occlusions, and to take into account the motion of a "stationary" person.

This function averages the motions in the passenger flow, and detects whether some part of the image, not being part of the background, remains motionless for more than a user-defined time threshold (usually two or three minutes) [6]. This can be a passenger (someone who has collapsed or fallen, a beggar, etc.), left luggage or an abandoned (potentially suspicious) package. All the above functions were designed taking into account the necessary immunity to changing lighting conditions (ageing fluorescent tubes, alternating sun & clouds, etc.) and to the possible changes in the background (e.g. thrown objects), which used to cause frequent false alarms in older similar systems, reducing their actual effectiveness.

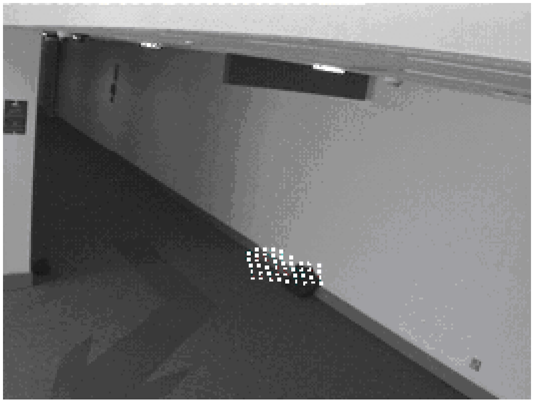


Figure 3. *Detection of left luggage.*

3.2 Passenger flow measurement functions

The purpose of these functions is to gather statistics related to the traffic of passengers, both as a means to ease the mid-term and long-term management of the Public Transport Operators (the data is added to a data base which can be exploited with software such as a spreadsheet), and as an immediate warning system to raise alarms when too many passengers are waiting at e.g. the ticket counters, so that appropriate steps can be taken more effectively. These "alarms" are, of course, of a different nature from those raised by incident detection functions, as they usually require sending commercial or customer services staff, rather than sending security or medical staff.

The measurement functions which have been incorporated in the local camera network are the following :

- **Passengers counting**
  The considerable development of passengers traffic in transport systems has quickly made it indispensable to set up specific methods of organization and management. For this reason, companies are very much concerned with counting passengers traveling on their transport systems. A passenger counting system is very important especially on the following points : best diagnosis on the characteristics of the fraud, optimization of line management, traffic control and forecast, budgetary distribution between the different lines, improvements of the quality of service.

  Our counting function uses an active infrared barrier, incorporating a video camera (Figure 4 and 5), in order to count the passengers passing through a specific location [7]. The system is able to give the number of passengers in both directions, and sends them to the operator's console and/or to a log file, at timely intervals. Optionally, it can also measure the speed of the passengers.

Figure 4. *The camera used for passengers counting (right), with its infrared lighting system (221 LED's). The camera on the left is used for reference manual counting*.

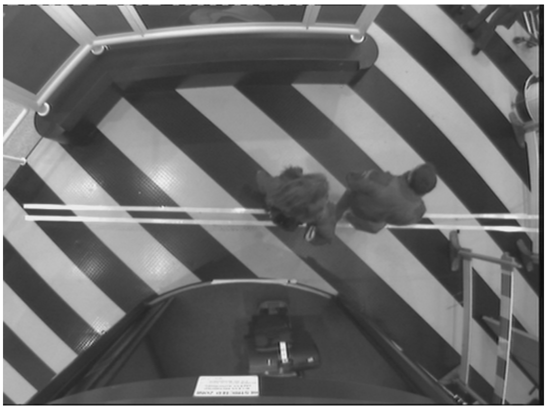


Figure 5. *Image shot by the camera of Figure 4. Two passengers very close to each other, crossing the barrier.*

- **Queue length measurement**
  This section will deal with automatic measurements of passenger queues using CCTV cameras. This type of research was initiated within the European CROMATICA (CROwd MAnagement with Telematic Imaging and Communication Assistance) project [4]. It continued within the European PRISMATICA project laying emphasis on performance assessment [1]. For this purpose the system was installed at Newcastle airport and several hours of video data were recorded during off peak and peak periods.

Measured information is both the queue length and the time spent in a queue. This is useful for, at least, three purposes :

- To increase passengers' comfort through information about the average time they will wait in the queue.
- To get statistics and, thus, an estimate on the crowd management efficiency.

These statistical data may be used, *a posteriori*, to improve the site design and layout, to distribute the employees in a more suitable way, to raise the efficiency of ticket offices, cash dispensers, check-in desks and custom desks, and to check continuously if the crowd management is effective and reacts immediately to any disturbances.

Figure 6. *Measurement of a queue.*

This function uses a camera overlooking the counters with a very wide angle lens (Figure 6), and can measure simultaneously the length of several queues, using a user-defined position of the start of each queue (this is done once for all, using a simple user interface that runs on the supervising PC, so no tricky intervention in the field is necessary). It can then send these lengths to the supervising PC, for immediate display and/or to a log file for later use. Optionally, a length threshold can be defined by the user to raise an alarm when one or more queues become too long.

4 Conclusion

This paper presented the distributed camera network developed as part of the PRISMATICA project, and four applications that have been implemented on this network for the purpose of the incident detection and statistical data gathering of passenger flows in public transport systems. The whole system was successfully tested in real-life conditions in airports and underground metropolitan railways. It must be mentioned, however, that the network and the functionalities are independent from each other : the network can support many other applications, and the functions can be implemented on a completely different architecture.

The work described here was conducted as part of the EU project PRISMATICA (GRD1 – 2000 – 10601). The project involved major European metro operators (RATP-Paris, LUL-London, ATM-Milan, STIB-Brussels, PPT-Prague, ML-Lisbon), research centers (King's College London, University College London, Kingston University, INRETS-France, CEA-France) and commercial companies (TIS-Portugal, SODIT-France, FIT-Italy, ILA-Germany, Thales-France). The authors are grateful to London Underground, the Paris Metro and Newcastle International Airport for providing access to their sites and staff.

References:

1. http://www.prismatica.com
2. Velastin S.A., Lo B., Sun J., Vicencio-Silva M.A, Boghossian B (2005). *PRISMATICA: Toward Ambient Intelligence in Public Transport Environments*, IEEE Systems, Man and Cybernetics (SMC), Special Issue on Ambient Intelligence, Vol. 35/1, pp. 164-182.
3. J. Sun, S.A. Velastin, B. Lo, M.A. Vicencio-Silva and L. Khoudour (2004), "A Distributed Surveillance System to Improve Personal Security in Public Transport", in *Knowledge-Based Media Analysis for Self-Adaptive and Agile Multimedia Technology* (P Hobson, E Izquierdo, Y Kompatsiaris & N E O'Connor: Eds), Proceedings of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT-04), ISBN 0-902-23810-8, 25-26 November 2004, London, pp. 7-14
4. Khoudour L., Deparis J.-P. (1997). Project Cromatica. *9th International Conference on Image Analysis and Processing*, Florence, Italy, September 17-19, 1997.
5. Deparis J.-P., Khoudour L., Cabestaing F., Bruyelle J.-L. (1997). Fall on the tracks and intrusion detection in tunnels in public transport. *EU-China Conference on ITS and Transport Telematics Applications*, Beijing, 3-6 June, 1997.
6. Bouchafa S., Aubert D., Bouzar S. (1997). Crowd motion estimation and motionless detection in subway corridors by image processing. *ITSC 97*, Boston, USA, November 9-12, 1997.
7. Khoudour L. (1997). Analyse spatio-temporelle de séquences d'images lignes. Application au comptage des passagers dans les systèmes de transport. *PhD thesis*, Université des Sciences et Technologies de Lille, January 1997.

Zebra fish, a new behavioral model system

Robert Gerlai [1] , Stéphane Berghmans [2], Irina V. Zhdanova [3], Laure Bally-Cuif [4], Su Guo [4]
[1] Symposium Organizer, *Department of Psychology, University of Toronto, Toronto, Canada*
[2] *DanioLabs Ltd., Cambridge, UK*
[3] *Department of. Anatomy and Neurobiology, Boston University, Boston, USA*
[4] *Zebrafish Neurogenetics Department, GSF-Research Center, Neuherberg, Germany*
[5] *Department of Biopharmaceutical Sciences, University of California at San Francisco, San Francisco, USA*

Abstract

Zebra fish has been a favored species in developmental biology and numerous genetic tools have been developed for this species over the past two decades. Zebra fish is small, easy to maintain, and prolific, attributes that make this species particularly amenable for high throughput screening. Indeed, numerous mutagenesis studies have been conducted in which mutation induced changes could be detected after the screening of thousands of fish. Most of these studies, however, focused on embryological questions. Given that zebra fish is a vertebrate organism with brain structure and function similar to other vertebrates, it should be an excellent study species for neuroscience and behavioral neuroscience too. However, the behavior of this species is rarely studied and behavioral test and quantification methods were previously unavailable. Behavioral tests that can tap into a diverse set of brain functions and the development of high throughput or automatable versions of such tests would be an important step. These tests would allow one to conduct large scale mutagenesis studies with a focus on brain function and would also allow one to test a large number of candidate compounds (small molecules) in pre-clinical drug screening studies.

The symposium will bring together a number of experts in the rather small but rapidly developing zebra fish behavioral neuroscience field. These experts will present their latest behavioral test designs and discuss the technical aspects of how to make the tests high throughput. Furthermore, they will present their latest discoveries that could be made only with the use of the novel test paradigms.

Keywords

Addiction, behavioral phenotyping, drug screening, sleep, zebra fish

1 Relevance for the Conference

The behavior of zebra fish is virtually uncharted. Novel behavioral paradigms presented at the symposium are expected to have a significant impact on how we can identify drug and mutation induced changes in brain function in a fast and reliable manner. The talks presented at this symposium should be of interest to a broad range of scientists including those studying substance abuse, sleep, cognition, and numerous other CNS functions and abnormalities. The technological challenges associated with the development of these behavioral tests and the development of high throughput behavioral quantification methods make the symposium particularly appropriate for the conference.

2 The outline of the symposium

Five talks will be presented at the symposium. The talks are briefly summarized in the order of presentation below.
Stéphane Berghmans will present on the establishment of neurobehavioral paradigms and behavioral quantification methods with which his company, Danio Labs Ltd., can detect drug induced alterations in brain function of the embryo or the adult fish. Dr. Berghmans' paradigms tap into a range of behaviors including simple startle response or locomotor activity to more complex characteristics such as drug addiction or sleep. Dr. Berghmans will discuss plans to establish a collection of scalable behavioral paradigms highly relevant to high throughput and automatable screens to be used for drug discovery, drug reprofiling and safety pharmacology and will present on the latest progress in neurobehavioral test designs and their utility in drug screening using the zebra fish.

Irina V. Zhdanova investigates sleep-related processes in zebra fish, a diurnal vertebrate with a robust circadian pattern of daytime activity and nighttime rest. Dr. Zhdanova's results show that rest state in this fish has important similarities with sleep in mammals and can be considered a sleep-like state. She uses high-throughput image-analysis techniques to continuously record locomotor activity patterns in hundreds of zebrafish larvae individually housed in the wells of 96-well plates. She will explain that sleep-related compounds are often small size molecules and thus they easily penetrate larval skin and can be administered directly into the water, providing an effective non-invasive drug delivery which facilitate pharmacological analysis of sleep in this species. She will also argue that the relative ease of developing transgenic fish lines and the possibility of continuous monitoring of reporter-linked gene expression in transparent transgenic zebra fish as well as the ease with which chemical mutagens induced random mutations may be generated make it possible to discover and study the role of novel genes in sleep process in zebra fish.

Laure Bally-Cuif will report on the development of a reliable Conditioned Place Preference (CPP) test in adult zebra fish, and demonstrate that adult zebra fish show robust conditioned place preference induced by the psychostimulant, D-amphetamine. Dr. Bally-Cuif will also show that conditioned place preference is dramatically reduced by genetic disruption of acetylcholinesterase (AChE) function in ache/+ heterozygotes, a reduction that cannot be accounted for by concomitant defects in exploratory activity, learning, and visual performance. She will demonstrate that the cholinergic system is a modulator of drug-induced reward in zebra fish, and thereby will validate the zebra fish as a model to study the molecular neurobiology of addiction in vertebrates. Dr. Bally-Cuif's results also demonstrate that AChE is a promising target for systemic therapies against addiction to psychostimulants, and open the way to screening for dominant mutations affecting psychostimulant-induced reward in zebra fish.

Robert Gerlai will present novel behavioral tests developed to measure alcohol induced changes in a simple and fast manner including the novel tank, the social preference, the aggression, and the predator model tasks. These tests are designed to enable automation, and to tap into diverse brain functions potentially influenced by alcohol. Behavior is quantified using event recording, to measure motor and posture patterns, and video-tracking, to analyze quantitative aspects of swim paths. Dr. Gerlai will compare these two methods with the development of automated behavioral quantification in mind. The results suggest that a video-tracking system, e.g. EthoVision, is capable of properly tracking and quantifying swim path patterns of the small zebra fish and that behavioral measures quantified with the use of event recording and with the use of video-tracking do not generally correlate with each other but both methods are capable of detecting changes induced by alcohol. Dr. Gerlai will discuss the implications of these findings for high throughput screening.

Su Guo will focus on the quantification of psychostimulant modulated locomotor activity, conditioned place preference, and "self-administration"-like behaviors. Dr. Guo will present video tracking data on the effects of alcohol on locomotor activity of larval zebra fish demonstrating that alcohol exhibits a biphasic response on activity: at low concentrations, it increases locomotor activity, and at high concentrations it depresses it. Using conditioned place preference adopted from rodent studies, Dr. Guo found that both food and morphine can elicit a significant place preference. Dr. Guo's choice chamber assay is also able to monitor morphine and food preference behavior in larval zebrafish. She will discuss the current and future efforts to make these assays robust and high throughput for the purpose of large-scale genetic screening.

Zebrafish: a novel source of behavioral paradigms for safety pharmacology and drug discovery

S. Berghmans, J. Hunt, P. Goldsmith, A. Fleming
DanioLabs Ltd., Cambridge, United Kingdom

Abstract
DanioLabs have designed a panel of neurobehavioural tests using zebrafish larvae. Here we describe the utility of these tests for medium to high throughput screening for safety pharmacology, drug discovery and drug reprofiling. Methods for the assessment and quantification of behavioral paradigms are used to monitor zebrafish larvae drug induced alteration of the brain function. The behaviors examined range from the simple "startle response" to more complex behaviors such as epilepsy.

Keywords
drug discovery, locomotor activity, neurobehavior, safety pharmacology, zebrafish

1 Introduction
The zebrafish model system is amenable to high throughput screening because of numerous advantages, including the relative ease of maintaining large stocks of animals; its high fecundity, which provides the investigator with large numbers of animals to analyze; and the rapid embryonic development *ex utero*, which facilitates experimental manipulation and allows the direct observation of tissue formation and organogenesis *in vivo* [8]. The organization of the genome and the genetic pathways controlling signal transduction and development are highly conserved between zebrafish and man [5]. These properties have established the zebrafish as an excellent model system that is relevant to studies of human diseases [3]. Recently, zebrafish have become the focus of neurobehavioral studies since larvae display learning, memory and behavior phenotypes that are quantifiable and relate to those seen in man. It is also known that the zebrafish brain structure and function are similar to other vertebrates, lending further support to its use as a useful model for evaluating vertebrate behavior [7]. In addition, since the larvae can live in as little as 200 µl of fluid; only micrograms of compound are needed for screening. This makes compound screening possible with a different order of magnitude of synthetic chemistry resource, enabling in vivo analysis of compound action to be undertaken at much earlier stages in the drug development process, and at a higher throughput than hitherto possible. This is facilitated by the fact that zebrafish are DMSO tolerant and readily absorb compounds from the water.

Our company's goals are to develop zebrafish models of human disease and to identify candidate therapeutic compounds by screening compound libraries against the assays we develop. Based on our expertise in modeling diseases, we have also developed safety pharmacology assays to identify undesirable adverse effects of compounds early in their development process. We have focused on establishing assays addressing neurobehavioral paradigms and have developed quantification methods with which one can detect drug induced alterations in brain function of the zebrafish larvae. These models range from basic behavior such as locomotor activity (LA) to more complex ones such as drug addiction or sleep. We are establishing a collection of scalable behavioral paradigms highly relevant to high throughput and automatable screens to be used for drug discovery, drug reprofiling and safety pharmacology. Our progress in neurobehavioral test designs using the zebrafish is discussed here with a special focus on their utility in safety pharmacology.

2 Materials and Methods
Zebrafish maintenance, breeding and staging of the WIK strain were performed as described [4]. Seven days post fertilization (dpf) zebrafish were placed individually in 96-well Multiscreen Nylon Mesh Plates (Milipore, France) containing 200 µl E3 medium. For analysis of LA, behavior was monitored using an apparatus (Tracksys, United Kingdom) containing an infra-red light source and an infra-red high resolution camera (Sony XC-EI50). 60 minute recording sessions of 96 larvae were analyzed using EthoVision 3.1 locomotion tracking software Noldus Information Technology, Inc. (Wageningen, The Netherlands).

3 Results
Spontaneous LA is a general starting point for CNS safety pharmacology evaluation in rodents. Our zebrafish assay evaluates the effects of a drug on LA by recording and analyzing the behaviour of 7 dpf larvae using video-tracking (Ethovision 3.1, Noldus, Wageningen, The Netherlands). LA is used in safety pharmacology to evaluate drug adverse effects on the CNS as it integrates function of several central nervous system structures. In addition, this assay can also identify effects on other components of locomotion such as the integrity of the peripheral nervous system or musculature. From a drug discovery perspective, such an assay is central in establishing efficacy assays where compounds are screened for specific therapeutic indications such as anti-convulsant activity or sleep promotion.
Drugs known to cause hyperactivity in rodents such as apomorphine or caffeine have similar effects in zebrafish larvae as identified in our assay. Figure 1 shows that a biphasic dose-dependent increase in LA, as expressed by the relative distance moved (distance moved in relation to the amount of distance moved by control fish), is detected for caffeine. In addition to monitoring the fish behavior, startle responses triggered by a series of auditory stimuli are elicited at the end of the assay as an indicator of arousal level.

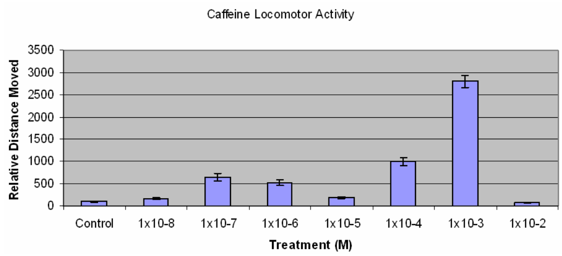


Figure 1. Caffeine hyperactivity dose reponse (N=12)

Figure 2 shows the dose-dependent increase in arousal threshold for the same larvae as those described in figure 1 when treated with increasing doses of caffeine. The monitoring of a startle response can also confirm that the low LA found for the highest caffeine dose in figure 1 is caused by toxicity as larvae do not respond to the auditory stimulus.

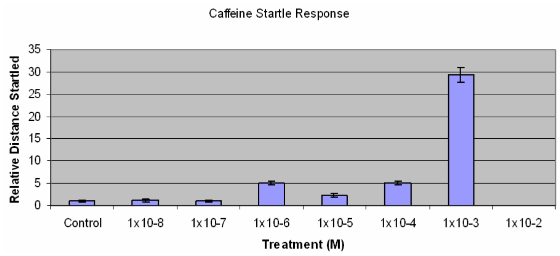


Figure 2. Caffeine dose-dependent arousal level (N=12).

Sedation is a key adverse effect screened for in new drug candidates. Safety pharmacology currently uses rodent assays to evaluate the potential hypnotic effect of drugs and zebrafish would be an ethical upgrade. Zhdanova and colleagues have established zebrafish larvae as a model for sedation using drugs like melatonin, pentobarbital and diazepam [9]. The sedative effect of these drugs as well as other sedative drugs such as chlorpromazine or ketamine (figure 3) can be detected in our assay.

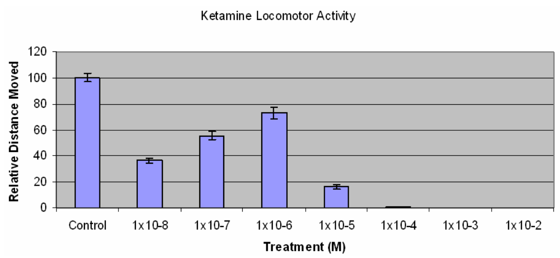


Figure 3. Ketamine dose-dependent sedative effect (N=12)

The assay is of further relevance to safety pharmacology as it can identify the potential undesirable pro-convulsant effect of a drug early in the development process. Seizures are elicited in zebrafish larvae by exposure to commonly used convulsant agents like pentylenetetrazole [1], pilocarpine and picrotoxin. Figure 4 shows that concentration-dependent epileptiform seizures induced by pentylenetetrazole (PTZ) result in stereotyped behavior expressed by increased LA.

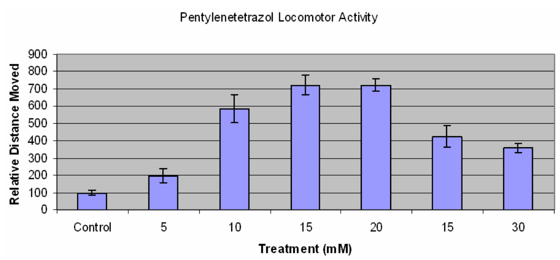


Figure 4. Pentylenetetrazole pro-convulsant dose reponse (N=12)

The speed at which the first seizures appear in larvae after exposure to PTZ is concentration-dependent [1]. A time response analysis applied to the larvae from figure 4 shows shorter latency to initial convulsions with higher PTZ concentrations (figure 5): LA is the highest during the first 15 minutes of recording for 30 mM whereas the highest LA at 5 mM occurs in the last 15 minutes of the assay.

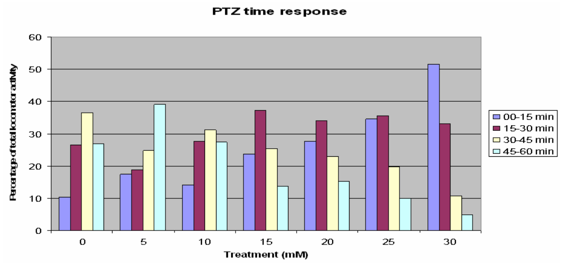


Figure 5. Concentration-dependent time response to PTZ (N=12)

Finally, the potential addictive effect of drugs destined to man is an important aim of safety pharmacology studies. It is known that zebrafish are affected by alcohol, cocaine [2] and nicotine [6]. The fact that our assay is able to detect a change in LA in zebrafish larvae exposed to drugs of addiction such as alcohol and amphetamine (figure 6) is extremely promising. Indeed, this opens the possibility of developing sensitization assays that would allow detection of novel drugs' addictive properties in zebrafish. It is also beneficial when considering launching drug discovery screening programs to identify molecules to treat drug addiction, a field in dire need of a scalable model.

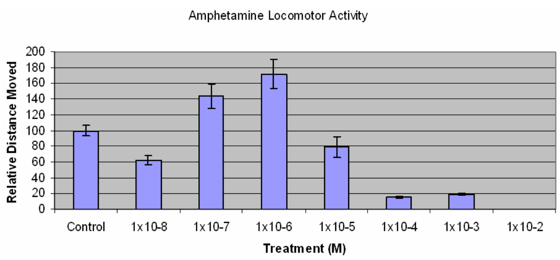


Figure 6. Amphetamine dose-dependent LA (N=12)

4 Conclusion

Performing safety pharmacology in zebrafish provides ethical advantages because it represents a substantial progress toward the "three Rs" principle (reduction, replacement and refinement of the use of animals in research). Zebrafish are also economically pertinent since turn-around time is rapid and compound requirement is small, allowing early screening to be undertaken very cost-effectively. Zebrafish are therefore a powerful pharmacological tool that is proving to be an attractive source of neurobehavioral paradigms for safety pharmacology as well as drug discovery.

References

1. Baraban S.C., Taylor M.R., Castro P.A., Baier H. (2005) *Neuroscience*, **131**, 759-68.
2. Gerlai, R. (2003) *Behav. Genet.*, 2003, **33**, 461-8.
3. Grunwald, D.J. & Eisen, J.S. (2002) *Nat. Rev. Genet.*, **3**, 717-24.
4. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. & Schilling, T.F. (1995) *Dev. Dyn.* **203,** 253-310.

5. Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., & Talbot, W.S. (2000) *Genome Res.*, **10**, 1890-902.
6. Svoboda, K.R., Vijayaraghavan, S., Tanguay, R.L. (2002) *J. Neurosci.*, **22**, 10731-41.
7. Tropepe V., Sive H.L. (2003) *Genes, Brain and Behavior*, **2**, 268-281.
8. Westerfield, M. (1993) *The Zebrafish Book*. Eugene, OR: University of Oregon Press.
9. Zdhanova I.V., Wang S.Y., Leclair O.U., Danilova N.P. (2001) *Brain research*, **903**, 263-8.

Event recording and video-tracking: towards the development of high throughput zebra fish screens

Robert Gerlai
Department of Psychology, University of Toronto @ Mississauga, Mississauga, Ontario, Canada

Abstract

Zebra fish is a popular subject of developmental biology and genetics. However, its brain function has rarely been studied and its behavioral characteristics are virtually uncharted. This paper is a feasibility study in which methodological aspects of behavioral analysis of zebra fish are considered in four paradigms: a novel open tank, social preference, aggression, and predator model tasks, tests that were originally designed to quantify acute alcohol responses. Quantification of behavior in these is conducted using manual recording of location & activity, computerized video-tracking, and observation based event recording of motor patterns. Video-tracking is found to reliably measure activity parameters and the location of fish thus allowing automated quantification of behavior. Motor and posture patterns did not generally correlate with video-tracking measures. It is argued that the simplicity of the behavioral tests coupled with automated video-tracking-based behavioral quantification will make them scaleable and thus appropriate for screening of mutation or drug induced changes in the zebra fish brain.

Keywords

behavioral phenotyping, event recording, forward genetics, video-tracking, zebra fish

1 Introduction

Zebra fish is a diploid vertebrate, a simple model organism with a complex brain. Its advantages are its small size and prolific nature and the wealth of information obtained about its development and genetics (for review see e.g. [3]). One of its drawbacks, however, is the paucity of proper behavioral paradigms that could detect alterations in its brain function. Given that the ability to detect and quantify such alterations is crucial for forward genetic analysis of mutation effects on brain function and for pharmacological analysis of the function of novel drugs, this is a serious problem. To address this, we have started to characterize zebra fish behavior and developed simple tests ([2]) including a novel open tank, a social preference task, an aggression test, and a predator model test. In the present paper we compare different methods of quantification of behavior in these tests: 1, manual recording of swim location and locomotor activity; 2, video-tracking-based analysis of swim paths; and 3, computer aided analysis of motor and posture patterns (event recording). Comparison of the manual recording and video-tracking data suggested that the latter could reliably measure the location and intensity of swim in zebra fish. Computerized video-tracking thus makes quantification of behavior automatable and will allow the experimenter to run multiple tests in parallel, i.e. to scale up the test, a necessary requirement for high throughput mutation or drug screening. Furthermore, analysis of motor & posture patterns and the video-tracking behavioral measures suggested that, in general, posture patterns did not correlate with location and activity measures. These results demonstrate the need for increasing the capabilities of video-tracking to enable it to capture some of the posture patterns but at the same time also suggest that computerized quantification of zebra fish behavior will be achieved with precision, and high throughput behavioral screening is feasible in this species.

2 Methods

2.1 Animals and Housing

118 adult, 3-6 month old, male and female zebra fish (*Danio rerio*) of an outbred genetically variable long-fin stock (Pet Pacifica, Honolulu, HI) were obtained and kept in community tanks with water temperature (26° Celsius) and light cycle (on at 7:00 and off at 19:00 h) controlled. Water was filtered by Fluval 404 canister filters containing filter foam (mechanical filtration), activated carbon (removal of organic waste and small particles) and BioMax rings (biological filtration). Fish were fed twice daily with a 50/50 mix of ground TetraMin flakes (Melle, Germany) and freeze dried krill (Aquatic Ecosystems Inc, Apopka, FL, USA).

2.2 General Experimental Procedures

The behavior of fish was recorded between 10:00 and 17:00 h in four test paradigms as described previously ([2]), i.e. first in the novel open tank, then in the group preference paradigm, and an aggression (mirror) test, and finally in a predator model task. In each test fish were placed individually into the experimental tank (20x25x12 cm, length x depth x width) and were monitored for 10 min. Inter-test interval was 2 min. A CCD camera (Panasonic WV-CP470) fed the live image (frontal view) into a computer (Dell Dimension 8300, pentium IV) which processed it using the EthoVision 3.0 video-tracking software (Noldus Info. Tech., Wageningen, The Netherlands). A second camera (Sony DCR TRV 70) recorded the frontal view of the experimental tank onto MiniDV tapes. These recordings were later downloaded onto the computer for manual analysis of activity and location and for event recording (Observer Color Pro, Noldus Info. Tech., Wageningen, The Netherlands)

2.3 Behavioral Tests

The tests have been described elsewhere ([1, 2]) and will be summarized here only briefly. Novel Open Tank. Exposure to a novel test chamber, as well as handling by the experimenter, is an inherent part of most laboratory animal behavioral tests. The novel open tank (21x12x24 cm) is intended to analyze behavior in response to these factors. Zebra fish placed singly in the open tank may exhibit elevated activity that habituates with time ([1]) and also show fear related behaviors ([2]). Subsequently, group preference is tested. Zebra fish is a social species. In nature and in the aquarium it forms schools, a group of individuals swimming close to one another. Individual zebra fish are motivated to join a school. This preference for the group was the basis of a behavioral test in which the effect of alcohol was investigated ([2]). The present test is a modification of this previously employed paradigm. Here, during a 10 min session, a single test fish can view a stimulus tank in each side of its test tank, one containing 10 stimulus zebra fish, and the other just fresh water. The positioning of the stimulus fish, i.e. whether they were presented on the left or the right side of the test tank was random across test fish. Subsequently, the aggression test (the inclined mirror task) was performed. Solitary zebra fish encountering another individual exhibit agonistic behavior, a response different in form and alcohol dose response characteristics from social behavior ([2]). A mirror placed at a 22.5 degree angle to the back of the aquarium resulted in a lateral view mirror image of the test fish that appeared closer when it swam on the left side of the tank allowing the experimenter to quantify how close it wanted to swim to its "opponent", a measure of aggressive tendencies that was previously shown to respond to acute alcohol treatment ([2]). The last test was the predator model test. Antipredatory behavior of zebra fish is adaptive and may be influenced by genetic factors (Gerlai, 1993; Csányi, 1986). Alcohol affects predator model elicited behavioral responses ([2]). Thus, this task may allow the detection of mutation or pharmaceutical agent induced functional changes in the brain. A predator model (a black falcon tube with eye spots) similar in size and shape to that used before by Gerlai et al. ([2]) was employed. It was placed into the stimulus tank (either on the left or the right side for each test fish) and was moved using a transparent plastic rod attached to its back during the 1st and 10th minute of the 10 min long test session.

2.4 Quantification of behavior

Manual quantification. As conducted previously ([2]), first zebra fish behavior was analyzed using manual quantification of location and activity of the fish. Video tapes were replayed on a Sony (DVCAM, DSR-11) MiniDV digital cassette player connected to a 14" JVC TV monitor. A transparency with a grid pattern placed on the monitor allowed the experimenter to record the duration of time fish spent in the upper or the lower half of the tank and on the left and the right side of the tank. The time the fish spent on the side opposite to the stimulus tank that contained the stimulus fish or the predator model and the time the fish spent in the upper half of the tank were statistically analyzed. In addition, the total number of times the fish entered the left, the right, the upper, and the lower half of the tank (shuttling activity) was also analyzed.

Video-tracking. Video-tracking (EthoVision Color Pro 3.0, Noldus, Wageningen, The Netherlands) is expected to allow quantification of swim path parameters more precisely than manual recording could and without the need for an experimenter to view video-tapes. Live input from a video camera fed into the computer through a piccolo video card was read by EthoVision and the background image taken before placing the test fish in the tank was compared to the live image (subtraction method) 10 times per sec. The altered pixel cluster with the largest surface area was interpreted as the target object, the test fish, and the X/Y coordinates of the center of this object were recorded. If no object of at least 25 pixels was located, the program recorded the coordinates of the last known location of the object (less than 5% of the samples). Parameters quantified were as follows: Mean distance from bottom: the distance between the test fish and the bottom of the tank measured every 0.10 sec and averaged for the 10 min session in each task. Previous observations (e.g. [2]) suggested that proximity of zebra fish to the bottom of the tank may represent a measure of fear vs. habituated state. Mean distance from stimulus: The distance between the test fish and the glass wall adjacent to the stimulus tank was recorded every 0.10 sec and was averaged for the session. Note that the stimulus (the group of conspecifics and the predator model) was presented at the same side for a given experimental fish but the side randomly changed among experimental fish. Also note that in the novel open tank no stimulus is presented on either side of the tank, and the side from which distance is quantified is chosen to be the same as the one in which the group of stimulus fish or the predator model would be presented. This measure allows us to analyze social cohesion, aggression, and the effect of the predator model. Total distance moved: to quantify locomotor activity, the total distance moved was recorded in cm. Mean heading: the angle of movement relative to the vertical plane perpendicular to the video camera. The subject's location was measured every 0.10 seconds, and a vector was calculated between that location 'n', and the most recent point 'n-1'. The average angle of this vector relative to the vertical reference line was taken for the entire session and for the first and last minute. Turn angle: it was calculated as the difference between 2 consecutive heading values, taken every 0.10 seconds and averaged across the session.

Event recording. Motor/posture patterns may reflect aspects of behavior different from what the above defined activity and location parameters quantify. Such motor patterns are recorded by a human observer. The ethogram, i.e. a complete list of species-specific motor and posture patterns, is not yet established for zebra fish. Here we recorded and quantified only 6 basic motor patterns using the Observer event recording software (Noldus, Wageningen, The Netherlands) as a proof of concept analysis. These included: Swimming, a continuous locomotion faster than 1 cm/sec with the use of the pectoral and caudal fins; Thrashing, a forceful back-and-forth swimming against the glass wall; Floating, very slow or no locomotion using the caudal fin, but pectoral, dorsal, and anal fins may open and close, or beat, with a low and stable frequency (no more that 1 beat per sec); Freezing, a motionless state during which only the gills or the eyes move; Erratic movement, a fast (more than 3 cm/sec) and seemingly aimless zig-zagging with frequent changes of direction; Creeping, a slow (less than 1 cm/sec) movement during which the caudal dorsal and anal fins are motionless and only the pectoral fins beat. The duration relative to session or interval length (%) was calculated for all behavioral units.

2.5 Statistical analysis

Data analysis was conducted using SPSS (version 12.0.1 for the PC). Behavior of fish across multiple tests was analyzed with repeated measure variance analysis (ANOVA). In case of significant results, differences across test situations were further analyzed using post hoc Tukey Honestly Significant Difference (HSD) test. To investigate potential correlations among swim path parameters and motor and posture patterns, bivariate Pearson correlation coefficients were calculated and the correlation matrices were subjected to Principal Component Analysis (not shown). The Component matrices were Varimax Rotated using Kaiser

Normalization. Retention of Components was set at the minimum eigenvalue of 1.

3 Results

Zebra fish exhibited different behavioral responses to the four test situations. These differences were detected similarly by the manual recording of location and activity of the fish and by computerized video-tracking. For example, manual quantification of the location of the fish (Figure 1A) showed a significant test effect (ANOVA F(3, 351) = 17.89, p < 0.001) with fish in the novel open tank and in the aggression task spending significantly (p < 0.01) less time in the upper half of the tank than in the other test situations (Tukey HSD). The pattern of results obtained with video-tracking (Figure 1B) is highly similar to the above: distance from bottom differed across tests (ANOVA F(3, 351) = 23.07, p < 0.001), and Tukey HSD showed that fish in the open tank and aggression task were closer to the bottom than in the other two tasks (p < 0.01).

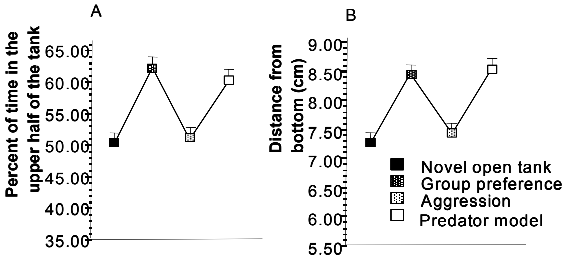


Figure 1. Zebra fish spend differing amount of time near the water surface depending on the test situation. Panel A: manual recording, panel B: video-tracking recorded data. Mean + SE are shown. Note the similar pattern of results obtained with the two recording methods

The amount of time in the half of the test tank further away from the stimulus (manual recording Figure 2A) and the distance from stimulus (video-tracking, Figure 2B) again showed virtually identical patterns across the tests (ANOVA percent time away from stimulus F(3, 351) = 199. 92, p < 0.001; distance from stimulus F(3, 351) = 568.94, p < 0.001). For both measures Tukey HSD showed that the sight of a group of conspecifics dramatically reduced the amount of time spent away or the distance from the stimulus side (p < 0.01).

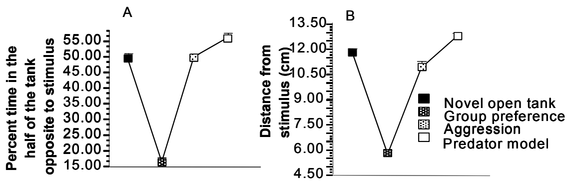


Figure 2. Zebra fish spend differing amount of time near the target stimulus. Panel A: manual recording, panel B video-tracking recorded data. Mean + SE are shown. Note the similar pattern of results obtained with the two recording methods.

Also, fish in the predator model task spent significantly more time in the opposite side of the tank and their distance from the stimulus was also significantly larger compared to the values obtained in the other tests(Tukey HSD, p < 0.05), i.e. fish avoided the predator model.

Locomotor activity (Figure 3) measured as the total number of transitions among the four quadrants of the test tank (shuttling activity), showed significant differences across tests (ANOVA F(3, 351) = 42.72, p < 0.001) with activity being highest in the novel open tank (p < 0.05 compared to all other test situations) and lowest in the group preference task (p < 0.01 compared to all other test situations, Tukey HSD). The video-tracking results are similar (ANOVA F(3, 351) = 22.23, p < 0.001) except that the total distance moved appears higher in the group preference task than seen in the manual data. This difference is due to that fish in the group preference task spent most of their time swimming in the half of the tank closer to the stimulus fish and thus performed less shuttling between the left and right side of the tank while still being quite active, which could be properly detected with video-tracking.

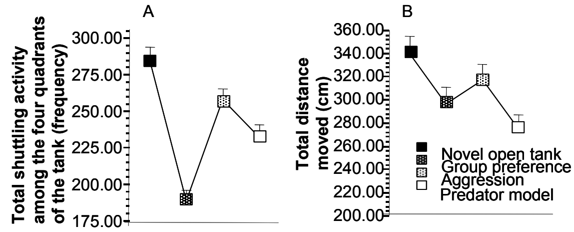


Figure 3. Locomotor activity of zebra fish as measured by total shuttling activity recorded manually (Panel A) and total distance moved recorded by video-tracking (Panel B). Mean + SE are shown. Note again that the pattern of results obtained with the two methods are comparable. Also note that the total distance moved was erroneously quantified for the aggression task by video-tracking due to technical problems (see text) and the swim paths had to be manually corrected before analysis.

Mean heading direction could only be quantified using video-tracking (Figure 4) but not by the manual technique. No significant differences were found among test situations (ANOVA F(3, 351) = 0.91, p > 0.40), i.e. all values were around 180 degrees demonstrating that, on average, fish swam in all directions in each task.

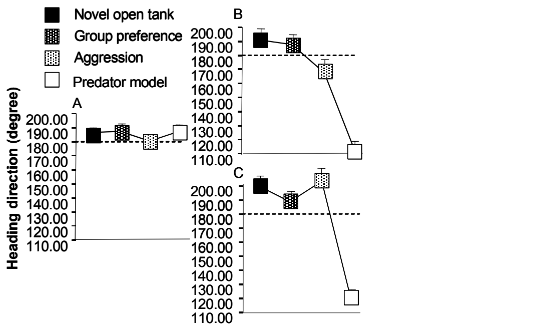


Figure 4. Heading direction quantified by video-tracking in the four test situations. Mean + SE are shown. Panel A: average heading direction (entire session). Panel B & C: heading direction during the 1st and 10th (last) minute of the session. The average heading direction (180°) is indicated by the dashed line. Note the lack of significant difference in overall heading direction and the significantly reduced values during the 1st and 10th minute of recording in the predator model test.

Figure 4B and C shows, however, that heading direction during the first and last (10th) minute of the session did differ among tests (ANOVA first min F(3, 339) = 20.24, p < 0.001; last minute F(3, 342) = 23.65, p < 0.001) likely because the predator model presented during these periods significantly (Tukey HSD, p < 0.01) reduced the heading direction values, i.e. fish moved away from the stimulus.

The mean turning angle (Figure 5) was largest in the two paradigms associated with social interaction, i.e. the group preference and the aggression task (ANOVA test effect F(3, 351) > 111.61, p < 0.001; Tukey HSD, group preference and aggression test values differ from the other two tasks p < 0.01, and no other differences are significant p > 0.05). The increased turning angle reflects the intense Thrashing (swimming against the glass, i.e. towards the group of conspecifics in the group preference task, see

Figure 7C) and the aggressive dance (not quantified) seen in the aggression task (personal observations).

In addition to the above behavioral parameters motor and posture patterns were also measured. These patterns

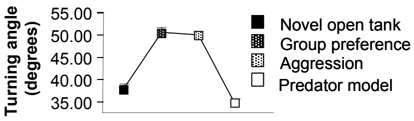
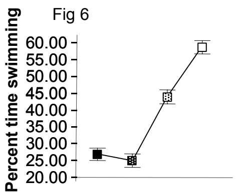


Figure 5. Turning angle quantified by video-tracking in the four test situations. Mean + SE are shown. Note the robust turn angle increase (smaller turning radius and thus increased angular change in movement) in the group preference and aggression tasks.

may reflect unique features not captured by the traditional activity parameters. Here we quantify 6 motor patterns and ask whether they are redundant as compared to video-tracking measures. Figure 6 shows "Swimming" in the 4 tests. ANOVA found a significant effect of test situation ($F(3, 351) = 137.27$, $p < 0.001$) with swimming highest ($p < 0.01$) in the predator model test, second highest in the aggression test ($p < 0.01$), and lowest ($p < 0.01$) in the group preference and novel open tank (Tukey HSD).



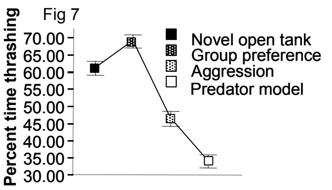


Figure 6. Percent of time zebra fish swam in the four test situations. Mean + SE are shown. Note that this behavioral measure was quantified using observation based event recording. Also note that the test specific differences show a pattern dissimilar to any activity parameters recorded by video-tracking.

Figure 7. Percent of time zebra fish performed "Thrashing" in the four test situation. Mean + SE are shown. Note that the pattern of differences among Thrashing values of the four test situations do not correspond to patterns of activity parameters recorded by video-tracking.

Thrashing also differed across tasks (Figure 7; ANOVA $F(3, 351) = 120.92$, $p < 0.001$) with a pattern different from that of Swimming. It was highest in the Group preference paradigm and lowest in the predator model test (Tukey HSD $p < 0.05$ all groups differ from each other).

Figure 8 depicts motor patterns that occurred rarely or for shorter periods. These include Erratic movement (Figure 8A), Floating (Figure 8B), Creeping (Figure 8C), and Freezing (Figure 8D). Despite the low occurrence and thus the relatively higher variability compared to the mean, these behaviors also showed test dependent significant differences (Erratic movement ANOVA $F(3, 351) = 30.24$, $p < 0.001$; Creeping ANOVA $F(3, 351) = 7.67$, $p < 0.001$; Floating ANOVA $F(3, 351) = 3.45$, $p < 0.05$; Freezing ANOVA ($F(3, 351) = 3.41$, $p < 0.05$).

Unlike in the case of video-tracking and manual recording, generally no correlation is apparent between the means of behavioral measures across the 4 tests obtained with video-tracking & event recording, i.e. the pattern of changes across tests is different between video-tracking and event recording measures. Nevertheless, inter-individual correlation may still exist between these variables. To address this question we conducted Principal Component Analyses (PCAs) to reveal correlation groups of behaviors. The results (not presented here but only at the conference due to space limitations) suggest that, in general, video-tracking and event recording measures do not correlate with each other, i.e. they are not redundant.

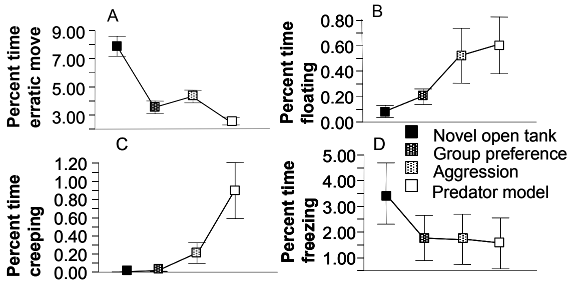


Figure 8. Motor and posture patterns Erratic movement (Panel A), Floating (Panel B), Creeping (Panel C) and Freezing (Panel D) (Mean + SE), behaviors that occurred less frequently and for a shorter duration but they also reveal significant test differences.

4 Discussion

The behavioral paradigms used in this paper were previously developed with simplicity and automation in mind ([2]). Quantification of behavior in these paradigms with video-tracking and event recording, were now conducted along with manual recording of activity and swim location. Comparison of the results obtained with these methods suggests that video-tracking can appropriately quantify activity as well as the location of zebra fish and it also allows better precision and more detailed analysis of path characteristics. Nevertheless, video-tracking may be inferior to manual, observation-based recording methods if the test set up is not appropriate. For example, in our studies, the human observer could easily recognize the experimental fish and tell it apart from its mirror image in the aggression task whereas the video-tracking system could not easily differentiate the two. Similarly, optimal illumination, lack of reflections, good water clarity are some examples of factors important for tracking that may not influence the quality of observation-based data recording.

Another finding that emerged from our studies we will discuss at the conference is that video-tracking measures generally did not correlate with event recording measures. This seemingly surprising result may be due to that video-tracking is capable of quantifying the intensity of behaviors while event recording may be less able to. For example, it is difficult to judge for an observer how fast a fish is swimming. Conversely, video-tracking, as conducted here, is less able to differentiate qualitatively different motor patterns. Thus the two methods appear complementary in our current study. However, it must be noted that video-tracking may be enabled to detect some of the motor and posture patterns of zebra fish. In the future, we will identify particular swim path patterns that correspond to motor/posture patterns and will program EthoVision to detect and quantify these behaviors, a topic that will be discussed in more detail by the participants of the Measuring Behavior 2005 conference.

References:

1. Gerlai, R. (2003). Zebra fish: An uncharted behavior genetic model. *Behavior Genetics*, **33**, 461-468.
2. Gerlai, R., Lahav, M., Guo, S., Rosenthal, A. (2000) Drinks like a fish: Zebra fish (*Danio rerio*) as a behavior genetic model to study alcohol effects. *Pharmacol Biochem Behav*, **67**, 773-782.
3. Guo, S. (2004). Linking genes to brain, behavior and neurological diseases: what can we learn from zebrafish? *Genes Brain and Behavior*, **3**, 63-74.

Behavior mediated by odors: challenges and solutions to elucidate the function, mechanisms and chemicals behind olfactory responses

T.Turlings
University of Neuchâtel, Neuchâtel, Switzerland

Two groups of organisms, mammals and insects, have been the subjects of intensive research on how odors modulate their behavior. Recent progress has been significant for both groups of organisms, but exchange of information between the two areas has been virtually non-existent. The current symposium is a first attempt to correct this by addressing the latest developments for both fields in a multi-disciplinary setting. The symposium brings together scientists that address comparable questions in their research, but they work on different organisms and study odors and their effects on behavior in totally different ways. With such communicative efforts scientists can be introduced to methodologies used in, for them, new research fields, exchange information between these fields, and thus generate new ideas on how to tackle their respective research questions. Hence, the following papers intend to exchange information on behavioral assays, neurophysiological measuring techniques, powerful statistical tools, as well as odor collection methods and it is expected that scientists in various related fields will benefit from this information and learn about novel methods that might be applicable to their own research.

From 160 to 20 to 1 odorant: Steps in a chemo-ethological strategy to identify a mammalian pheromone

B. Schaal, G. Coureaud and D. Langlois

This paper presents the strategy used to fractionate a complex odorant mixture into 20 candidate compounds, to finally single out a compound which activity explains the activity of the whole mixture. The joint chemical-behavioral approach is outlined, and followed by a description of the experiments permitting to enter the isolated key-compound into the category of biologically-active compounds termed "pheromone".

Bioassays for attractants, to what end: aiding identification, measuring adaptive value, or verifying orientation mechanisms?

R.T. Cardé

Laboratory bioassays of insect attractants serve many masters. In identification, laboratory bioassays are used to monitor fractionation procedures for active components and ultimately to verify correct identification. Some identification bioassays, such as those commonly employed in wind tunnels, attempt to mimic most aspects of field behavior, whereas others rely on selected responses as being representative of the full behavioral repertoire. Do such truncated laboratory bioassays fall short, either by failing to reveal active components or by claiming behavioral activity for odors that in the field seem inactive? A second issue, measuring the adaptive value of attraction and attractive odorants, often involves simple, two-choice bioassays. These can produce valuable insights into an organism's behavioral capabilities, but they also can present artificial choices that would rarely, if ever, be encountered in the field. Comparisons using no-choice assays and choice assays with positive controls offer complimentary approaches. Varying the duration of assays also can influence their outcome. A third area of inquiry is establishing the maneuvers and multiple sensory inputs that culminate in location of the odorant's source. The term "attractant" conveniently denotes the 'end point' of displacement, and this level of analysis may be sufficient for understanding the ecological and evolutionary value of attractants. Several very different orientation mechanisms, however, can modulate finding of odorant source in wind and still air. To date, in-flight maneuvers have been documented mainly with video in 2-D planar view; however, a 3-D perspective can be critical to understanding orientation maneuvers. These principles are illustrated with selected examples with moth and beetle attraction to pheromone and female mosquito attraction to host odorants.

Chemosensory recognition among mice

K. Yamazaki and G.K. Beauchamp

Body odor plays a prominent role in regulating social, sexual and endocrine responses of many species and specialized structures have evolved to produce and detect odorous signals. The major histocompatibility complex of genes (MHC) imparts to each mouse an individual odor, called an odor type, which reflects its MHC genotype. A prime experimental method we have used for identifying MHC odor types is a specially designed Y maze in which mice are trained, by water deprivation and reward, to distinguish odors from MHC dissimilar mice or their urines. An alternative or supplement to the Y maze is an automated olfactometer, a computer-programmed and fully automated apparatus in which a mouse or rat is trained for odor type distinctions, again by water deprivation and reward. In this presentation advantages and cautions associated with these methods are discussed.

Ensemble recordings in rat orbitofrontal cortex during olfactory discrimination learning: Encoding of reward magnitude

E. van Duuren, F.A. Nieto-Escámez, R.N.J.M.A. Joosten, R. Visser, A.B. Mulder and C.M.A. Pennartz

The orbitofrontal cortex (OBFc) is known to be involved in the encoding of the motivational significance of stimuli and in applying this information to the guidance of goal-directed behavior, as well as in the encoding of predictive information regarding upcoming reinforcers: neurons in the OBFc show differential firing activity during the anticipation of appetitive and aversive outcomes. To examine whether neurons in the OBFc are also able to discriminate between different amounts of reinforcers ensemble recordings were performed in rats engaged in associative learning during an odor discrimination 'go – no go' task. In this task animals learned to associate a particular odor stimulus with fluid reinforcement of varying magnitude or sign and adjust their behavior accordingly. A set of five different odors was used for any discrimination session: three odors associated with a particular amount of a positive reinforcement (10% sucrose solution, i.e. 0.05, 0.15 and 0.30 ml), one odor with no reward (non- reinforced condition), and one odor associated with a negative reinforcement (0.15 ml of a 0.015 M quinine solution). This task design allows the examination of the encoding of both actual and expected reward magnitude in the OBFc since reinforcements are preceded by the odors as predictive stimuli.

How to study the neuronal mechanisms underlying chemosensory coding and learning in herbivorous insects

H. Mustaparta

We are searching for the neuronal mechanisms underlying the behavior of herbivorous insects, heliothine moths, in their interactions with plants. Host location and selection is to a large extent based on olfaction and taste. Olfactory learning also plays an important role in this process. How the information is detected by the receptor neurons and processed in the neuronal network in the brain leading to behavioral responses, is studied by the use of chemical analyses, electrophysiological recordings, morphological tracing of the neuronal pathways and behavioral experiments.

Separation of naturally produced plant volatiles by two-column gas chromatography linked to electrophysiological recordings form single olfactory receptor neurons is suitable for identifying the active compounds. This is followed by chemical identification of these compounds using linked gas chromatography with mass-spectrometry and verification by retesting authentic materials on the sensory neurons. The identified compounds are considered as biologically relevant odorants for the insect species. They are then used to study how the olfactory information is processed in the brain, by physiological and morphological characterization of inter neurons, using intracellular recordings combined with fluorescent staining. This is followed by confocal laser scanning microscopy and 3-D reconstruction of the neurons and the innervated brain structures in the software AMIRA. Taste receptor neurons and their projections are studied by extracellular recordings combined with fluorescent staining to trace the taste pathway in the central nervous system. The connection between the olfactory and taste pathways being the pathways of the conditioned and unconditioned information, respectively, in appetitive learning is searched for by the use of intracellular recordings and staining. By placing the various identified neurons of each preparation in an averaged brain model the neuronal network can eventually be described, like in the honeybee brain (Krofczik et al 2004). Behavioral experiments, carried out to elucidate the importance of the chemical information and learning ability, include conditioning of the proboscis extension response, observation of walking and flight behavior as well as egg laying performance.

Above- and belowground olfactometers for high throughput bioassays

T. Turlings, A. Davison, I. Ricard, C. Tamo, S. Rasmann and M. Held

Customary procedures to determine attraction to odor sources and to identify the substances that are responsible for the attraction require tedious and time-consuming bioassays, as well as separate experiments using specialized methods for the collection, analyses, and identification of the chemicals. The inclusion of several different odor sources and treatments in such studies can make them exceedingly lengthy and impossible to accomplish within a reasonable time period. We have developed two types of six-arm olfactometers to facilitate our research on the chemical signals emitted by insect-damaged plants that serve to attract parasitic wasps and entomopathogenic nematodes, which are natural enemies of the damaging insects. One olfactometer is used to study aboveground and another to study belowground attraction. Using log-linear models for the statistical treatment of the data allows for the testing of multiple individuals at the same time and the models control for overdispersion and positional effects. In addition to allowing us to test the relative attractiveness of up to six odor sources at the same time, the equipment is also designed to simultaneously sample part of the volatiles emitted by the odor sources, which can subsequently be identified and quantified. The 6-arm olfactometers offer several advantages over conventional olfactometer designs and have proven to be highly effective in aiding us to identify key attractants involved in tritrophic interactions. The above- and belowground olfactometers can be attached to each other and we currently use this feature to investigate how aboveground herbivory affects belowground interactions and visa versa.

From 160 to 20 to 1 odorant: Steps in a chemo-ethological strategy to identify a mammalian pheromone

B. Schaal [1], G. Coureaud [1], D. Langlois [2]
[1] *Centre des Sciences du Goût, CNRS-Université de Bourgogne-Inra, Dijon, France*
[2] *Laboratoire FLAVIC, Inra, Dijon, France*

Abstract
This paper presents the strategy used to fractionate a complex odorant mixture into 20 candidate compounds, to finally single out a compound which activity explains the activity of the whole mixture. The joint chemical-behavioral approach will be outlined, and followed by a description of the experiments permitting to enter the isolated key-compound into the category of biologically-active compounds termed "pheromone".

Keywords
European rabbit (*Oryctolagus cuniculus*), Newborn, Milk, Pheromone, Gas Chromatography-Behavior Assay.

Introduction
To survive, infant mammals face the convergent obligation to obtain colostrum and milk from their mother with minimal delay after birth. Reciprocally, mammalian mothers have coevolved convergent means to favor the initial suckling success of their immature offspring. The most conserved of these means among the different lifestyles of mammalian neonates relies on odor cues. These cues are emitted in, on or near the mammillary structures, and promote general activation, guidance and incentive to orally attach to a nipple and maintain sucking in offspring. The identification of the volatile factor(s) releasing suckling behavior in newborn mammals has been the focus of continued interest since the sixties. Empirical attention has been most directed to available laboratory and stock species of carnivores, rodents and ungulates (e.g., 12, 13, 25, 26, 30, 31). These studies came up with diverse substrates eliciting approach and sucking in newborn rats and lambs, such as saliva, amniotic fluid, and secretions from peri-mammary skin glands. In the rat, one compound from maternal saliva was identified (dimethyl-disulphide) and assayed for its unconditional reactogenic potency in newborn pups (24). But, so far, no volatile compound produced in, on or around the mammae have been identified and systematically assayed with neonates.
The rabbit is another species of choice to investigate female-newborn interactions based on olfaction. Like kittens and rat pups, rabbit newborns are functionally deaf and blind, and hence exclusively governed by odor and tactile cues. But unlike cats and rats, rabbit litters are left isolated in a burrow (or a nest box) to be nursed only once per 24 h for less than 5 minutes. They are thus under strong selective pressure to quickly find nipples and efficiently suck when the rare nursing opportunities occur. To cope with this situation, they react instantly to odor cues from the visiting doe by highly competent rooting and oral grasping responses. Intensive research has developed in different groups (11, 14, 28) to find out the nature of the signal(s) that control this behavior. Some technical steps of this endeavor are summarized here, which collectively made possible the characterisation of a single compound in rabbit milk releasing rabbit newborns' behavior.

Step 1. Finding out the most productive physiological state
When approached to the abdomen of rabbits, pups do respond discriminatively to males, virgin or lactating females. These responses come out whether the contact is direct (implying stimulation by tactile, and volatile and non-volatile chemical stimuli; 15) or indirect (implying only stimulation by volatile odorants; 3), suggesting that odors are active by themselves, without additional effect of tactile cues. But lactating females are by far the most efficient individuals to release orientation behavior in pups. Moreover, rabbit pups orient preferentially to the abdomen of females in early, rather than late, lactation, and in the pre-nursing, rather than the post-nursing, phase (4). The volatile factor(s) appear thus to be tightly linked with lactation in general, and with the nursing cycle in particular, being more active and/or more concentrated on the abdominal surface right before nursing. find

Step 2. Getting the source of active cues
The question arises then whether the active cues emanate from skin glands on the does' ventral surface or from milk itself. First, it became clear that the nipples were the main source of active cues on the abdomen (6, 14, 22, 23).Second, when putting the odor of lactating does' abdomen in competition with milk odor in a simultaneous choice test, it came out that both sources are equally attractive to pups (6). Otherwise, milk presented on a glass-rod elicits responses of similar pattern than those expressed on the female's abdomen (19, 23). This confirms that milk is a major, if not the major, biological carrier of the potent odorous factor which elicits activation, orientation, searching and nipple attachment in rabbit pups. Milk being the most obvious and easy-to-get secretory output in lactating rabbits, it became the focus of all subsequent analyses. We greatly benefited from the perfection of milking devices adapted for rabbits by F. Lebas (21) and W. Schley (28).

Step 3. Separating individual-specific from species-specific odor cues
The next difficulty to resolve was the multiplicity of the pathways conveying volatile olfactants into milk. A range of individual-specific pathways (diet, stress) and species-specific pathways (genetic, immunogenetic constitution) may be involved in the nested cues collecting in milk. Rabbit pups detect species-specific odor cues as they do not respond in the typical way when put on the abdomen of anaesthetized lactating rats, cats, or hares (16, 23). Thus, female rabbits are unique in sending out cues which affect rabbit pup behavior. Further, when put on the belly of anaesthetized lactating female rabbits, pups express their typical searching-grasping responses regardless of whether the test female was the mother or unfamiliar females fed the same diet (23). Lactating females emit thus some invariant cue(s) which surpass individually variable cues in attractiveness. Further, when exposed to glass-rods dipped in samples of milk obtained from females having been fed contrasted diets during pregnancy, rabbit pups seize more often the milk sample corresponding with their prenatal experience, indicating their sensitivity to individual-specific odor cues; but pups also seize any milk rather than the control stimulus (water), indicating that they react to a factor that is present in rabbit milk in general, regardless of its content in dietary aromas (8). Finally, Müller (23) demonstrated that rabbit milk odor is unique in eliciting the responses of newborn rabbits, as bovine, ovine, porcine and feline milks are behaviorally inert. In sum, the volatile constituents of rabbit milk which are responsible for such invariable activity are targeted by our analyses.

Step 4. Separating active fractions

Milk is a multifaceted mixture from either the physical, chemical and biochemical points of view. This complexity was reduced in focusing on the volatile fraction, i.e. the headspace developing over standing milk. Pups exposed to a nitrogen flow that bubbled through fresh milk respond in the same way as when they are directly exposed to milk on a glass-rod, or to a nipple in situ. Thus, the volatile fraction of milk is sufficient to explain the release of the typical pup response. But a difficulty resides in the fact that the activity of rabbit milk is constrained by temporal factors. Müller (23) evidenced that the behavioral activity of the rabbit milk decreases considerably within 20-30 minutes after milking, a finding later corroborated (19, 27). The loss of behavioral activity of milk can be slowed down by freezing (-80°C) or cooling (on ice). Thus, all analyses were conducted on the headspace of milk within 10-15 min after milking, or on milk snap-frozen after milking.

Step 5. Capturing the odor cues from the milk headspace

Then the question arose whether the milk headspace could be captured in a vessel or on a adsorbing surface. Two strategies were taken to ascertain that point. First, vacuum distillation ensured that the volatile fraction containing the active odorant could be extracted from the milk matrix (2). We then developed a trapping method which consisted in adsorbing the volatiles on a porous polymer (Tenax). To that aim, freshly obtained milk (30 g) was poured into a flask immersed in a 30°C Bain Marie (27). A nitrogen stream was flown for 2 hours through the sample (rate: 100 ml/min), and directed to a Tenax trap (200 mg TA, 20-35 mesh). The trap was then immediately fitted to a Chrompack injector (TCT-CP4010, Middleburg, NE) placed on a HP 5890 GC (Hewlett-Packard, Palo Alto, CA). Volatiles were desorbed by heating the Tenax to 240°C during 20 min and subsequently condensed into a liquid nitrogen-cooled (-130°C) section of capillary tubing placed before a DB Wax column (30 m, 0.32 mm, 0.5 µm, J and W Scientific, Folsom, CA). This section of tubing was then rapidly heated at 250°C, and the GC run (carrier gas: H2, velocity: 50 cm/s) in temperature-programmed mode (40 to 240°C at 5°C/min). Despite of its reduction to the volatile fraction, the GC profile of the headspace of rabbit milk remains exquisitely complex, as more than 160 peaks can be individualized. A way had thus to be found to further reduce the complexity of the milk mixture in pinpointing the peaks which correspond to the highest behavioral impact.

Step 6. Locating active compounds in the headspace of milk

A test method associating GC and olfaction (GCO) was developed for rabbit pups (27), designed after current methods used in human food flavour chemistry. At the end of the capillary column, the stream of volatiles was split into two equal flows directed by silicium capillaries to the flame ionisation detector of the GC and to a sniff-port made of an insulated copper tubing fixed on the GC (length: 20 cm, i.d.: 3 mm; heated at 240°C to avoid condensation), with a glass funnel inserted at its end. The silicium capillary (0.32 mm i.d.) affixed on the GC column ended at the entry of the glass funnel. Air humidified on charcoal (flow rate: 20ml/min) arrived into the glass funnel downstream of the end of the capillary carrying the volatiles. This approach allowed to direct a flow carrying the active compounds simultaneously to the nose of rabbit pups held to the sniffing port of the GC and to the flame ionisation detector. In total, 875 pups were assayed with the GCO (25 runs) of the headspace of fresh milk from 25 females. Each GCO run lasted 35 minutes, one pup being presented per minute at the GC sniff-port. Thus, 35 hungry pups were assayed (5 pups from 7 litters; aged 1-4 days, 75% being tested once, 25% twice; cf. Schaal et al., 2003). To be presented at the sniff-port of the GCO device (muzzle at a distance of 5 mm), rabbit pups were hand-held in a standardized way. Pups were alternated every minute at the sniff-port with a minimal interruption period (<2 s). A ±30-second lag was set at the beginning of each next GCO run so that this interruption was not coincident with the same time in all GCO runs. Three experimenters ran the procedure: A piloted the GC and recorded pup responses on both computer and GC paper-tracing; B and C (blind to the progress of the GC) exposed the pups to the GC effluent and gave a duplicate report of the pups' responses to A.

Two behavioral criteria were considered as responses: high frequency cephalic searching movements directed to the sniffing-port, and attempts to orally seize it (27). The cumulative frequency of both responses allowed to determine several active regions on the GC profiles. After 25 such GCO tests, we obtained a pattern of 21 reactogenic regions of the chromatogram, which could be directly linked with more acute pup responses. Most active compounds corresponding to the peaks were then identified by GC-mass spectrometry (GC-MS) from the headspace volatiles of fresh rabbit milk. A Hewlett-Packard 5890 GC (TCT-CP4010 injector) with the same capillary column under the same conditions than those used for the GCO assays, and coupled with a mass spectrometer (R10-10C, Nermag, Rueil-Malmaison, F) were used for identification (ionisation energy: 70 eV; scan frequency 25-300 a.m.u. in 800 ms). The mass spectra of the candidate compounds were identified by comparison with those of the Wiley library (27).

Step 7. Behavioral assay of candidate compounds

The 21 candidate odorants were purchased in reagent grade quality (Aldrich) to be assayed in a test consisting to serially present pups with glass-rods carrying 30 µl of dilute aqueous solutions of these compounds (concentration: 10-6 g/ml; pups aged 2-4 days). The a priori choice of this concentration relied on routine procedure in the screening of food aromas in humans. The variables recorded were the frequency of pups responding by head-searching movements and/or oral grasping of the glass-rod. Among these 21 candidates, one revealed to be by far the most effective, as more than 90% of pups responded to it by typical searching-grasping actions (27).

Step 8. Chemical and Etho-chemical verifications

To further ascertain that 2MB2 is a key-compound in the headspace of rabbit milk further verifications were needed (27). 1) GC-detectable impurities co-occurring with commercial 2MB2 were screened for activity with the glass-rod assay, without record of notable effectiveness. 2) When pure 2MB2 was GCO-assayed using a polar capillary column (DB wax, 1 µl 2MB2 injected at concentration: 2.5x10-5 g/ml in dichloromethane, same GC conditions as above), the timing of the pup searching-grasping responses corresponded with the retention time of 2MB2. 3) An apolar GC column (DB5 column, 30 m, 0.32 mm, 1 µm, J-W Scientific, Folsom, CA) was used to check for co-eluting compounds hidden under the 2MB2 peak. The pup response time matched with the retention time of 2MB2, designating 2MB2 as a compound that releases the typical responses in pups. 4) The observation (cf. step 4) that the behavioral activity of milk drops drastically after milking was suitable to assess the contribution of 2MB2 to this activity in dosing its headspace concentration in parallel with behavioral testing of milk activity after ejection. The concentration of 2MB2 strongly correlated both with the standing duration of milk and with its behavioral effectiveness in pups (27), indicating that the 2MB2 content parallels with the releasing potency of milk. In addition, when fresh milk was rendered inactive by bubbling nitrogen through it (during 2 h, flow rate: 100 ml/min), its behavioral activity was reinstated by adding 2MB2 (concentration: 10-6 g/ml). 2MB2 is thus a key-component among the volatiles carried in rabbit milk.

Step 9. From a common "odor cue" to a "pheromone"

At that stage, 2MB2 cannot be considered a pheromone. It is only a candidate substance for pheromonal mediation. It was thus submitted to a series of behavioral tests which aimed to assess the adequacy of entering it into that class of biologically-active substances. Methods are governed by concepts, and we decided not just to follow the initial, very loose Karlson-Lüscher (18) conceptualisation of pheromones, but relied on subsequent fine-tuning of the concept based on several restrictive operational criteria (1, 17). These criteria were that the candidate compound(s) is (are) 1) chemically simple (one or a very small set of compounds), 2) release(s) unambiguous, morphologically invariant, and functionally obvious behavioral responses in a receiver 3) in a highly selective and 4) species-specific manner; 5) the coupling between the stimulus and the response should not (or minimally) depend on previous sensory exposure (learning). Finally, we added a subsidiary point (criterion 4bis) concerning the species-specificity of emission of the odorant.

Regarding criterion 1, it is clear that 2MB2 is the chemically "simplest" possible signal, being composed of a single molecular compound.

Criterion 2 has been addressed in comparing the form and frequency of responses triggered by pure 2MB2 and milk. The macroscopic structure of pup responses could not be differentiated between both stimuli, indicating that a single key-compound from milk can elicit the same immediate response than milk itself. Further, 2MB2 was highly efficient to trigger the typical responses regardless of the mode of presentation: pups orient, search, grasp and attempt to suck when exposed to 2MB2 either on a glass-rod, at the outlet of a GC, in the nest or in a 2-choice arena) (9, 27).

Criterion 3 was ascertained by comparing pup responses elicited by 40 reference odorants either chosen arbitrarily (artificial odorants), present in rabbit milk or skin gland secretions, or finally known to be active in rat newborns. These reference odorants were ineffective, or clearly less effective than 2MB2, to release the typical responses in pups at any concentration (9, 10). Thus, the behavioral effect of 2MB2 does not arise from novelty or non-specific arousal effects which any odorant would generate.

Criterion 4 addressed species-specificity of 2MB2 at the level of reception (27). The generality of 2MB2 behavioral activity was positively demonstrated in the pups from 5 different colonies raising the same breed (New Zealand x California) but on different diets. That pup responsiveness to 2MB2 was independent of maternal dietary influences was further confirmed by direct comparison of 2MB2 activity on pups born to 2 groups of pregnant/lactating females of a same rabittry fed distinct diets. In addition, pups of 8 domestic breeds and of wild-type rabbits (20) demonstrated very high reactivity to 2MB2, attesting its releasing potency for the Oryctolagus genus in general. Finally, the presentation of 2MB2 to newborn rats, mice, cats, and even brown hares (Lepus europaeus), came out with negative results, indicating its species-specific activity.

Species-specificity of 2MB2 emission (criterion 4bis) in milk was ascertained in different ways (27): 1) GC-MS analyses indicated that 2MB2 is invariably present in rabbit milk, regardless of breed and local diet. 2) To assess the presence of 2MB2 in mammalian milks, rabbit pups were presented with colostra/milks from rats, ewes, cows, mares or women. In no instance did these milks release any significant response, suggesting that 2MB2 was either absent, present in subliminal concentrations or masked by other key compounds. At least in bovine milk, the absence of 2MB2 was ascertained by GC-MS (27).

Finally, criterion 5 assessed that pup responses to 2MB2 develop in independence of previous exposure to 2MB2 (27). In theory, neonatal responsiveness to a given odorant can be ascribed to prenatal acquisition, to facilitated learning during the natal process, or to rapid learning immediately after birth. To rule out the intervention of postnatal or natal processes in the development of 2MB2 activity, pup responses were compared after i) normal birth and exposure to the doe (prenatal, natal or postnatal experience), ii) normal birth with separation from the doe (prenatal or natal experience), and iii) Caesarean delivery and postnatal isolation (prenatal experience). Pups from these 3 groups responded at the same high rate to 2MB2, indicating that natal and/or postnatal processes do not differentially affect 2MB2 activity. To further discard the possibility that 2MB2 activity depends on fetal experience, pups were tested with glass-rods carrying either amniotic fluid (from mid- and late pregnancies) or blood from pregnant does. The outcome was negative, and additionally reinforced by the fact that 2MB2 could not be detected by GC-MS in these substrates. Thus, the development of 2MB2 activity does apparently not derive from prenatal induction (as was noted for various food aromas; e.g., 8).

The above results confirm that 2MB2 from rabbit milk is eligible as a pheromone, both in the sense of the initial definition of Karlson & Lüscher (18) derived from insect evidence, and in the sense of the later definitions proposed by Beauchamp et al. (1) and Johnston (17) to revise the concept as it applies to the complexity of mammalian behavior. As 2MB2 appeared to be produced in the mammary tract, and emitted in milk it was named "Mammary Pheromone" (MP; 27).

Step 10. From pheromone identification to testing pheromone function(s)

Some authors have insisted on the fact that the biological activity of pheromones should be viewed in the background of their communicative/adaptive meaning. Accordingly, to be qualifiedly named pheromone any given compound should not only fulfil a set of chemical or biological features such as those reported above, but it should also be demonstrated that it functions as a signal in communicative exchanges which promote mutually beneficial behavioral or endocrine consequences in the interacting conspecifics. So far, the functional involvement of the MP in the beneficial reciprocity of the rabbit doe and her litter can only be stated at a general level, at which it is straightforward. On the offspring side, the MP may have a key-role in the immediate causation of a vigorous arousal response when the doe enters the nest, then in the localisatory performance of a nipple, and finally in sustaining searching and sucking. Through its facilitation of milk intake, the MP obviously promotes survival and normal thriving until the transition to solid food. On the females' side, the MP is one signal which may release abdominal poking and searching in pups, and then boost sucking and associated tactile stimulations known to be key-mechanism to lactational physiology. Ultimately, the MP may thus be related to the individual fitness of pups, and hence to the inclusive fitness of females.

It is agreed that such a beneficial reciprocity remains close to conjecture at this general level, but it opens a ground for future experiments regarding the involvement of a pheromone in the behavioral development of newborns and juveniles, and in the regulation of maternal investment. Future efforts should be devoted to investigate the extent to which the MP and other odor cues are causal to successful performance in functional domains related to feeding and socialization. For example, it could be shown that individual pups who do not react to the MP in an oral-seizing test on postnatal day 1 had lower chances to survive during the following 4 weeks (5), indicating that initial responsiveness to the MP is linked with long term viability. The mediating factors of such a phenomenon, possibly related to perceptual or behavioral deficits in individual pups, are under scrutiny.

Conclusions

The 10 steps to the identification of the rabbit mammary pheromone took about 25 years since the first reports by P. Schley (1976) and K. Müller (1978) that rabbit pups are particularly reactive to odor cues from their lactating does' belly and milk. The testing techniques devised by these pioneers, and further important refinements brought about by R. Hudson's, F. Lebas's, and our own, groups, greatly facilitated the achievement of steps 6-9 during an additional slice of 7 years. Thus, even with an easily accessible and manipulable species expressing easily quantifiable responses, and with an easily-collectable odor substrate, identifying a pheromone and testing its function(s) is a protracted venture in a mammalian species. Nevertheless, the obvious and (almost) imperturbable motor responses of immature rabbit pups were critical in sniffing out the active peaks during the GC runs. A similar technique has since been successfully applied to isolate active compounds from urine in mice by recording electrophysiological responses of individuals exposed to the outlet of a GC, and then assaying behaviorally the compounds triggering neural responses (32). Reliable behavioral work is also necessary for another purpose: It is not the most easy task to persuade chemists to engage into long and uncertain chemo-ethological analyses, and in our case only the display of the compellingly strong responses of rabbit pups to the milk odor won their favor. Thus, the design of rigorous behavioral analyses and the set up of repeatable behavioral assays is a prerequisite for starting to interact with chemists.

Finally, the present research makes clear that precise, empirically-testable criteria should be re-defined to replenish a restrictive meaning for the concept of pheromone (1, 17, 33). A pheromone is far from being demonstrated when a functional response to an undefined olfactory agent is depicted and wishfully termed pheromone. Although definitions of the concept are multiple, they all share the recognition that some materially-based substance has a 'special' biological activity. Thus, to have an existence, a pheromone needs to be chemically substantiated with a kit of functional measures, and the 'special' properties of the purported substance need to be scrutinized in a systematic way.

References

1. Beauchamp, G.K., Doty, R.L., Moulton, D.G., Mugford, R.A. (1976). The pheromone concept in mammals: A critique. In Doty, RL. (Ed.), *Mammalian Olfaction, Reproductive Processes, and Behavior*, p. 143-160. New York, Academic.
2. Coureaud G. (2001). *Régulation Olfactive de la Prise Lactée Chez Le Lapereau: Caractérisation Ethologique et Chimique d'un Signal Phéromonal*. Unpubl. PhD thesis, Univ. Paris 13.
3. Coureaud, G. Schaal B. (2000). Attraction of newborn rabbits to abdominal odors of adult conspecifics differing in sex and physiological state. *Dev. Psychobiol.*, **36**, 271-281.
4. Coureaud, G., Schaal, B., Langlois, D., (2000) The interaction between pheromone-elicited and odor-elicited behaviour in the newborn rabbit. *Chem. Senses*, **26**, 781.
5. Coureaud, G., Fortun-Lamothe, L., Langlois, G; Schaal, B., (2000). *Odeurs lactées, survie et croissance chez le lapereau.* 1er Colloque d'Ethologie Appliquée (Villetaneuse, France).
6. Coureaud, G. Schaal B., Langlois D., Perrier G. (2001). Orientation response of newborn rabbits to odors emitted by lactating females: Relative effectiveness of surface and milk cues. *Anim. Behav.*, **61**, 153-162.
7. Coureaud, G., Langlois, D., Schaal, B. (2002). Rabbit Pups and Their Mother's Abdominal and Milk odors: Age-Specific Changes in Emission and Reception. *Dev. Psychobiol*., **41**, 74
8. Coureaud G., Schaal B., Hudson R., Orgeur P., Coudert P. (2002). Transnatal olfactory continuity in the rabbit. *Dev. Psychobiol.*, **40**, 372-390.
9. Coureaud G., Langlois D., Perrier G., Schaal B. (2003). A single key-odorant accounts for the pheromonal effect of rabbit milk. *Chemoecol.*, **13**, 187-192.
10. Coureaud G., Langlois D., Sicard G., Schaal B. (2004). Newborn rabbit reactivity to the mammary pheromone: Concentration-response relationship. *Chem. Senses*, **29**, 341-350.
11. Drewett RF., Kendrick KM., Sanders DJ., Trew AM. (1982). A quantitative analysis of the feeding behavior of suckling rabbits. *Dev. Psychobiol.*, **15**, 25-32.
12. Ewer R.F. (1961). Further observations on suckling behaviour in kittens, together with some general considerations of the interrelations of innate and acquired responses, *Behaviour* **17**, 247-260.
13. Hofer, M.A., Shair, H., Singh, P.J., (1976). Evidence that maternal ventral skin substance promote suckling in infant rats. *Physiol. Beha.,* **17**, 131-136.
14. Hudson R. (1985). Nipple location by newborn rabbits: evidence for pheromonal guidance. *Behaviour*, **85**, 260-275.
15. Hudson R., Distel H. (1990). Sensitivity of female rabbits to changes in photoperiod as measured by pheromone emission. *J. Comp. Physiol. A*, **167**, 225-230.
16. Hudson, R (1985), Do newborn rabbits learn the odor stimuli releasing nipple-search behavior? *Dev. Psychobiol.*, **18**: 575-585.
17. Johnston RE. (2000). Chemical communication and pheromones. In TE Finger, WL Silver & D Restrepo (Eds.), *The neurobiology of taste and smell, 2nd Edition*, Wiley-Liss, New York, pp. 101-127.
18. Karlson P., Lüscher M. (1959). "Pheromones": a new term for a class of biologically active substances. *Nature*, **183**, 55-56.
19. Keil, W., von Stralendorff, F., Hudson, R. (1990). A behavioral bioassay for analysis of rabbit nipple-search pheromone. *Physiol. Behav.*, **47**, 525-529.
20. Kurz C., Rödel H., Coureaud G., Schaal B., von Holst D. (2005). *Wild rabbit pup responsiveness to the pure rabbit mammary pheromone*. 98th Annual Meeting of the German Zoological Society, Bayreuth, Germany.
21. Lebas, F. (1970). Description d'un e machine à traire les lapines. *Ann. Zootechnie*, **19**, 223-228.
22. Moncomble, A.S., Coureaud G., Quennedey B., Langlois D., Perrier G., Brossut R., Schaal B. (2005). The mammary pheromone of the rabbit: where does it come from ? *Anim. Behav.*, **69**, 29-38.
23. Müller K. (1978). *Zum Saugverhalten von Kaninchen unter besondere Berücksichtigung des Geruchsvermögen*. Unpublished PhD Thesis, University of Giessen, Germany.
24. Pedersen PA., Blass EM. (1981). Olfactory control over suckling in albino rats. In RN. Aslin, JR. Alberts, MR. Pedersen (Eds.), *The development of perception: Psychobiological perspectives* (pp. 359-381). New York, Academic.
25. Rosenblatt JS. (1983). Olfaction mediates developmental transitions in the altricial newborn of selected species of mammals. *Dev. Psychobiol.*, **16**, 347-375.
26. Schaal B., Orgeur P., Arnould C. (1995). Olfactory preferences in newborn lambs: possible influence of prenatal experience. *Behaviour*, **132**, 351-365.
27. Schaal B., Coureaud G., Langlois D., Giniès C., Sémon E., Perrier G. (2003). Chemical and behavioural characterization of the rabbit mammary pheromone. *Nature*, **424**, 68-72.
28. Schley P. (1976). *Untersuchungen zur künstlichen Aufzucht von Hauskaninchen*. Unpublished Habilitationschrift, University of Giessen, Germany.
29. Schley P. (1979). Olfaction and suckling behavior in young rabbits. *Proceedings of the 1st World Lagomorph Conference* (pp 291-294), Guelph, Canada.
30. Teicher M.H., Blass E.M. (1977). First suckling response in the newborn albino rat: the roles of olfaction and amniotic fluid. *Science*, **198**, 635-636.
31. Vince, M.A., Ward, T.M., 1984. The responsiveness of newly-born clun-forest lambs to odor sources in the lamb. *Behaviour*, **89**, 117-127.
32. Lin da Y., Zhang S., Block E., Katz L.C. (2005). Encoding social signals in the mouse olfactory bulb. *Nature*, **434**, 470-477
33. Doty, R.L. (2003). Mammalian pheromones: Fact or fantasy? in RL. Doty (Ed.), *Handbook of Olfaction and Gustation*, pp. 345-383, New York, Dekker.

Bioassays for insect attractants, to what end: aiding chemical identification, measuring adaptive value, or verifying orientation mechanisms?

R.T. Cardé
Department of Entomology, University of California, Riverside, USA

Abstract

Laboratory and field bioassays of odourants that induce attraction serve many masters. In identification, bioassays often are used to monitor fractionation procedures and then verify correct identification. Some bioassays, such as those commonly employed in wind tunnels, attempt to mimic most aspects of natural orientation behaviour in the field, whereas others rely on selected ('key') responses as representative of the full behavioural repertoire. Do such truncated bioassays fail in revealing active components? The second issue, measuring the adaptive value of attraction and attractive odorants, often involve simple, two-choice bioassays. These can produce valuable insights into an organism's behavioural capabilities, but they also can present artefactual choices that would be rarely, if ever, encountered in the field. Comparisons of no-choice assays and choice assays with positive controls offer complimentary approaches. The duration of assays also can influence the outcome of comparisons. The third area of inquiry is establishing the manoeuvres and multiple sensory inputs that induce orientation manoeuvres that culminate in location of the odorant's source. The term "attractant" conveniently denotes the "end point" of such displacement, but several very different orientation mechanisms can modulate finding of an odorant source in wind. These principles will be illustrated with selected cases with moth and beetle attraction to pheromone and female mosquito attraction to host-emitted odorants.

Keywords

Attractants, bioassays, orientation, olfactometer, wind tunnel, anemotaxis,

1 Introduction

As J.S. Kennedy [1] pointed out in 1978, a chemical 'attractant' is a portmanteau term that signifies an end point of several kinds of orientation manoeuvres. Attraction is usually measured by the accumulation of responders at or near the chemical source, whether in a laboratory or field bioassay. Kennedy [2] had earlier cautioned that 'chemical attractant' is a teleological concept, because to move towards a chemical source, an organism needs to employ positional information (e.g., where it is relative to the stimulus or where it is relative to its last sensing of the stimulus) to manoeuvre, and therefore its trajectory is not dictated simply by the distribution of the odorant in space. Attraction may also include a strong component of the equally portmanteau term 'arrestment', in which following arrival at the chemical source, movement is greatly reduced or turning is increased. Indeed, arrival at the source could be an entirely random event—in the sense that the organism blundered to the vicinity of a source—in which case aggregation would be mistaken for attraction.

An attractant chemical therefore **induces** these reactions, but the chemical itself is not 'attractive'. These reactions include directed movements in which the long axis of the responding organism is generally aligned with the direction towards the stimulus (collectively called taxes) and undirected reactions (called kineses) in which the rate of turning, speed of movement, or both, are altered by stimulus intensity [3].

'Attractant' remains a widely accepted term and it will be used here, understanding the caveats posed by Kennedy. Behavioural ecologists find measuring such 'end points' of behaviour nonetheless to be useful benchmarks. Attraction bioassays, for example, can determine how varying a natural blend of chemicals alters response.

Here I will review selected examples of bioassays that quantify attraction, asking: when they are used for identification, do they adequately mimic natural field behaviours? Can they quantify differences in response that mimic the consequences of choices in the field? And are they instructive in establishing which kinds of manoeuvres result in attraction?

2 Assays Aiding Identification

2.1 What behaviour to monitor?

For the most part I will consider assays that are intended to simulate directed upwind movement (called 'anemotaxis') towards the odorant's source by flying insects. A walking organism can rather simply gauge wind direction with mechanoreceptors, in many species via antennal deflection. An in-flight organism, however, cannot determine wind direction by this means, because it is iterant within the medium. To gauge upwind instead it appraises its trajectory relative to the flow of its visual surround. If for example the flow of the visual field seen below is front to rear, then the organism is aligned with the flow. If there is some transverse image flow, then it is headed off the windline. This orientation mechanism is termed 'optomotor anemotaxis'.

Laboratory assays that attempt to mimic conditions in the field for upwind flight vary considerably. An 'ideal' assay would enable the full repertoire of behaviours to be expressed and monitored, including having the test insect encounter the odorant plume while it is in 'foraging' flight. In practice, however, in assays insects are usually released singly or sometime in groups directly downwind of the odour to be tested and their progress towards the source monitored. The test milieu ranges from wind tunnels [4,5], in which flight can extend over several meters, to Y-tube olfactometers with wind flow, in which in-flight insects often contact the walls of the olfactometer [6], to choice chambers which push odor plumes into a still air flight chamber [7]. Nowadays, these assay systems are used mainly to verify the behavioural activity of putative attractant chemicals, not to monitor their activity during the process of isolation and identification. Such assays require an extensive investment of time, given the need for replication. Analysis of the actual movement patterns from video records would be prohibitively time consuming and likely is unneeded.

There has been an increasing tendency to bypass behavioural bioassays in the fractionation and chemical identification process. One approach involves 'simply' identifying all of the major constituents in an airborne collection or extract and then testing these in bioassays in varying combinations. Although this method often works, particularly when the extract or collection is comprised of relatively few constituents as can be the case with pheromones, in some systems this approach yields a daunting number of chemicals, and therefore combinations, concentrations and ratios to evaluate (e.g., >300 chemicals in human skin head space) [8]. Consequently this approach has not yet proven valuable in the identification of compounds involved in host finding of mosquitoes.

A second method uses the electroantennograms or 'EAG' linked directly to gas chromatography. This has proven to be a rapid and diagnostic method to bypass behavioural assays to monitor gas chromatographic separations for behaviourally active compounds. It has been particularly successful in the rapid characterization of female pheromones attractive to male moths [9-11].

We may now ask how important is it to use elaborate bioassays such a wind tunnel? The evidence, as we might expect, is equivocal. Wood and colleagues [12,13] successfully used a simple walking, multiple-port choice assay as a surrogate for flight to trees in their quest to unravel the attractive pheromones used by *Ips* bark beetles in location of conifer trees for pioneer colonization, mate finding, and eventually, mass attack. This suggests that mimicking the natural behavioural milieu was unneeded. Other examples include substituting wing fanning as a readily quantifiable behaviour and suitable surrogate for upwind flight in genetic assays [14].

In contrast, wind tunnel assays with host-seeking mosquitoes have uncovered important odorant-mediated reactions. By comparing the in flight reactions of host-seeking female *Aedes aegypti* females, Dekker et al. [15] found that a single, very brief contact with a filament of CO_2 immediately lowered their sensitivity to human skin odours presented an odour cloud. Healy et al. [16] found a series of oxocarboxylic acids from human sweat evoked landing in female *Anopheles gambiae*. Landing, of course, can be considered a final component of attraction. It remains unresolved in the case of *An. gambiae* whether these compounds also mediate upwind flight. These two examples suggest that we may fail to define the suite of odorants mediating the full behavioural repertoire unless we take account of the natural sequence of behaviours involved in orientation.

2.2 Effects of odour plume structure

The distribution of odour with a plume from a small source dispersed in wind is patchy because of the action of turbulent forces [17]. Does the temporal pattern of encounter with odourant alter response? Upwind flight in several moth species seems to require the successive encounter of filaments of pheromone [18,19], and so this ought to be taken into account in the assay system. In fact, most experiments do not account for this possibility, and there have been few attempts to characterise or to optimise the odour plume's structure (*cf.* Justus et al. [20]). In the case of *Ae. aegypti*, [6, 15] plume structure affects upwind movement. An intermittent CO_2 plume evokes upwind displacement, whereas a relatively homogeneous CO_2 cloud does not. Conversely, a relatively homogeneous human skin-odour cloud improving orientation over an intermittent skin-odour cloud.

3 Measuring adaptive value

Assays for attraction are often comparative: which blend of chemicals yields the highest level of attraction? Such comparisons are obviously useful and, when conducted in the field, it is often feasible to compare many candidate blends or doses. In the laboratory, however, the question is often posed as a choice bioassay: which of two blends is 'best' or, oftentimes, is a test stimulus better than the experimental control of no odor? There are two issues to examine. First, choice assays can accentuate the apparent behavioural significance of the favoured stimulus [21]. When the same stimuli are compared in a no-choice situation, both could turn out to be equally attractive. Thus it is clear from this assay that the insect is behaviourally competent to distinguish between the two stimuli, but in the field it is most improbable that it will encounter a mixed plume of odours from two such sources at the same concentrations which upwind fortuitously split into two plumes, permitting this choice. The second issue is one of whether in a choice assay the control should offer a 'positive', that is to say, at least weakly attractive stimulus.

A final question to pose is the duration of the assay. Attraction ought to occur fairly rapidly—certainly in the case of insect responding to a sex pheromone, it is presumed that there is considerable selective pressure to react immediately [22]. However some assays are conducted over long intervals, which Kennedy [23] pointed out, can alter their outcome. Dekker et al. [7], for example, evaluated the attractancy of various odours to host-seeking mosquitoes with two choice ports delivering test odourants into a large still-air chamber. Assays were run for 15 minutes. If the port entries were tallied after 3 minutes, one treatment was clearly attractive; however, by the end of 15 minutes, several additional treatments were equally attractive. It is tempting to conclude that the early responders represent the most field-relevant reaction to putative host odours. However, there may be less selective pressure to locate a sedentary host rapidly.

4 Understanding orientation manoeuvres

4.1 Analysis of flight paths

Understanding the sensory inputs and control manoeuvres that flying insects use to orient along a plume of odor has remained a challenging endeavour. Kennedy pioneered in 1940 [24] the use of wind tunnels to establish the principle of optomotor reaction. The advent of video recording permitted a rigorous analysis of the moment-to-moment reactions of insects [25,26], but it has been impossible to establish just when they are within an odour's plume envelope and, even when they are clearly within its boundary, just what kind of an intermittent signal they are encountering. Moreover, until recently nearly all of the recordings of flight paths were conducted in horizontal planar view. Most studies have been conducted with moths and it has been assumed that their height of flight above ground is held fairly constant, while they range mostly laterally, often assuming a zigzag path. In this 2-D world, the ground speed can be measured directly, speed of air moment through the tunnel is known, and its airspeed then can be calculated

In turn these permit us to calculate the apparent image flow of ground patterns directly below the insect, decomposing these into longitudinal and transverse image flow. Do we need to know more about where the flying insect is peering?—directly below as in these calculations, or, perhaps, ahead? When an insect turns with a banking manoeuvre, does it move its head to maintain the same orientation of its visual field with respect to ground as it would by simply yawing?

However, most pressing to resolve, is what do we miss in understanding orientation manoeuvres by relying on a 2-D perspective? In the case of mosquito flight along odour plumes, movement occurs in all 3-planes, as seen in flight tracks of *Culex* mosquitoes flying along a plume of carbon dioxide [27] (Figure 1).

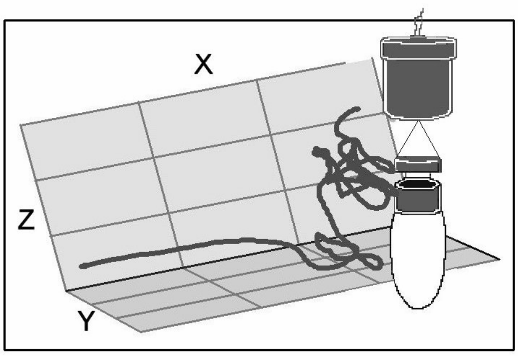


Figure 1. Three-dimensional flight path of a host-seeking female Culex tarsalis *flying along a plume of carbon dioxide from a CDC-style trap in a field wind tunnel. Upwind is to the right. Details in Cooperband [27].*

4.2 Importance of other sensory modalities in assays of attraction

Chemical ecologists are prone to assume that, given the complete characterization of a pheromone or host odor, attraction to a synthetic copy requires little else beyond its proper formulation. Indeed, this supposition is usually correct. But important exceptions abound.

In the case of the oxocarboxylic acids mediating landing in *An. gambiae*, mentioned earlier, these must be presented within ±2°C of human skin temperature to be effective [16]. In tsetse flies (*Glossinia*), the attractive component of ox breath, octenol, must be coupled with a dark visual silhouette, if it is to draw flies close to the odour's source [28]. The female pheromone of the oriental fruit moth, *Grapholita molesta*, does not evoke the male's hairpencil courtship display, unless the female's chemical signal is accompanied by a suitable visual representation of the female [29]. In each of these examples, a failure to incorporate the appropriate non-chemical sensory inputs into the assay would have resulted in either a considerable lowering or elimination of response.

5 Conclusion

The development of diagnostic behavioural assays for semiochemicals that induce attraction in flying insects has been largely incremental and very much dependent on devising unique methods to answer specific questions. The single most important bioassay technique for understanding orientation manoeuvres has been the wind tunnel coupled with a video analysis of movement, first used together by Kennedy and Marsh [25]. However, this approach has been mainly used with moths. The adoption of wind tunnels by those interested in characterizing pheromone blends [5] ushered in further refinements in analysis which are mainly considered whether or not the pheromones source was contacted. The understanding of the spatial/temporal structure is the newest frontier in devising assays that mimic the natural milieu.

References

1. Kennedy, J.S. (1978) The concepts of olfacotory 'arrestment' and 'attraction'. *Physiological Entomology* **3**, 91-98.
2. Kennedy, J.S. (1972) The emergence of behaviour. *Journal of the Australian Entomological Society* **11**, 168-176
3. Kennedy, J.S. (1986) Some current issues in orientation to odour sources. pp 11-25 In *Mechanisms in Insect Olfaction*, T.L. Payne, M.C. Birch and C.E.J. Kennedy, eds. Clarendon Press, Oxford.
4. Baker, T.C. and Linn, C.E. (1984) Wind tunnels in pheromone research. pp. 75-110. In *Techniques in Pheromone Research*, H.E. Hummel and T.A. Miller, eds. Springer-Verlag, New York.
5. Baker, T.C. and Cardé, R.T. (1984). Techniques for behavioral bioassays. pp. 45-73. In *Techniques in Pheromone Research*, H.E. Hummel and T.A. Miller, eds. Springer-Verlag, New York.
6. Geier, M. (1999) *Journal of Experimental Biology*
7. Dekker, T., Takken, W. and Cardé, R.T. (2001) Structure of host-odour plumes influences catch of *Anopheles gambiae s.s*. and *Aedes aegypti* in a dual-choice olfactometer. *Physiological Entomology* **26**, 124-134.
8. Bernier, U.R., Kline, D.L., Barnard, D.R., Schreck, C.E. and Yost, R.A. (2001) Analysis of human skin emanations by gas chromatography / mass spectrometry. 2. Identification of volatile compounds that are candidate attractants for the yellow fever mosquito (*Aedes aegypti*). *Analytical Chemistry* **72**, 747-756.
9. Roelofs, W.L. Electroantennogram assays: rapid and convenient screening procedures. pp. 131-159. In *Techniques in Pheromone Research*, H.E. Hummel and T.A. Miller, eds. Springer-Verlag, New York.
10. Strubble, D.L. and Arn, H. Combined gas chromatography and electroantennograms recording of insect olfactory responses. pp. 161-110-189. In *Techniques in Pheromone Research*, H.E. Hummel and T.A. Miller, eds. Springer-Verlag, New York.
11. Bjostad, L.B. (1998) Electrophysiological methods. pp. 339-375 In *Methods in Chemical Ecology. Chemical Methods*, J.G. Millar and K.F. Haynes, eds. Chapman and hall, New York.
12. Wood, D.L. and Bushing, R.W. (1963). The olfactory response of *Ips confusus* (LeC.) (Coleoptera: Scolytidae) to the secondary attraction in the laboratory. *The Canadian Entomologist* **95**, 1066-1078.

13. Payne, T.L., Hart, E.R., Edson, L.J., McCarty, F.A., Billlings, P.M., and Coster, J.E. (1976) Olfactometer for assay of behavioral chemicals for the southern pine beetle, *Dendroctonus frontalis* (Coleoptera: Scolytidae) *Journal of Chemical Ecology* **2**, 411-419.
14. Collins, R.D. and Cardé, R.T. (1989) Wing fanning as a measure of pheromone response in the male pink bollworm, *Pectinophora gossypiella* (Lepidoptera: Gelechiidae). *Journal of Chemical Ecology* **15**, 2635-2645.
15. Dekker, T., Geier, M. and Cardé, R.T. (2005). Carbon dioxide instantly sensitizes female yellow fever mosquitoes to human skin odours. *Journal of Experimental Biology* (in press).
16. Healy, T.P., Copland, M.J.W., Cork, A. Przyborowska, A. and Halket, J.M. (2002) Landing responses of *Anopheles gambiae* elicited by oxocarboxylic acids. *Medical and Veterinary Entomology* **16**, 126-132.
17. Murlis, J., Elkinton, J.S. and Cardé, R.T. (1992) Odor plumes and how insects use them. *Annual Review of Entomology* **37**, 505-532.
18. Mafra-Neto, A. and Cardé, R.T. (1974) Fine-scale structure of pheromone plumes modualtes upwind orientation of flying moths. *Nature* **369**, 142-144.
19. Vickers, N.J. and Baker, T.C. (1974) Reiterative responses to single strand of odor promote sustained upwind flight and odor source location by moths. *Proceedings of the National Academy of Sciences U.S.A.* **91**, 5756-5750.
20. Justus, K.A., Murlis, J., Jones, C. and Cardé, R.T. (2002) Measurement of odor-plume structure in a wind tunnel using a photoionization detector and a tracer gas. *Environmental Fluid Mechanics* **2**, 115-142.
21. Wagner, W.E., Jr. (1998) Measuring female mating preferences. *Animal Behaviour* **55**, 1029-1042
22. Cardé, R.T. and Haynes, K.F. (2004) Structure of the pheromone communication channel in moths. pp. 283-332 In *Advances in the Chemical Ecology of Insects*, R.T. Cardé and J.G. Millar, eds. Cambridge University Press, Cambridge.
23. Kennedy, J.S. (1977) Behaviorally discriminating assays of attractants and repellents. pp. 215-229 In *Chemical Control of Insect Behavior. Theory and Application*, H.H. Shorey and John J. McKelvey, Jr., eds. Wiley-Interscience, New York.
24. Kennedy, J.S. (1940) The visual response of flying mosquitoes. *Proceedings of the Zoological Society of London*, Ser A **109**, 221-242.
25. Kennedy, J.S. and Marsh, D. (1974) Pheromone-regulated anemotaxis in flying moths. *Science* **184**, 999-1001.
26. Marsh, D., Kennedy, J.S. and Ludlow, A.R. (1978) An analysis of anemotactic zigzagging flight in male moths stimulated by pheromone. *Physiological Entomology* **3**, 221-240.
27. Cooperband, M.F. (2005) Carbon Dioxide-Baited Traps and Bird-Associated Odors: Orientation Behavior and Chemical Ecology of Host-Seeking *Culex* Mosquitoes. Ph.D. Thesis University of California, Riverside, 243 pp
28. Vale, G.A. (1982) The trap-orientated behaviour of tsetse flies (Glossinidae) and other Diptera. *Bulletin of Entomological Research* **72**, 71-93.
29. Baker, T.C. and Cardé, R.T. (1979) Courtship behavior of the oriental fruit moth (*Grapholitha molesta*): experimental analysis and consideration of the role of sexual selection in the evolution of courtship pheromones in the Lepidoptera. *Annals Entomological Society of America* **72**, 173-188

Chemosensory recognition among mice

K. Yamazaki and G.K. Beauchamp
Monell Chemical Senses Center, Philadelphia, PA USA

Abstract
Body odor plays a prominent role in regulating social, sexual and endocrine responses of many species and specialized structures have evolved to produce and detect odorous signals. The major histocompatibility complex of genes (MHC) imparts to each mouse an individual odor, called an odor type, which reflects its MHC genotype. A prime experimental method we have used for identifying MHC odor types is a specially designed Y maze in which mice are trained, by water deprivation and reward, to distinguish odors from MHC dissimilar mice or their urines. An alternative or supplement to the Y maze is an automated olfactometer, a computer-programmed and fully automated apparatus in which a mouse or rat is trained for odor type distinctions, again by water deprivation and reward. In this presentation advantages and cautions associated with these methods are be discussed

Introduction
Body odor plays a prominent role in regulating social, sexual and endocrine responses of many species and specialized structures have evolved to produce and detect odorous signals. Individual recognition, often communicated through genetically-determined body odors, may be critical in mate choice, incest avoidance, parental care, and other inter-individual interactions. We have found that the same set of genes that code for individual identity in the immune system (i.e. those genes that insure rejection of foreign tissue and organ transplants) also provide an animal with a unique odor. These genes, termed Major Histocompatibility Complex (MHC) genes, form a linked set in all vertebrates and they include the most polymorphic of any known genes. They provide each animal with what we have termed an MHC odortype as has been demonstrated by our group and others in studies with mice, rats and humans.

Methods
A prime experimental method we have used to investigate MHC odortypes is the specially designed Y-maze (Fig.1). The purpose is to determine if mice can discriminate between body odors of 2 mice that differ only by genetic variation in specific MHC genes. If they can, this is proof that these genes modulate body odors. For training and testing in the Y-maze, gates are raised and lowered in timed sequence of up to 48 consecutive trials, the paired urine samples being changed for each trial. Reward for correct response is a drop of water, the trainee mouse having been deprived of water for 23 hr. Each mouse is first trained to discriminate between urine donors of two unrelated mouse strains, that differ in many genes throughout the genomes. Next, mice differing only at the MHC loci serve as odors stimuli. If the trained mouse is able to discriminate between these mice then interspersed unrewarded trials (about one out of every four) are included to familiarize the mice with occasional absence of reward following a correct response. The trained mice perform on the unrewarded trials with the same accuracy as on rewarded trials. Training and testing continue as described above but samples from interspersed unrewarded (generalization) trials are now supplied from new panels of H-2 types.

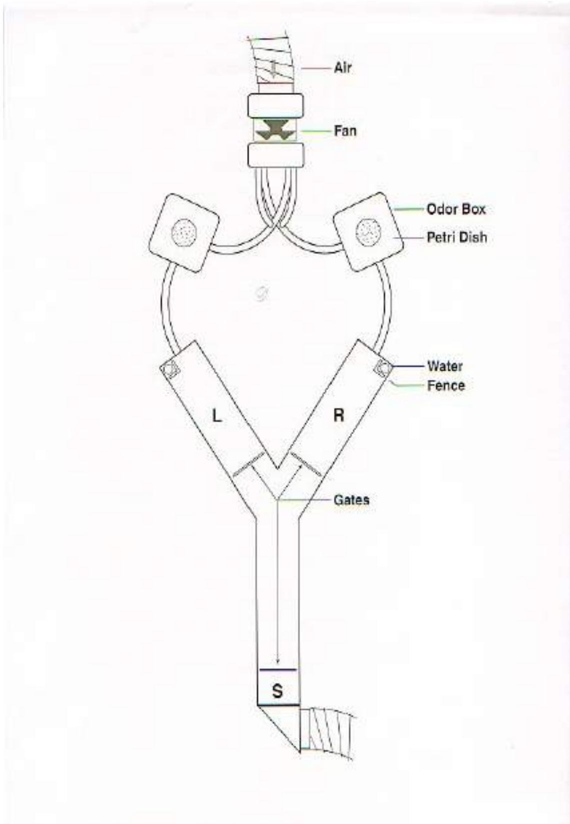


Figure 1. *Y-maze. Air drawn by a fan through a tube whose inlet is near the input vent supplying the laboratory is conducted through the left and right odor boxes. Each odor box has a hinged lid to admit a Petri-dish containing urine, the odor source. The air currents then pass to the left (L) and right (R) arms of the maze, which have hinged transparent lids. Each arm of the maze is fitted with a plastic tube perforated at the bottom to make one drop of water available. Each water tube is guarded by a fence that is raised only if the mouse enters the arm scented by the odor concordant with its training. Each arm of the maze is fitted with a gate that is lowered once the mouse has entered. If the choice is discordant, the fence is not raised, and the mouse is returned to the starting compartment (S). If the choice is concordant, the fence is raised to give access to the drop of water. The time interval in the starting compartment is set at 30 seconds to allow for changing the Petri-dishes in the odor boxes and for replacing the drop of water (if indicated); after this, on a timed signal, all three gates are raised to commence the next trial. Left-right placing is decided by a series for random numbers suited to the sample size. The time taken for trained mouse to make a choice is 2or 3 seconds; the choice is made without pause, or after sniffing at the entrance to the arms, or sometimes with brief retracing from one arm to the other.*

An alternative or supplement to the Y maze is the automated olfactometer, a computer-programmed and fully automated apparatus in which the rat is trained for odortype distinctions, again by water deprivation and reward. In this presentation advantages and cautions associated with these methods are discussed.

Supported by NSF Grant #0112528.

Above- and belowground olfactometers for high throughput bioassays

Ted Turlings[1], Anthony Davison[2], Ingrid Ricard[2], Cristina Tamo[1], Sergio Rasmann[1] and Matthias Held[1]
[1] *Institute of Zoology, University of Neuchâtel, Neuchâtel, Switzerland*
[2] *EPF-Lausanne, Mathematics/Statistics, Lausanne, Switzerland*

Abstract
We designed and tested two types of olfactometers, one for aboveground and one for belowground measurements of the attraction of small organisms to plant-produced signals. The equipment can be used for high throughput assays and allows for the simultaneous collection of the odors of the test sources. The olfactometers can be attached to each other, which permits us to study cross effects between above- and belowground interactions.

Keywords
Olfactometers, parasitoids, nematodes, plant odors, tritrophic interactions

Insect foraging behavior has long been the subject of many investigations because insight into how insects locate resources maybe exploited to manipulate the behavior of pest and beneficial insects to our advantage. Most insects use olfaction as a principal means of assessing their environment and their behavior is largely determined by what they perceive with their olfactory system. Therefore, in order to understand the behavior of insects it is pertinent to understand how they perceive bio-active volatile molecules and how they respond to them.
Customary procedures to determine attraction to odor sources and to identify the substances that are responsible for the attraction require tedious and time-consuming bioassays, as well as separate experiments using specialized methods for the collection, analyses, and identification of the chemicals. The inclusion of several different odor sources and treatments in such studies can make them exceedingly lengthy and impossible to accomplish within a reasonable time period. In most devices only one or two odors are tested at once, although various studies with small insects have employed four-arm olfactometers in which the behavioral responses of walking insects can be observed to multiple odors [1, 2].

We have expanded the number of odors that can be tested simultaneously with the development of two types of six-arm olfactometers to facilitate our research on the chemical signals emitted by insect-damaged plants that serve to attract parasitic wasps and entomopathogenic nematodes, which are natural enemies of the damaging insects. One olfactometer is used to study aboveground [3] and another to study belowground attraction [4]. Using log-linear models for the statistical treatment of the data allows for the testing of multiple individuals at the same time and the models control for overdispersion and positional effects [3]. In addition to the advantage of testing the relative attractiveness of up to six odor sources at the same time, the equipment is also designed to simultaneous sample part of the volatiles emitted by the odor sources, which can subsequently be identified and quantified. The 6-arm olfactometers offer several advantages over conventional olfactometer designs and have proven to be highly effective in aiding us to identify key attractants involved in tritrophic interactions. The above- and belowground olfactometers can be attached to each other and we currently use this feature to investigate how aboveground herbivory affects belowground interactions and visa versa.

References:

1. Pettersson, J. (1970) An aphid sex attractant. I. *Biological studies. Entomologica Scandinavica*, **1**, 63–73.
2. Vet, L.E.M., van Lenteren, J.C., Heymans, M. & Meelis, E. (1983). An air-flow olfactometer for measuring olfactory responses of hymenopterous parasitoids and other small insects. *Physiological Entomology*, **8**, 97–106.
3. Turlings, T. C. J., Davison, A. C. & Tamo, C. (2004). A six-arm olfactometer permitting simultaneous observation of insect attraction and odour trapping. *Physiological Entomology*, **29**, 45–55.
4. Rasmann, S., Köllner T. G., Degenhardt J., Hiltpold I., Toepfer S., Kuhlmann U., Gershenzon J., & Turlings Ted C. J. (2005). Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature*, **434**, 732-737.

3-D flight track analyses of *Culex* mosquito orientation to oviposition-related odors

M.A.H. Braks, M. F. Cooperband and R.T. Cardé
Department of Entomology, University of California, Riverside, USA

Abstract
A novel wind tunnel bioassay was developed for studying the oviposition-related behaviors of mosquitoes. A 3-D video recording and analysis system, initially created for studying the host-seeking behavior of mosquitoes, was adapted for analysis of upwind flight of mosquitoes in this assay. Individual gravid mosquitoes were released downwind of a turbulent odor (hay infusion or its control) plume and its flight track was recorded, digitized, and analyzed. Video analysis showed that a higher proportion of gravid female mosquitoes found and returned to hay infusion than to humid air. After obtaining the 3-D flight coordinates of a standardized section of the initial upwind flight track, multiple flight parameters were calculated and compared between the two stimuli. Here, we report two examples of important differences: the X,Y and X,Z tortuosities of females orientating to hay infusion were significantly greater than for the control. With this bioassay and recording system, events and flight patterns to different olfactory stimuli can be studied. By digitizing a standardized, relatively short segment of each flight track, orientation behavior to different odor stimuli can be distinguished.

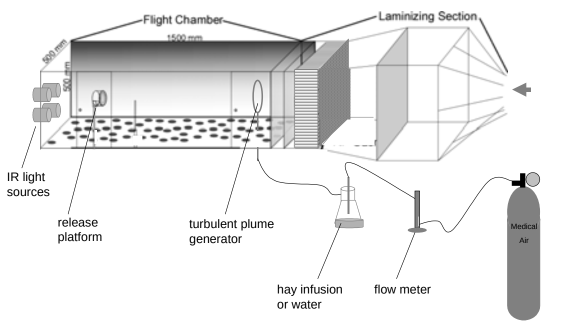
Keywords
Mosquito, oviposition, orientation, 3-D analyses, wind tunnel

1 Introduction
Egg-carrying or gravid mosquitoes use a combination of sensory modalities to find and determine the suitability of water-containing sites to lay their eggs. Wind-borne odorants appear to be particularly important in this process. The odors promoting egg laying and finding of suitable egg-laying sites have been classified as attractants, arrestants, and stimulants, whereas those that inhibit these reactions have been termed deterrents and repellents [1]. A number of semiochemicals for the southern house mosquito, *Culex pipiens quinquefasciatus* (a vector of human lymphatic filariasis in humid tropical countries and West Nile virus in the USA) are known, including the "egg raft pheromone" [7], as well as compounds identified from the headspace above water from suitable oviposition sites [4, 6]. Studies designed to identify these semiochemicals have depended mainly on bioassays measuring end-point behaviors, that is, did a given odor increase or decrease the number of eggs deposited. The preference for this simple bioassay is, at least in part, motivated by the facts that it requires little labor and provides a readily quantifiable metric. However, little is known about on the role of these cues in the orientation of mosquitoes to a suitable oviposition site.
As in any behavioral study, a diagnostic bioassay is crucial. Here, we describe the initial experiments assessing a novel bioassay for the examining orientation behavior of the *C. p. quinquefasciatus* to oviposition cues in a wind tunnel. Video recording and analysis in 3-D, originally developed to examine orientation maneuvers of host-seeking mosquitoes [2, 3], were adapted for this purpose. To appraise the suitability of the bioassay, the orientation behavior to a known oviposition medium, hay infusion, commonly used as an attractant in gravid traps (portable battery powered devices to capture gravid females in arboviral surveillance programs) was compared with its control.

2 Material and Methods

2.1 Mosquitoes
The *C. p. quinquefasciatus* strain originated from Riverside County, CA, (courtesy of Prof. W. Walton) and was maintained under standard conditions (27°C, 70% RH). The mosquitoes were held in a light-dark cycle, with a 14 h day, including a 1 h artificial dawn and dusk period (1 incandescent light bulb of 25W), lasting from 5 PM to noon, which facilitated assaying of these nocturnal mosquitoes.

a.



b.

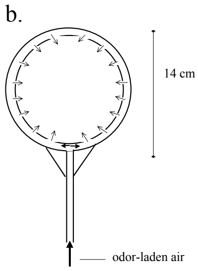


Figure 1. a. *Wind tunnel set up, which provides laminar flow through the flight chamber.* ***b.*** *Turbulent plume generator.*

Eggs were set up to hatch biweekly and the larvae were reared on an artificial diet (3:1 rat chow: yeast mixture) in 3.8 L containers. Pupae were collected daily and allowed to emerge in 30 by 30 by 30 cm screen cages (Bioquip®) where they were provided with a 10% sucrose solution. A week to ten days after emergence, the mosquitoes were offered the opportunity to feed on white leghorn chicks for 5 hours. Seven days after blood feeding, females, which were deprived from water and sugar for 24 hrs, were used in the bioassay. Twelve gravid and non-gravid females were selected and transferred to release cages (8 cm diameter, 6 cm height) 2 hours before testing. Each release cage contained one mosquito. Experiments were conducted during the first 5 hours of the scotophase.

2.2 Experimental set-up and flight recording

Flight behavior of mosquitoes was examined in a wind tunnel (Fig. 1a) with a transparent Plexiglas flight chamber of 150 cm (length) by 50 cm (width) by 50 cm (height), the sides of which were covered with Medium Red light filter (Roscolux, which blocks light <600 nm) to prevent orientation of the mosquitoes to visual cues outside of the tunnel. Transparent red dots (6 cm Ø, approx. 100 m^{-2}, Medium Red, Roscolux) randomly arranged on the floor of the wind tunnel provided "non-directional" optomotor cues. Air was taken from outdoors, filtered by an activated charcoal, and pushed by a centrifugal in-line duct fan through the wind tunnel. A turbulence-free airflow of 50 cm/s was created by passing the air through an activated carbon air filter, an aluminum honeycomb laminizer (consisting of cells of 1.5 cm diameter and 15 cm long), and two stainless-steel screens (150 mesh). Mosquitoes were illuminated against the background by four infrared lights (UFL 694, Rainbow) at the tunnel's downwind end. Each had sixty 940 nm LEDs, equipped with cut-off filters permeable to light of >950 nm. During the experiments, light level in the human visual spectrum from diffuse incandescent lights was approximately 15 lux.

Odors were delivered into the wind tunnel as a broad turbulent plume by pushing the odor-laden air (5 L/min) into a plume generator, placed 27 cm downwind from the two laminizing screens. To minimize visual cues, the turbulent plume generator consisted of a glass ring (5 mm Ø) with 19 1 mm holes on the inner side of the ring equidistant from each other, through which the odor-laden air would exit into the flight chamber (Fig. 1b). A release cage was transferred to a release platform 14 cm above the wind-tunnel floor, 130 cm downwind from the upwind screen (100 cm upwind from the plume generator), such that the cage was inside the broad odor plume. The opening of the cage faced upwind and was placed against a screen. After 1 minute, the screen was removed and the mosquitoes were allowed to initiate flight for 3 minutes. The flight was recorded with two Sanyo black and white video cameras (VCB-3112T, shutter speed set at 1/60 s) equipped with 6 mm lenses, one view from the side and one from the bottom of the wind tunnel. The camera views were synchronized with an Event & Video Control Unit (Peak Performance Technologies, Centennial, CO, USA), overlaid and recorded on one video record. Twelve calibration points (5 mm Ø) were present on the floor (6) and ceiling (6) of the flight chamber.

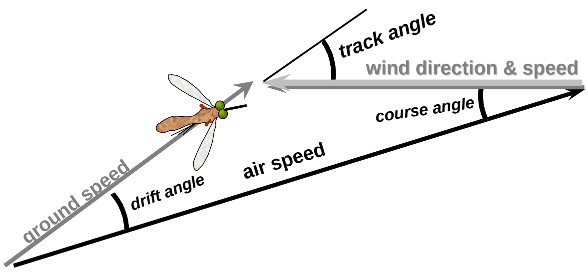
2.3 Testing procedures

Odor-laden air was produced by introducing clean air obtained from a pressurized cylinder of medical air via a flow meter over either 20 mL of hay infusion or water in a 100mL Erlenmeyer flask. Hay infusion was prepared with the formula used by the Mosquito Control Research Laboratory in Parlier, CA (University of California, Davis) as an oviposition medium in gravid traps: 5.6 grams of dried and shredded Bermuda grass and 0.34 g of bakers yeast per 2.5 L water. The hay infusion in a 3.8 L glass jar was placed in a hood in the laboratory. After 7 days, the infusion was filtered though a cloth to remove the grass fragments. The infusion was divided into 200 ml aliquots in Ziplock® bags and frozen. On the morning of an experiment, one bag was placed in warm water until the hay infusion thawed. Each experimental day consisted of four sets of six individual mosquitoes per treatment: gravid and non-gravid females were both tested to hay infusion and water. The orders of the sets were randomized within each experimental day. Here, we present only the data of three days and only of gravid females assayed for hay infusion (N=18) and water (N=16).

2.4 Data analyses

The numbers of mosquitoes activated (exiting the release cage), and the proportions of activated females finding the source (flying within the 1 cm vicinity of the odor source), landing on the odor source, and returning (flying back to downwind side after passing the odor source before reaching the upwind screen of the wind tunnel) were determined.

a.



b.

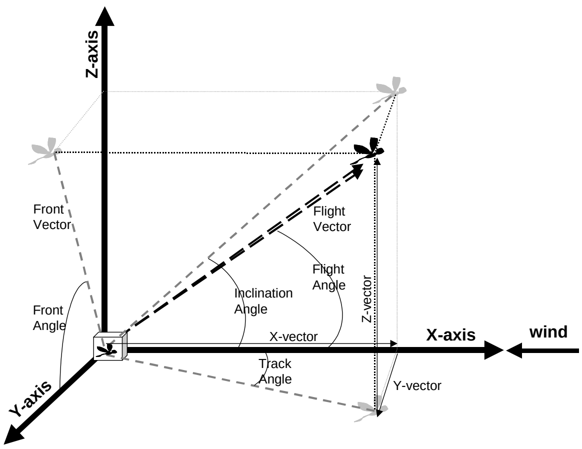


Figure 2. a*. The triangle of velocities looking down from above the mosquito (in the x-y plane)[8]****. b****. Visualization of the angles and vectors used to calculate 3-D flight parameters with the mosquito's location in the three planes.*

For each track, the resulting pairs of 2-D flight tracks from the two camera views were digitized using Motus 32 (Peak Performance Technologies) software at 30 frames/s. By means of the fixed coordinates from the 12 calibration points present in the flight chamber, accurate 3-D flight coordinates of mosquitoes flying through the same space could be calculated with Motus. In addition to the mosquito flight, the position of the release cage, and four points of the glass plume generator (top, bottom, left and right edge) and the four central calibration points, creating a virtual vertical plane at 90 cm down wind of the laminizing screens, were digitized in each frame. Due to the large variation in the duration of the flight tracks, only a standardized segment was digitized. The flight was digitized from the moment the mosquito crossed the virtual plane to when it either reached the end of the tunnel or, in case the mosquito returned (see above), arrived again downwind of the odor source. Not all tracks were recorded with sufficient clarity for analysis of all frames. Because two cameras were used from different angles, sometimes the mosquito was visible in one, but not in the other video record. Because no 3-D coordinates could be calculated for those frames, this would appear as a gap. If more than 30% of all frames of a track were gaps, that track was excluded from further analyses. In the remaining tracks, the missing coordinates of the gaps were interpolated using Motus.

The 3-D coordinated were entered into an in-house computer program (Track 3-D, version 2.2.6, written by Joseph Bau), which calculated variety of flight parameter in one, two or three dimensions, such as: X speed (along wind velocity), Y speed (across wind velocity), Z speed (vertical velocity, track angle (X,Y), course angle (X,Y and 3-D), drift angle (X,Y and 3-D), inclination angle (X,Z), front angle (Y,Z), ground speed (X,Y), airspeed (X,Y and 3-D) time elapsed, flight angle (3-D), flight speed (3-D), and tortuosity (X,Y,Z and 3-D). Figure 2 provides diagrams of 2-D and 3-D flight parameters. Tortuosity was calculated dividing the distance traveled between the starting point and the end by the straight distance between those two points. The tortuosity index is always ≥ 1, and, the higher the value, the more turning. Flight tracks were divided into 5 cm increment bins along the X-axis, starting at the position of the plume generator/odor source, which is in the middle of bin (0).

Table 1. a. *The proportion of* C. p. quinquefasciatus *females tested, activated, finding the odor source, returning and landing (see also text).* ***b.*** *Duration ± SD of different segments of the track (asterisk indicates significant difference between the two stimuli, P<0.05).*

a. Proportion of mosquitoes (N)	**Stimulus** Hay Infusion	**Water**
Tested	(16)	(14)
Activated	0.94 (15)	0.79 (11)
Finding source	0.60 (9)	0.09 (1)
Returning	0.67 (10)	0.27 (3)
Landing	0.07 (1)	0.00 (0)
b. Track duration^ (SD)		
t_0 to t_{exit}	55.4 (11.6)*	12.3 (4.8)
t_{exit} to t_{hoop}	10.6 (1.7)	13.7 (4.4)
t_{plane} to $t_{odor\ source}$	5.9 (2.7)	3.3 (0.8)

^ t_0 = time of opening release cage
t_{exit} = time of mosquito exiting release cage
t_{plane} = time of mosquito crossing virtual vertical plane,
$t_{odor\ source}$ = time of mosquito arriving upwind of odor source

Track duration for flight tracks were compared using Games-Howell *post hoc* analysis of variance with *P* =0.05 [5] using StatView 5.0.1. This test was selected because it does not assume equal sample sizes or equal variances. Means of flight parameters per bin were quantified for individual mosquitoes grouped by odor stimuli and compared and tested for significance using Games-Howell *post hoc* analysis of variance with *P* =0.05. Here we present only two flight parameters X,Y and Z,Y tortuosities, which were compared between odor stimuli.

3 Results and Discussion

3.1 Events

Large proportions of gravid *C. p. quinquefasciatus* females flew upwind to both a complex oviposition attractant (0.94) and its control (0.79) (Table 1a).

High activation levels are essential for effectively testing individual mosquitoes to various odor stimuli. With hay infusion, despite the fact females took longer to initiate flight (Table 1b), the majority of the females (0.60) found the odor source within three minutes. Eight females flew in close vicinity (7 cm) to the odor source in their initial upwind flight with three females actually flying through the plume generator (Figure 3). After arriving upwind of the odor source, a larger proportion of gravid females returned to the hay infusion plume than to the control plume, before reaching the upwind screen of the wind tunnel (Table 1b). Nine of the fifteen activated females stayed in close vicinity of the odor source for more than 30 s and one female landed on the glass hoop of the plume generator and stayed there till the end of the experimental trial.

a.

b.

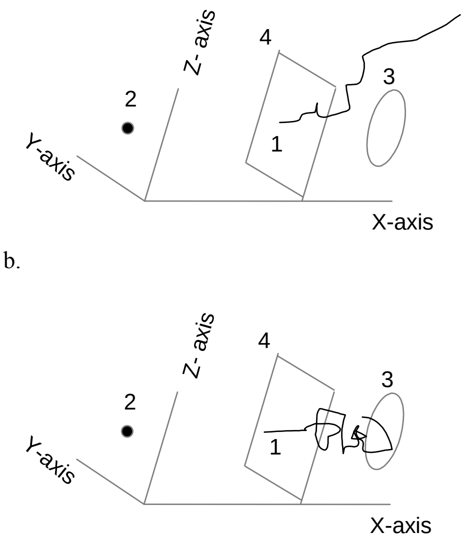


Figure 4. *Representative examples of tracks of a* C. p. quinquefasciatus *female flying upwind to water (**a**) and hay infusion (**b**). 1: mosquito track, 2: release cage, 3: odor source, and 4: virtual vertical plane.*

Figure 4 shows typical 3-D flight tracks of *C. p. quinquefasciatus* females flying upwind to hay infusion and the water control. After obtaining the 3-D flight coordinates of a standardized section of the initial upwind flight track, multiple flight parameters were calculated and compared between the two stimuli. Here, we report only two fight parameters, X,Y and Y,Z tortuosities. X,Y tortuosity of females orientating to hay infusion was significantly higher than ones of females flying to the control stimulus. With hay infusion, the peak of X,Y tortuosity occurs just upwind of the odor source, which is a reflection of the high proportion of females returning to downwind of the odor plume after exiting the plume. However, considering Y,Z tortuosity, a significantly higher peak with hay infusion occurs just downwind of the odor source, which suggests that females hovered at this position.

Traditionally, insect flight along odor plumes has been recorded and analyzed in the X,Y plane [8]. Pile *et al.* [9, 10] used this technique for analyzing orientation behavior of *C. p. quinquefasciatus* mosquitoes to oviposition cues in a small wind tunnel. Although a planar analysis of moth flight along pheromone plumes provides useful information, largely because there is comparatively little vertical movement, this does not hold for mosquito flights. Mosquitoes move and turn very rapidly in all 3 planes and crucial information is lost in 2-D analyses. This study is the first report of 3-D flight analyses of orientation behavior to oviposition-related odors.

a

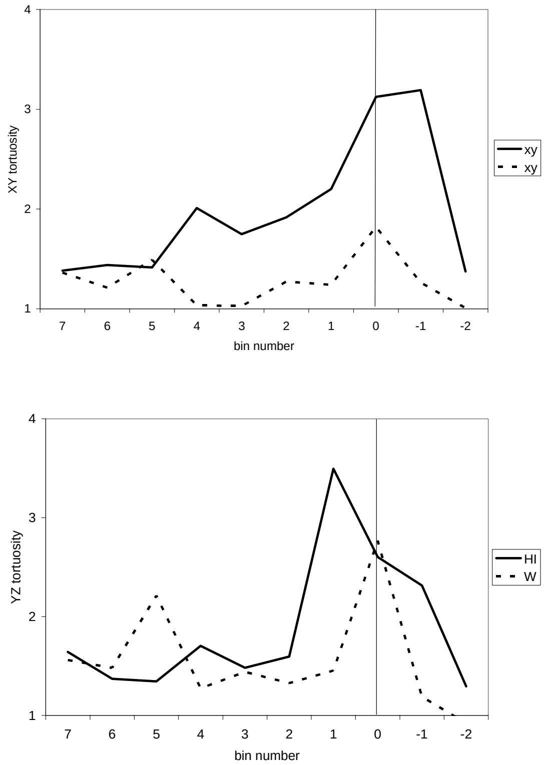


Figure 5. *Average X,Y* ***(a)*** *and Y,Z* ***(b)*** *tortuosities of* C. p. quinquefasciatus *females at distances of 5 cm increments from the odor source, bin (0); HI= hay infusion, W= water control.*

We thank J. Bau for programming the Track 3-D software. This research was part of project entitled "Control of urban and peri-urban Culex mosquitoes" funded by National Institutes of Health (1-U01-A10058267).

References

1. Clements, N.A. (1999) The Biology of Mosquitoes, Vol. 2. CABI Publishing. Wallingford, U.K.
2. Cooperband, M.F. (2005) Carbon Dioxide-Baited Traps and Bird-Associated Odors: Orientation Behavior and Chemical Ecology of Host-Seeking *Culex* Mosquitoes. Ph.D. Thesis University of California, Riverside, 243 pp
3. Dekker, T. (2002) Influence of plume structure, composition and concentration on orientation behavior of mosquitoes to host odors. Ph.D. Thesis University of California, Riverside, 305 pp.
4. Du, Y. and Millar J.G. (1999) Electoantennogram and oviposition bioassay responses of *Culex quinquefasciatus* and *Culex tarsalis* (Diptera: Culicidae) to chemicals in odors from Bermuda grass infusions, *Journal of Medical Entomology,* **36**, 158-166.
5. Games, P.A. and Howell, J.F. (1976) Pairwise multiple comparison procedures with unequal n's and/ or variances: A Monte Carlo study. *Journal of Educational Statistics*, **1**, 113-125.
6. Isoe, J. and Millar, J.G. (1995) Characterization of factors mediating oviposition site choice by *Culex tarsalis*. *Journal of the American Mosquito Control Association*, **11,** 21-28.
7. Laurence, B.R. and Pickett, J.A. (1982) Eythro-6-acetoxy-5-hexadecanolide, the major component of a mosquito oviposition attractant pheromone. *Journal of the Chemical Society - Chemistry Communications*, **1**, 59-60
8. Marsh, D., Kennedy, J.S. and Ludlow, A.R. (1978) An analysis of anemotactic zigzagging flight in male moths stimulated by pheromone. *Physiological Entomology* **3**, 221-240.
9. Pile, M.M. Simmonds, M.S.J. and Blaney, W.M. (1991) Odor-mediated upwind flight of *Culex quinquefasciatus* mosquitoes elicited by a synthetic attractant. *Physiological Entomology* **16**, 77-85.
10. Pile, M.M. Simmonds, M.S.J. and Blaney, W.M. (1993) Odor-mediated upwind flight of *Culex quinquefasciatus* mosquitoes elicited by a synthetic attractant--a reappraisal. *Physiological Entomology* **18**, 219-221.

A video-based movement analysis system to quantify behavioral stress responses of individual and groups of fish

A.S. Kane[1,2,] J.D. Salierno[1,2], G.T. Gipson[2]
Aquatic Pathobiology Center
[1]Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine
[2]Virginia-Maryland Regional College of Veterinary Medicine, 8075 Greenmead Drive, College Park, MD 20742 USA

Abstract

Group behavior, aerial predator avoidance, mummichog
Behavioral alterations can serve as quantitative endpoints of sub-lethal toxicity, and provide a valuable tool for environmental risk assessment. A video analysis system and software algorithms were designed to investigate stress- and contaminant-induced behavioral alterations in fish, using a mummichog (*Fundulus heteroclitus*) model. The videography system automatically recorded individual and group fish behavior at 30 fps from up to twelve 20-L exposure arenas throughout an experimental period. Analytical endpoints included velocity, burst frequency, total distance traveled, angular change, percent movement, space utilization, fractal dimension (path complexity), interactions, schooling, shoaling, and aggregations. Movement endpoints discerned by this system can be used to identify responses to a variety of stressors at environmentally-relevant concentrations, and assist in risk assessment and the development of more sensitive 'lowest observable effect levels' and 'no observable effect levels' for a variety of sentinel species.

Keywords

Fish, behavior, toxicology, movement analysis, sublethal

1 Introduction

Fish are ideal sentinels and test organisms for behavioral assays of various stress factors and toxic chemical exposure due to their ecological relevance in many natural systems [1]. In addition small fish species may bioaccumulate various contaminants and/or play a role in food web biomagnification. Behavioral endpoints in fish, as well as other organisms, are valuable tools to discern and evaluate effects of environmental stress. These endpoints of exposure are important because they can integrate endogenous and exogenous factors that can link biochemical and physiological processes, and can provide insights into individual- and community-level effects of environmental contamination [2, 3]. Quantifiable behavioral changes in organisms associated with stress and toxicant exposure provides novel information that cannot be gained from traditional toxicological methods, including short-term and sublethal exposure effects, and the potential for mortality [4-7]. Ecological relevance of such exposures may stem from altered vigilance, startle response, schooling, feeding, prey conspicuousness, migration, and diurnal rhythmic behaviors [4, 8].

We designed and developed a system to quantify individual and group behaviors, using a mummichog (*Fundulus heteroclitus)* model. The system software, in conjunction with a videography system [9, 10], can quantify a variety of individual and group behaviors for up to twelve individual or groups of fish simultaneously. The ultimate goal is to utilize this system to quantify alterations in swimming speed, angular movement, path tortuosity, space utilization, and aggregative and interactive behaviors in association with exposure to stressors such as contaminants, toxins, and environmental fluctuations. In this current study, alterations in mummichog behaviors over time, associated with laboratory acclimation or exposure to an anesthetic (MS-222), are exemplified.

2 Methods

Twelve 20-liter exposure arenas were constructed from 14-inch (35.6cm) diameter polyvinyl chloride (PVC) pipes and end caps. Each arena had two 0.25-inch (0.64cm) threaded nipples that served as an input and drain. The drain line lead to an adjustable height standpipe, which was set to achieve a 5 L exposure volume. The input was bifurcated to accept both toxicant and dilution water flow lines. Alternately, glass aquaria of various sizes with overflow standpipes can implemented to collect different data from different viewpoints (above, beneath, side). Toxicant and dilution water flow were electronically controlled with digital, multi-channel peristaltic pumps (Masterflex L/S, Cole-Parmer, Vernon Hills, IL) that were supplied by multiple, aerated 600-liter carboys. The exposure arenas were illuminated with shadowless fluorescent lighting. Twelve color CCD cameras with manual iris and focus control were mounted above the respective arenas and were connected to dedicated VCR decks with time-lapse and dry contact closure capability for recording (Figure 1). All twelve VCRs were connected to a multiplexer and monitor that supported real-time observation. VCR recording and stop functions were synchronously activated remotely using X-10 technology.

Analog video data were then digitized in real time at 3 frames per second, on a Macintosh platform (PowerPC G5, 2 GHz, 4 GB SDRAM). Digitized video data was analyzed using a commercial tracking and image processing program, Videoscript Professional, version 3.1©. The tracking system generated x,y data points from the middle of the fish body for consistent tracking over time. For groups of fish, a least squares method was applied to reduce tracking error when individual fish cross paths.

The x,y coordinate data was then analyzed using software designed at the Aquatic Pathobiology Center to obtain the desired behavioral endpoints (Table 1). Sample baseline and MS-222 exposure data for an individual fish are presented in Figure 2. Sample baseline data for an acclimation study with 10 groups of 5 fish are presented in Figure 3.

Individual Endpoints	Definition
Percent Movement	The # of seconds the fish satisfies movement criteria divided by the total # seconds spent swimming, multiplied by 100.
Velocity	Average velocity (cm/sec) while the fish is moving during the experimental period.
Angular Change	The difference (0-180°) between the angular components of two consecutive 1 second movement vectors (degrees/sec) divided by the total # of consecutive 1 second movement events. Angular change was only calculated when two consecutive movement vectors met the movement criteria.
Path Tortuosity (Fractal dimension)	Fractal dimension (D) is calculated using the hybrid divider method (Hayward et al., 1989) and is an indicator of path complexity. A series of path generalizations are created at various step sizes and the data are mathematically drawn on a Richardson plot (Log path length vs. Log step length). D is then calculated as 1 minus the slope of the Richardson plot. D=1 if the path is a straight line and 2 if it completely fills the 2 dimensional plane.
Space Allocation	The # of frames fish spend in predefined regions of the exposure arena divided by the total # frames.
Distance from Center	Sum of individuals distance from the center of the exposure arena (cm) divided by the total # frames. This is a measure of how close the fish swims to the walls of the arena.
Relative Burst Frequency	The # of frames that velocity is > 3 SDs above mean velocity.
Startle Response	Response latency, response duration, percent responding, and burst swimming (Response to vibratory/auditory stimulus)
Anti-predator response	Percent fish halting movement, latency to response, percent exhibiting startle-type response, direction of movement (toward or away visual stimuli), and group endpoints (Response to overhead “fly-by” of bird silhouette).
Group endpoints	Definition
Interactions	The number of times two fish swim within 0.1 body lengths of each other.
Percent Shoaling	Number of frames satisfying shoaling criteria divided by the total number of frames.
Shoal NNA	Angle of trajectory between 2 fish in a shoal, must be greater than 45°.
Shoal NND	Distance between nearest and second nearest neighbor for each fish in a shoal (minimum 3 fish).
Percent Schooling	Number of frames satisfying schooling criteria divided by the total number of frames.
School NNA	Angle of trajectory between 2 fish in a shoal, must be less than or equal to 45°.
School NND	Distance between nearest and second nearest neighbor for each fish in a shoal (minimum 3 fish).
Percent Solitary	Number of frames not satisfying shoaling or schooling criteria divided by the total number of frame
Solitary NND	Distance between nearest and second nearest neighbor for each individual fish not in a shoal school.
Velocity	Speed of fish calculated in centimeters per second.

Table 1. *Individual and group endpoints for movement analysis.*

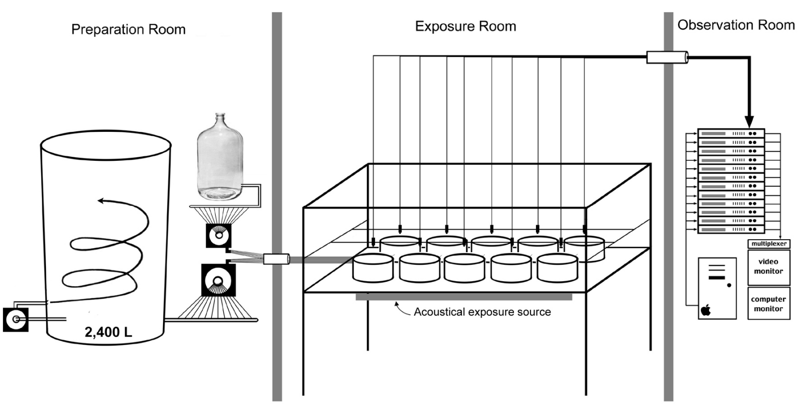


Figure 1. *Schematic showing flow-through exposure arenas with dedicated, overhead video cameras (shadowless lighting and drainage not shown) in the behavioral toxicology exposure suite at the University of Maryland Aquatic Pathobiology Center. Water is prepared, temperature adjusted and delivered from the adjacent preparation room, while video data is amassed by computer-controlled video decks in the observation room. Toxicants or aqueous stress agents, and flow-through dilution water, are pumped from the preparation room to the exposure room by computer-controlled peristaltic pumps. An instantaneous acoustic/vibratory stimulus, or a visual stimulus, can be provided to discern differences in startle response behavior. In this diagram, an acoustical exposure source, used to examine startle response time and duration, is indicated under the platform supporting the exposure arenas.*

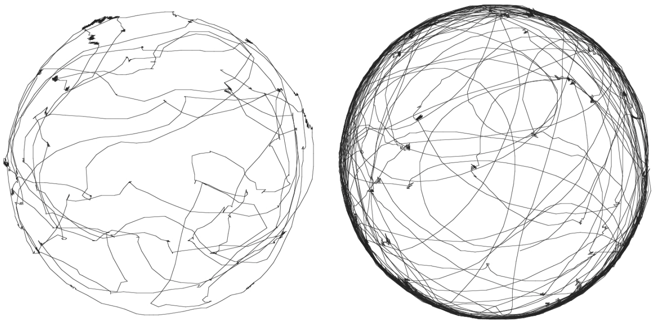


***Figure 2**. Paths of an individual killifish (Fundulus heteroclitus) before (left) and after (right) exposure to a sub-anesthetic dose (60 mg/L) of MS-222. Analysis of these 30-min paths indicate that exposure to low doses of this anesthetic agent causes an increase in percent time in motion (from 12 to 49%) and movement velocity (9.1 to 11.9 cm/sec), a decrease in path complexity (fractal dimension of 1.082 to 1.027), and a tendency to swim close to the arena periphery (change in distance from center). All of these endpoints describe quantifiably significant alterations in movement associated with exposure. Functionally, exposed animals tended to increase their speed and stay in motion to compensate for slight loss of equilibrium. The "intoxicated state" was depicted by "hugging" the vessel walls during movement and failing to maintain vigilance (loss of path complexity). MS-222 exposure also significantly altered the startle response of exposed fish such that there was a decrease in the number of responses, a decline in the response duration, and an increased response latency (P<0.02; data not shown).*

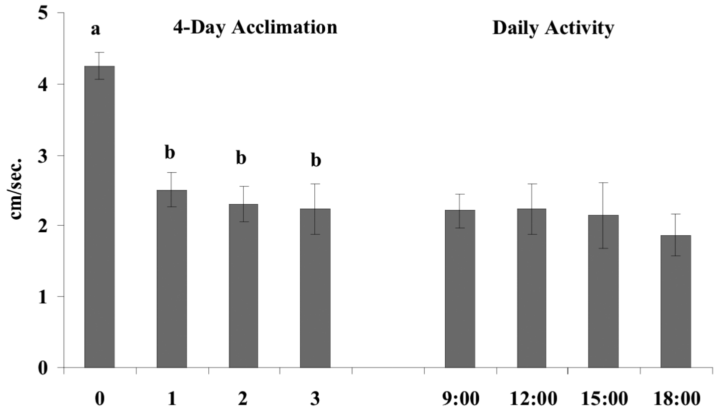


***Figure 3.** Shoaling velocity data for acclimationof groups of mummichog over a 3 day acclimation period (Day 0, 1, 2, and 3) and the final day (0900, 1200, 1500, and 1800) experiments. Data are means ± S.E. Times with different letters are significantly different at the p≤ 0.001 level.*

2.1 Statistical Analyses

For all experiments, the experimental unit consisted of an individual fish or a group of 5 fish in each arena. A completely randomized statistical design was used with behavior as the response variable and time (e.g., day 0, 1, 2, 3…n) as the categorical variable. Non-normal data were either log or square root transformed to meet the assumptions of the parametric ANOVA procedure. Data that could not be transformed to meet the assumptions were ranked prior to analysis. A repeated measures 1-way ANOVA (PROC MIXED, repeated, SAS, vs. 8.1, Cary, N.C.) was used to compare behavioral endpoints over time for both experiments. In addition, several covariance structures can be investigated to discern a best fit structure for different sets of data. The F and p values reported for each endpoint result from the structure that provided the best fit. Covariance structures included compound symmetry (CS), heterogeneous compound symmetry (CSH), auto regressive (AR-1), heterogeneous auto regressive (ARH-1), and ante-dependence (ANTE-1). A Tukey-Kramer *post-hoc* mean comparison test was used to evaluate differences (t value, $\alpha \leq 0.10$) between time periods in the event of a significant F statistic.

3 Summary

The video tracking and analysis system described has the ability to investigate changes in structure and function of group dynamics as well as movement patterns of fish through time. The endpoints used by this system are biologically relevant, and represent social and movement behaviors that occur on a continuous basis in nature. Changes in these behaviors, in association with stress or toxicant exposure, can directly affect individual fitness (i.e., altered reproduction, competition for resources, predator avoidance and habitat utilization). Movement endpoints discerned by this system can be used to identify responses to a variety of stressors at environmentally-relevant concentrations, and assist in risk assessment and the development of more sensitive lowest observable effect levels and no observable effect levels for a variety of sentinel species.

4 Acknowledgements

Support for this project came from the U.S. Environmental Protection Agency (Science to Achieve Results Program, STAR Award #R82-8224), and matching equipment funds from the State of Maryland through the Maryland Department of the Environment. We also thank Edward Little, Aaron DeLonay and Sandra Brewer of the U.S. Geological Service, Columbia Environmental Research Center, for their kind assistance and insight.

References

1. Little, E.E.;Archeski, R.D.;Flerov, B.A. and Kozlovskaya, V.I. (1990). Behavioral Indicators of Sublethal Toxicity in Rainbow-Trout. Archives of Environmental Contamination and Toxicology, 19, 3, 380-385.
2. Brewer, S.K.;Little, E.E.;DeLonay, A.J.;Beauvais, S.L.;Jones, S.B. and Ellersieck, M.R. (2001). Behavioral dysfunctions correlate to altered physiology in rainbow trout (Oncorynchus mykiss) exposed to cholinesterase-inhibiting chemicals. Archives of Environmental Contamination and Toxicology, 40, 1, 70-76.
3. Vogl, C.;Grillitsch, B.;Wytek, R.;Spieser, O.H. and Scholz, W. (1999). Qualification of spontaneous undirected locomotor behavior of fish for sublethal toxicity testing. Part I. Variability of measurement parameters under general test conditions. Environmental Toxicology and Chemistry, 18, 12, 2736-2742.
4. Little, E.E. and Finger, S.E. (1990). Swimming Behavior as an Indicator of Sublethal Toxicity in Fish. Environmental Toxicology and Chemistry, 9, 1, 13-19.
5. Henry, M.G. and Atchison, G.J. (1986). Behavioral-Changes in Social-Groups of Bluegills Exposed to Copper. Transactions of the American Fisheries Society, 115, 4, 590-595.
6. Bridges, C.M. (1997). Tadpole swimming performance and activity affected by acute exposure to sublethal levels of carbaryl. Environmental Toxicology and Chemistry, 16, 9, 1935-1939.
7. Saglio, P. and Trijasse, S. (1998). Behavioral responses to atrazine and diuron in goldfish. Archives of Environmental Contamination and Toxicology, 35, 3, 484-491.
8. Zhou, T. and Weis, J.S. (1998). Swimming behavior and predator avoidance in three populations of Fundulus heteroclitus larvae after embryonic and/or larval exposure to methylmercury. Aquatic Toxicology, 43, 2-3, 131-148.
9. Kane, A.S.;Salierno, J.D.;Gipson, G.T.;Molteno, T.C.A. and Hunter, C. (2004). A video-based movement analysis system to quantify behavioral stress responses of fish. Water Research, 38, 18, 3993-4001.
10. Kane, A.S.;Salierno, J.D. and Brewer, K.L. (2005). Fish models is behavioral toxicology: Automated techniques, updates, and perspectives. Techniques in Aquatic Toxicology. New York, Taylor and Francis.

Development and use of in-stream PIT-tag detection systems to assess movement behavior of fish in tributaries of the Columbia River Basin, USA

P.J. Connolly[1], I.G. Jezorek[1], and E.F. Prentice[2]

[1] *U.S. Geological Survey, Western Fisheries Research Center, Columbia River Research Laboratory, Cook, WA, USA*

[2] *National Marine Fisheries Service, Manchester Research Center, Manchester, WA, USA*

Abstract

We have developed a detector system for fish implanted with Passive Integrated Transponder (PIT) tags to assess their movement behavior and habitat use within fast flowing streams. Fish tested include wild rainbow trout *Oncorhynchus mykiss* and coastal cutthroat trout *O. clarki clarki*. Our first detector system became operational in August 2001, with subsequent improvements over time. Longitudinal arrangement of two single antennas and two paired sets of side-by-side antennas have allowed determination of direction of fish movement and allowed estimation of the efficiency for detecting PIT-tagged fish. In the first 21 months of operation with two single antennas, the detection efficiency of downstream-moving fish during low flows was 94%, but reduced to 69% during high flows. By adding antennas lateral to the primary antennas to create two paired sets, and along with the subsequent availability of PIT tags with increased read range, the detection efficiency of downstream-moving fish was increased to over 98% at all flows. Detection efficiency of upstream-moving fish was high (over 88%) in both the two- and paired-set antenna system during all flows. The primary antenna of each set was anchored to the substrate within the thalweg, while the secondary antenna was in shallower areas that may or may not have been wet during times of low stream flow. Modifications to the anchoring of antennas have increased the likelihood of the system performing during challenging flow and debris conditions. Our goal to have continuous monitoring of fish movement throughout the year has largely been achieved. Success has relied on the improvement of transceiver technology, on quality of electrical power to run the system, and our ability to adequately anchor the antenna in the stream. Detection of tagged fish passing our stationary PIT-tag detector is providing valuable information on how selected fish species behave and how they use a stream network.

Keywords

Fish movement, PIT tags, detection efficiency, interrogation system.

1 Introduction

Because of interest in monitoring the fate of individual fish for studies of habitat use, population structure, survival, and responses to environmental variables, the use of Passive Integrated Transponder (PIT) tags has large potential and appeal [5,10,11]. Because PIT tags can uniquely identify individual fish over many years, they have potential to provide information throughout the life span of a fish, which can be 10 years or more for some species. Use of PIT tags in fish research has recently increased, particularly in the Columbia River Basin in the northwest United States. These tags have become a primary tool for monitoring juvenile salmonid movement and for estimating survival past large hydroelectric dams [1,7,8,13].

Growth of individual fish and movement behavior can be assessed by subsequent recapture of previously PIT-tagged fish [5,9]. The amount of information gained, however, can be limited by the ability to recapture fish. Additional behavioral information has been gained by placing PIT-tag detector systems in streams to detect passing fish [2,3,4,15]. Our desire was to create a reliable, stationary system to detect fish passing for a continuous period of time, including times with severe conditions such as high flows and ice cover that are difficult to sample by conventional means.

To help identify behavioral and life history characteristics of rainbow trout *Oncorhynchus mykiss* and coastal cutthroat trout *O. clarki clarki* in our study stream, we PIT tagged 600-1,500 fish per year since 2001. We used 12 mm, full duplex PIT tags (134.2 kHz), which have the advantage of being detected at many of the mainstem dams in the Columbia River Basin. These dams have been retrofitted with PIT-tag interrogation systems to provide data for restoration of threatened and endangered fish stocks and for information as to how best manage fisheries and operate the hydroelectric system. We were particularly interested in determining timing and direction of fish movement past a single point near the mouth of our study stream, and we desired the ability to measure the efficiency of detecting passing fish. To date, use of stationary detector systems in flowing water have been highly customized to meet criteria of a particular study and have been generally limited by available technology. We needed a system that could be placed in a natural reach of stream rather than one that relied on a structural component such as a bridge or culvert. Because we had specific information needs and constraints and because we had new technology available to us, we designed and tested our own customized stationary PIT-tag detection system. We believe our efforts and results have utility elsewhere and have aspects that should lead to further development for increased capabilities. Since our first placement of a system in Rattlesnake Creek in August 2001, we have actively replaced or modified the transceiver, power, antenna, and data storage subsystems. The objectives of this paper are to describe and discuss our design, lessons learned, limitations, and successes with an in-stream PIT-tag detector system.

2 Study Area

The Rattlesnake Creek watershed is located in the northwestern USA in the south central part of the State of Washington. The watershed covers 143 km^2 and supports a third-order stream system. Rattlesnake Creek enters the White Salmon River at river kilometer (rkm) 13.8, which in turn enters the Columbia River at rkm 271. Elevations range from 114 m at the mouth of Rattlesnake Creek, which is at the watershed's western boundary, to 927 m at ridge tops near its eastern edge. The watershed's climate is temperate with 80% of the annual precipitation occurring between October and March.

3 Methods and Results

In August 2001, we installed our first detection system near the mouth of Rattlesnake Creek (rkm 0.2). Grid electrical power was provided to operate the PIT-tag interrogation system. Fiber optic cables were installed and connected to a computer with software capable of recording PIT-tag information. The onsite data storage system was located about 100 m from the transceiver(s). The software we used was "Minimon", which is public domain and available through Pacific States Marine Fisheries Commission (PSMFC; Portland, Oregon, USA). This program automatically transfers input data in the proper format to be uploaded to a regional data base (PTAGIS), which is maintained by PMSFC. In addition, the program provided diagnostics that were accessible onsite to assess how well the detector system was performing. With the later installation of a wireless phone and internet connection, the data were automatically transferred to the regional database. To download information on detections of PIT tagged fish, we queried the regional database at regular intervals.

Prior to installation, much attention was given to designing antennas that afforded as large a read distance as possible for a 12-mm PIT tag. The antenna wiring and capacitors were housed in 10-cm diameter PVC pipe arranged in rectangular shapes with internal open areas measuring 200-300 cm in length and 80-100 cm in width. Read distances for 12-mm PIT tags were variable between antennas and initially affected by electronic noise and transceiver tuning. In general, we were able to get read distances in the 10-25 cm range for optimally oriented PIT tags, which was considered acceptable and which allowed any tag passing through the antenna opening to be read. We were rarely able to exceed this range at the Rattlesnake Creek site likely because of Radio Frequency Interference (RFI) noise often exceeding 20%. At a second site in northern Washington, where RFI noise seldom exceeded 10%, we were able to regularly achieve read distances as high as 40 cm.

In 2001, direction of fish movement was determined by anchoring single antennas in the stream's thalweg about 15 m apart (Figure 1). Antennas were placed in various orientations. The upstream unit was initially placed perpendicular to the stream bed, presenting itself much like a window, and thus we call it a "pass through" design. In this orientation, only the bottom edge was secured to the substrate. Because this orientation proved vulnerable to being displaced by stream velocities and debris, we subsequently introduced a "hybrid" design. Use of anchored pivots allowed this otherwise floatable antenna to buoy itself with rising and falling water levels. The dowstream unit was laid flat, which we call a "pass by" design. To anchor antennas, swedge bolts with an attached ring were installed in large and heavy pieces of substrate to serve as tie-off points for rope that was wrapped around the corners of antenna. All antennas were provided with break-away, waterproof connectors to prevent damage or loss of the transceiver(s) in case the anchoring gave way, which proved to be warranted a couple of times during the study.

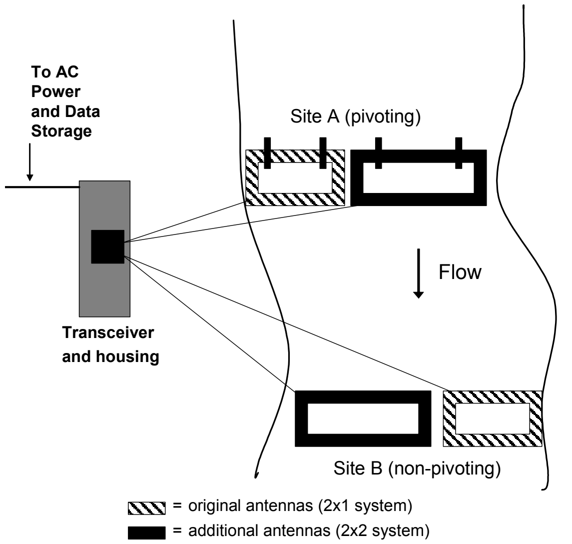


Figure 1. *Diagram of the in-stream PIT-tag detector system.*

The transceivers and data storage devices were located in weatherproof housing. We found that use of direct AC power caused high interference for the transceivers, but that DC had much less interference. The AC power for the transceivers power was inverted to DC, which provided 24 V to the transceivers.

Initially each antenna was connected to a separate transceiver, FS 1001-A prototype models from Digital Angel Corporation. With subsequent availability of a multiplexing prototype transceiver (FS 1001-M, auto tune, from Digital Angel Corporation), we were able to substitute one transceiver in place of two and able to add more antennas without much additional cost. A second 2-4 m long antenna placed laterally to an existing antenna enabled covering most the stream channel width at most flows. We have subsequently added a third pair of antennas, but sufficient data have not been collected yet on this configuration.

From the data collected, we derived estimates of overall detection efficiency and the error associated with the estimate. Fish-passage data used for efficiency estimates had to meet certain qualifications to be considered. We attempted to use those incidents where a fish actively migrated past the arrays. We made the assumption that fish used for efficiency estimates passed both antenna arrays even if not read on both. We only used fish that were tagged and released at least 50 m from the detector system and thus were not resident in the habitat immediately adjoining the site. If a fish was read on both antenna arrays, the time between reads could not exceed 18 minutes. We did not use fish that exhibited milling behavior (i.e., getting detected many times on one or both antenna arrays), nor did we use fish that passed and quickly returned. More than one passage incident per individual fish was used if the migrations happened at different life-history stages (e.g., juvenile outmigrant and spawning adult). The equations we used for these estimates were adapted from Skalski et al. [13], who developed the estimators for radio-tagged fish passing dams on the Columbia River. We used the USER program (v. 2.1) [6] to generate antenna array efficiencies, antenna array standard errors, and site efficiencies. The site standard error was calculated using the delta method in Seber [12].

In the first year of operation with our 2x1 design, the detection efficiency of downstream-moving fish during low flows (<14 cfs) was 94% (SE, 4.2; n, 18), but was reduced to 69% (SE, 19.8; n, 15) during high flows. By adding antennas lateral to the original antennas (2x2 design; Figure 1) and with the concomitant availability and use of a more powerful generation of 12 mm PIT tags (TX1400BE to TX1400ST; from Biomark, Inc.), the detection efficiency of downstream-moving fish was increased to over 98% at all flows (SE, 0.4; n, 173). Detection efficiency of upstream-moving fish was relatively high in both the 2x1 (88%; SE, 8.8; n, 17) and 2x2 (98%; SE, 2.4; n, 14) system during all flows.

The detector system had a date and time stamp feature, which recorded the time of detection to the nearest second. These data were imperative in order to determine direction of fish movement and detection efficiency.

4 Discussion

Our PIT-tag detector system was designed for a free-flowing stream in hope that we could monitor fish movement throughout all seasons of the year. A major desire was to have a system that enabled us to determine the direction of fish movement (upstream or downstream). Many improvements to the system were made to the system since its deployment in 2001. We rapidly infused new technology into the system, including change in prototype transceivers, upgrades to improve data storage and transmission, and use of improved PIT tags with better read distances. We added antennas to change from a 2x1 to a 2x2 design, and we changed the orientation and anchoring of the antennas. We are currently testing a 3x2 design (i.e., 3 pairs of antennas), but enough data have not been secured for reporting these results. The 2x2 system proved to be highly efficient for reading PIT-tagged fish in both the upstream and downstream direction, and it proved capable of continuous performance through a range of sometimes challenging seasonal conditions.

Because of confounding factors of site differences, flow conditions, and lack of controls, we could not ascribe the incremental change in improved detection capability to any single improvement made. However, adding antennas to cover more of the stream width appeared to have a definitive benefit. There is some possibility of simply increasing the length of antennas to cover more of the stream width, but because of power and performance limitations, we consider it unlikely that antennas could have been built much longer than the 3 m that we constructed.

The technology for detection of PIT-tagged fish is rapidly evolving. The frontier for PIT tags is for smaller tags with increased read distances. The frontier for detector systems is for development of reliable systems for application in larger streams. One of the many challenges is the ability to provide power with reduced labor to these systems in remote locations.

With collection of stream habitat variables, movement information can be readily related to life history events such as spawning migrations and to changes in environmental conditions such as stream flow and temperature. Integrating information from our detectors in tributary streams with that from detectors downstream at dams on the Columbia River has promise to be a powerful tool for monitoring movement patterns of anadromous and other migratory fish species, which will serve to increase our understanding of complex fish behavior and habitat use.

A number of people helped with this work. Steve Stampfli allowed use of his property for the system. Sandra Downing, Bruce Jonasson, and Ed Nunnallee of the National Marine Fisheries Service were instrumental with the hardware and software installation and troubleshooting. Brady Allen, Kyle Martens, Jodi Charrier, and Chris Schafer of the U.S. Geological Survey provided major help with fish collection and tagging. The Pacific States Marine Fisheries Service provided help with the electronic transmission of data. Bonneville Power Administration provided the major funding for the project.

Use of trade names does not imply endorsement by the U.S. government.

References

1. Achord, S.; Matthews, G.M.; Johnson, O.W.; Marsh, D.W. (1996). Use of passive integrated transponder (PIT) tags to monitor migration timing of Snake River chinook salmon smolts. *North American Journal of Fisheries Management* **16**:302-313.
2. Armstrong, J.D.; Braithwaite, V.A.; Rycroft, P. (1996). A flat-bed passive integrated transponder antenna array for monitoring behavior of Atlantic salmon parr and other fish. *Journal of Fish Biology* **48**:539-541.
3. Brannas, E.; Lundqvist; H.; Prentice, E.; Schmitz, M.; Brannas, K.; Wiklund, B. (1994). Use of the passive integrated transponder (PIT) in a fish identification and monitoring system for fish behavioral studies. *Transactions of the American Fisheries Society* **123**:395-401
4. Greenberg, L.A.; Steinwall, T.; Persson, H. (2001). Effect of depth and substrate on use of stream pools by brown trout. *Transactions of the American Fisheries Society* **130**:699-705
5. Juanes, F.; Letcher, B.H.; Gries, G. (2000). Ecology of stream fish: insights gained from an individual-based approach to juvenile Atlantic salmon. *Ecology of Freshwater Fish* **9**:65-73.
6. Lady, J.M.; Skalski, J.R. (2003). USER 2.1, User specified estimation routine. Project No. 8910700, Contract No. DE-B179-90BP02341. Prepared for Bonneville Power Administration, Portland, Oregon, USA.

7. Muir, M.D.; Smith, S.G.; Williams, J.G.; Sandford, B.P. (2001a). Survival of juvenile salmonids passing through bypass systems, turbines and spillways with and without flow deflectors at Snake River dams. *North American Journal of Fisheries Management* **21**:135-146.
8. Muir, M.D.; Smith, S.G.; Williams, J.G.; E.E. Hockersmith. (2001b). Survival estimates for migrant yearling spring chinook salmon and steelhead tagged with passive integrated transponders in the lower Snake and lower Columbia rivers, 1993-1998. *North American Journal of Fisheries Management* **21**:269-282.
9. Ombredane D.; Baglinière, J.L.; F. Marchand. (1998). The effects of passive integrated transponder tags on survival and growth of juvenile brown trout (*Salmo trutta* L.) and their use for studying movement in a small river. *Hydrobiologia* **371/372**:99-106.
10. Prentice, E.F.; Park, D.L.; Flagg, T.A.; McCutcheon, S. (1986). A study to determine the biological feasibility of a new fish tagging system, 1985-1986. Project No. 83-19, Contract No. DE-A179-83BP11982. Prepared for Bonneville Power Administration, Portland, Oregon.
11. Prentice, E.F.; Flagg, T.A.; McCutcheon, C.S. (1990). Feasibility of using implantable passive integrated transponder (PIT) tags in salmonids. *American Fisheries Society Symposium* **7**:317-322.
12. Seber, G.A.F. (1982). The estimation of animal abundance and related parameters. Macmillan, New York.
13. Skalski, J.R., Smith, S.G.; Iwamoto, R.N.; Williams, J.G.; Hoffman, A. (1998). Use of passive integrated transponder tags to estimate survival of migrant juvenile salmonids in the Snake and Columbia rivers. *Canadian Journal of Fisheries and Aquatic Sciences* **55**:1484-1493.
14. Skalski, J.R.; Townsend, R.; Lady, J.; Giorgi, A.E.; Stevenson, J.R.; McDonald, R.D. (2002). Estimating route-specific passage and survival probabilities at a hydroelectric project from smolt radiotelemetry studies. *Canadian Journal of Fisheries and Aquatic Sciences* **59**:1385-1393.
15. Zydlewski, G.B.; Haro, A.; Whalen, K.G.; McCormick, S.D. (2001). Performance of stationary and portable passive transponder detections systems for monitoring of fish movements. *Journal of Fish Biology* **58**:1471-1475.

Stochastic catastrophe analysis of effects of density and group size on behavior: an example in laying hens

P. Koene
Ethology Group, Department of Animal Sciences, Wageningen University, the Netherlands

Abstract

A nonlinear method of analyzing behavioral data is described using stochastic catastrophe analysis. This method is applied when bimodality, sudden transitions, hysteresis, inaccessibility and/or divergence are found in data normally analyzed using multiple or logistics regression. This paper shows that the effects of varying stocking density and group size on vocal behavior of laying hens were better described by a stochastic cusp catastrophe model than by regression analysis.

Keywords

Catastrophe Theory; cusp surface analysis; laying hen behavior; density; group size

1 Introduction

When an individual animal is simultaneously attracted and repelled by a goal or situation, the individual is said to be in an approach-avoidance conflict. Several models have been postulated to describe the dynamics of this type of conflict and its resolution. In a social situation individuals are attracted to each other for cooperation or safety and simultaneously repelled by competition. Our purpose was to try to describe this conflict in a catastrophe model (Thom, 1975).

2 Catastrophe Theory

Catastrophe Theory is a mathematical theory created by René Thom of the Institute of Advanced Sciences in France. The theory describes changes in nature. Science often studies gradual changes in variables. However, sometimes changes in nature happen swiftly rather than smoothly; these changes are not gradual but abrupt and discontinuous. Catastrophe Theory is able to describe these sudden changes mathematically; it deals with the stability and transformation of forms in space.

Only a few successful applications of Catastrophe Theory have been reported in the behavioral sciences. Baker and Frey (1980) used Catastrophe Theory to relate associative strength of conditioned stimuli to response probabilities. They showed that there is not a simple curvilinear relationship between association strength and response probability.

In another study using stochastic catastrophe analysis, Stewart and Peregoy (1983) asked their subjects to judge whether a drawing represented a man or woman. These authors presented different drawings in regular or random changing sequences. The authors analyzed the data using Cobb's program (1981) to estimate the fit of catastrophe models to their empirical data. The subjects showed a sudden transition in estimation from woman to man and from man to woman. The cusp catastrophe model – with two control and one behavior parameter - produced a better fit than classical linear regression models.

Energy in space based on two control dimensions is mathematically described by the potential function V(x) given by

$$V(x) = x^4/4 - bx^2/2 - ax + c$$

Stable points in this three dimensional space are found by setting the first derivative of V(x), designated by V'(x), equal to zero:

$$V'(X) = x^3 - bx - a = 0$$

The function of the cusp catastrophe describes the stable cusp surface as an elementary catastrophe with a folded behavioral surface and control parameters a and b (called the normal and the splitting factor, Figure 1). The preceding equation describes a cusp catastrophe, which is called this way because the projection of the behavioral surface on the control surface has a cusp like appearance (Figure 1).

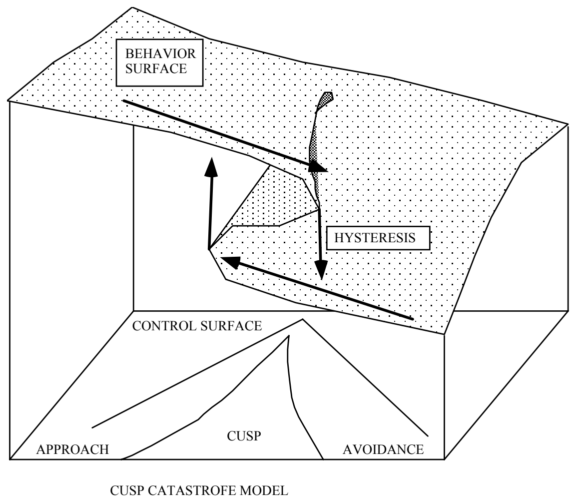


Figure 1. *Cusp catastrophe of approach-avoidance conflict (for further explanation see section 3). The control parameters are here not the normal factor and the splitting factor b, but the conflicting factors approach and avoidance). In the cusp part of the model two behavioral states are found for every combination of approach and avoidance parameters.*

The characteristics of the Cusp Catastrophe are:

- Bimodality: The behavioral surface is folded in such a way that two behavioral states are possible.
- Sudden transitions: A gradual increase of a control parameter can cause a sudden change in behavior; a catastrophic jump or catastrophe from one surface to another may occur.
- Hysteresis: The transition of behavior from the upper surface of the folded area into behavior of the lower surface takes place at another combination of control parameters than the transition of behavior from the lower surface into the upper surface.
- Inaccessibility: The behavior is unstable between the upper and lower surface.
- Divergence: A gradual increase in the factor causes an increasing difference in behavior when the cusp is entered.

3 Cusp model of conflict behaviour

In rats we showed that individual non-social conflicts result in either stable approach or stable avoidance behavior, i.e. bimodality and bistability in behavior (Koene and Vossen, 1994). In the next paragraph the experiment is shortly explained.

Data were gathered in Skinner boxes, in which a light of varying intensity was associated with food and a tone of varying intensity was associated with shock; the higher intensity was associated with more food, or higher shock. In the test phase rats had to press a lever for 5 seconds to acquire the reward. The criterion for reward was a higher light intensity than tone intensity, after which food was delivered. If the tone intensity was higher than the light intensity, shocks were given. The response of the animals was analyzed using the cusp fitting program of Cobb (1981). The resulting cusp surface formula resulted in a model that described short responses when the tone intensity was much larger than the light intensity (see fig. 1), and long responses when the light intensity was much larger than the tone intensity. In case light and tone intensity were equal or slightly different responses could be either short or long. Log linear analysis showed that the experience (either a reward or a punishment before the actual response) had a significant effect on the response duration. For individuals or groups the cusp model significantly explained the variation in responses better than a multiple regression model.

In the example of the approach avoidance cusp, the control parameters correspond to the two conflicting control dimensions 'approach' and 'avoidance,' respectively (Figure 1). The behavioral state value of a subject is described by a point in the folded behavior surface, dependent on approach and avoidance values in the control surface. In the folded part of the behavior surface two behavior states are possible (minimum of V'(x)), with an intermediate inaccessible region (maximum of V'(x)). The advantage of the cusp model is that it allows describing the bimodal response behavior of rats in an approach-avoidance situation and that it incorporates the effects of the preceding reinforcement history.

4 Laying hen vocal behaviour

Laying hens in non-cage systems are mostly housed in large groups and in high densities. Both affect the behavior of the laying hen. For instance, a high density may increase stress by decreasing the individual space of a hen. The high group size may increase safety. In literature, both effects are often difficult to separate. High density (# hens/m2) is considered to restrict behavior and reduce welfare of laying hens (Carmichael et al., 1999). Group size is also known to influence laying hen behavior (Nicol et al. 1999).

When density and/or group size changes, the animals will show different behavioral responses. Most obvious, interactions, such as pecking will change. Furthermore, vocal expressions may change too. Vocalizations may serve as indicators of a laying hen's welfare (Zimmerman and Koene, 1998; Manteuffel et al., 2004). Vocalizations as expressions of positive and negative experiences may be an indication of the welfare of animals at different stocking densities and group sizes. Furthermore, measuring vocalizations is non-invasive, which makes it an even more interesting tool for the estimation of welfare.

4.1 Methods

The effects of area size and group size on behavior of laying hens were measured. Sub samples of 40 Lohman Brown laying hens were tested in 25 crossed combinations (Figure 2) of group size (2, 4, 6, 8, and 10 birds) and density (1, 3.2, 10, 17.8 and 31.6 birds/m^2).

Group size / Log (density)	**2 hens**	**4 hens**	**6 hens**	**8 hens**	**10 hens**
0 (1 hen/m^2)					
0.5 (3.2 hens/m^2)					
1 (10 hens/m^2)					
1.25 (17.8 hens/m^2)					
1.5 (31.6 hens/m^2)					

Figure 2. *Experimental design of effect of density and group size on laying hen behavior.*

Vocalizations and interactive behaviors were recorded using behavior sampling on a Psion Organiser II hand-held computer with Observer software (Noldus, 1993).

The interactive behavior ethogram was composed during weeks of pilot measurements and included detailed descriptions of vocalizations (with sonograms) and other behaviors that were considered interactive.

As different names are given in literature to certain known vocalizations of chicken species (Collias, 1987), vocalizations in the current research are referred to with a number, to which in the ethogram structural definitions are attached.

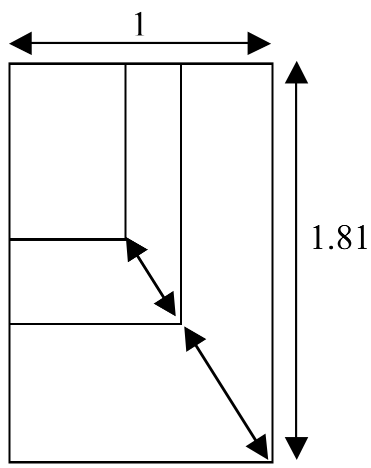


Figure 3. *Changing the size of the area leaving the shape intact. The two changing walls are made of plastic tubes in aluminium reels.*

For acquiring the grid of 5 x 5 combinations of density and group size, 25 different area sizes had to be designed that could be changed rapidly, keeping the shape of the rectangled area intact (Figures 3, 4 and 5).

In each of the 25 density/group size combinations a replication observation was done.

Figure 4. *Density 1 m²/hen, group size 4 hens.*

Figure 5. *Density 17.8 m²/hen, group size 4 hens*

4.2 Analysis and results

Multiple regression analysis showed that there were strong significant relations of vocal and overt behavioral elements with stocking density and group size. For example, an increase in group size led to less vocalizations per hen ($P<0.004$), while an increase in density resulted in more vocalizations per hen ($P<0.001$). Further results of the multiple regression analysis will be published elsewhere (Koene et al., in prep).

Table 1. *Data reduction by factor analysis. Factor scores for 25 combinations of density and group size are given.*

LOGDEN	GS	Voc1	voc2	voc3	beh1	beh2	beh3
0.00	2	-0.10985	-0.62516	-0.10593	-0.10600	-0.72592	0.31166
0.00	4	-0.61580	-0.98500	1.19616	0.30193	-0.91115	-0.72464
0.00	6	-0.88735	-0.86627	-0.34458	-0.30007	-0.91107	-0.34516
0.00	8	0.40976	-0.09537	1.54164	0.85194	-0.63465	-0.99128
0.00	10	-0.90555	-0.91360	0.36319	-0.27727	-0.63096	-0.79893
0.50	2	-1.41982	0.66458	-0.47265	0.97193	-0.32206	0.33297
0.50	4	-0.52148	2.74983	0.91340	4.20029	1.18004	1.43530
0.50	6	-0.82235	-1.15516	-0.43401	0.34047	0.01825	-0.84800
0.50	8	0.28635	-0.37828	0.50536	0.75947	-0.33840	0.27616
0.50	10	-0.48006	-0.28378	0.01425	0.04965	-0.83336	-0.48406
1.00	2	-0.22941	-0.94252	-0.15900	-0.86892	-0.74116	0.80697
1.00	4	2.68502	-0.60301	-1.57279	-0.81256	-0.32500	-0.33459
1.00	6	-0.40918	0.05899	-0.54033	-0.36610	-0.25267	0.69800
1.00	8	1.11286	0.41434	0.19354	0.68996	0.77551	0.10176
1.00	10	0.76792	-1.25934	1.83129	-0.29123	1.92759	-0.76284
1.25	2	3.03100	-1.26458	-0.12099	-0.90020	-0.27306	-0.08238
1.25	4	-1.05258	-0.76655	-0.81833	-0.81741	-0.05213	-0.65256
1.25	6	0.34317	-0.05979	-0.02194	-0.88876	1.03957	-0.15071
1.25	8	0.11020	2.04819	0.38632	0.57789	1.81205	0.68677
1.25	10	3.09874	0.54100	0.70285	-0.40926	3.62050	-0.32723
1.50	2	-0.93260	-0.72877	-0.34103	-0.97038	-0.03944	-0.48620
1.50	4	0.86759	0.02436	-1.15824	-0.63147	0.81420	-0.32342
1.50	6	0.06621	1.39374	0.74535	0.06432	0.84555	0.55846
1.50	8	0.02238	2.83546	-1.13649	0.79755	1.59695	0.51661
1.50	10	-0.61606	-0.37893	-0.68574	-0.67313	0.43397	-0.78024

In the 50 situations the interactive behaviors and the vocalizations were factor analyzed for data reduction (Factor analysis with varimax rotation). Factor analysis of the vocal behavioral elements revealed 3 vocal and 3 behavioral factors with an Eigenvalue above 1 (Table 1). The analysis was here continued with the scores on those 6 factors. The Cuspfit programme gave visual output of the location of the data in the control space (Figure 6). Each of the resulting factors (Eigenvalue above 1) with its 50 factor scores was used as input for the revised Cobb (1981) programme called Cuspfit that is available at http://users.fmg.uva.nl/hvandermaas/. The output of this programme was combined for comparison of the fits of the 6 factors. In the example of the first vocal factor a unimodal and a bimodal area in control space and hence behavior surface is shown (Figure 6).

```
location of data in the control space:

     -5    -4    -3    -2    -1     0     1     2     3     4     5

5.0   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0   5.0
            *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
4.5   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0   4.5
               *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
4.0   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0   4.0
                  *  *  *  *  *  *  *  *  *  *  *  *  *
3.5   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0   3.5
                     *  *  *  *  *  *  *  *  *  *  *
3.0   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0   3.0
                        *  *  *  *  *  *  *  *  *
2.5   0  0  0  0  0  0  0  4  4  4  0  0  0  0  0  0  0  0  0  0  0   2.5
                           *  *  *  *  *  *  *
2.0   0  0  0  0  0  0  0  4  6  4  0  0  0  0  0  0  0  0  0  0  0   2.0
                              *  *  *  *  *
1.5   0  0  0  0  0  0  2  6  8  2  0  0  0  0  0  0  0  0  0  0  0   1.5
                                 *  *  *
1.0   0  0  0  0  0  0  0  2  4  0  0  0  0  0  0  0  0  0  0  0  0   1.0
                                 *  *  *
0.5   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0   0.5
                                    *
0.0   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0   0.0
```

Figure 6. *Cusp surface analysis showing the cusp area (*) and the data of factor vocal1 partly inside and partly outside of the cusp area.*

The statistical output shows that for example for the first vocal factor (Table 2) the cusp analysis explains significant more variance ($R2 = 0.40$) than the multiple regression analysis ($R2= 0.08$).

Table 2. *Output of the Cuspfit program, comparing linear, logistic and cusp fit.*

```
****************************** STATISTICS ******************************

   MODEL          R^2        LOGLIKELIHOOD    #PAR         AIC          BIC
 -------------------------------------------------------------------------------
   Linear        0.0789      -0.7300E+02       4      0.1540E+03   0.1616E+03
   Cusp          0.4013      -0.5571E+02       8      0.1274E+03   0.1427E+03
   Logistic      0.1524      -0.6681E+02       5      0.1436E+03   0.1532E+03
 -------------------------------------------------------------------------------
 chi^2 test (a=0.05): linear vs cusp ===> cusp
 AIC: ==>cusp
 BIC: ==>cusp
```

Analysis of all 6 factors - using stochastic cusp catastrophe analysis - shows that in 4 of the 6 cases the explained variance by a cusp model was significantly higher than using linear multiple regression. In 2 other cases the logistic model explained more variance than the linear model.

4.3 Discussion

Vocal and overt behavioral factor scores in laying hens were better described by a stochastic cusp model than by a multiple regression model using density of the hens and group size of the hens as independent factors. This implies that part of the data shows linear relations and another part shows non-linear relations between vocalizations and density and group size, i.e. showing bimodality in behavioral states. That means that in situations of identical parameters of group size and density two opposite behavioral responses are found. According to Catastrophe Theory these different reactions of the hens are caused by their different motivational states and/or genetic make-up or by their preceding experience. In this experiment hens were assigned to a condition at random. It is therefore unknown what the previous experience of every individual hen exactly was. For this purpose experiments will be planned focusing on individual vocal and overt behavior of hens in a group, and not group behavior of the hens only. In general, research in behavior and - more specific - abnormal behavior in which different types of reactions upon stressful stimuli are found is a promising area for application of (statistical) Catastrophe Theory.

5 Conclusion

There is a future for stochastic cusp catastrophe analysis in the behavioral sciences when data show the occurrence of at least one of the five characteristics (divergence, catastrophic jumps, hysteresis, bimodality and inaccessibility) of cusp catastrophe. In order to facilitate this, special attention for individual behavior and distributions of individual data is necessary. Opposite types of behavioral reactions are expected to emerge in challenging situations such as approach-avoidance conflicts (rats) and social crowding with its attractive and repellent forces (laying hens). Applications of catastrophe models in such contexts show that such models have significantly more explanatory power than regression analysis, and are relevant contributions to the behavioral sciences.

Hanneke Nijland and Raymond de Heer are greatly acknowledged for their assistance in the laying hen experiment.

References:

1. Baker, J.S. and P.W. Frey (1980). A cusp catastrophe: hysteresis, bimodality and inaccessibility in rabbit eyelid conditioning. *Learning and Motivation*, **10,** 520-535.
2. Carmichael, N.L., A.W. Walker and B.O. Hughes (1999) Laying hens in large flocks in a perchery system: influence of stocking density on location, use of resources and behaviour. *British Poultry Science*, **40**, 165-176.
3. Cobb, L. (1981). Parameter estimation for the cusp catastrophe model. *Behavioral Science*, **26**, 75-78.
4. Collias, N.E. (1987).The vocal repertoire of the Red Junglefowl: a spectrographic classification and the code of communication *The Condor*, **89**, 510-524.
5. Koene, P. and J. M. H. Vossen (1994). A catastrophe model of approach-avoidance conflict. In: *Viability of mathematical models in the social sciences*.
6. Manteuffel, G., B. Puppe and P. Schön (2004). Vocalization of farm animals as a measure of welfare. *Applied Animal Behaviour Science*, **88**, 163-182.
7. Nicol, C.J., N.G. Gregory, T.G. Knowles, I.D. Parkman and L.J. Wilkins (1999). Differential effects of increased stocking density, mediated by increased flock size, on feather pecking and aggression in laying hens. *Applied Animal Behaviour Science*, **65**, 137-152.
8. Noldus (1993). The Observer. Noldus Information Technology, Wageningen, the Netherlands.
9. Stewart, N. and P.L. Peregoy (1983). Catastrophe Theory modeling in psychology. *Psychological Bulletin*, **94**, 336-362.
10. Thom, R. (1975). *Structural stability and morphogenesis*. Benjamin Addison Wesley, New York.
11. Zimmerman, P.H., and P. Koene (1998). The effect of frustrative nonreward on vocalisations and behaviours in the laying hen (*Gallus gallus domesticus*) *Behavioural Processes*, **44**, 73-79.

Use of ActiTrac™ in the training of rescue dogs

C. Wust, M. Schneider, M. Erhard
Institute of Animal Welfare, Ethology and Animal Hygiene, Ludwigs-Maximilians University, Munich, Germany

Abstract

Rescue dogs are employed in difficult, rough terrain (rubble search) and on spacious, broken areas (area search). Uninterrupted observation and analysis of the search activity of rescue dogs via video recording is therefore not possible. The objective of this study was to test and establish ActiTrac™ for recording motion in the field of working with rescue dogs. The physical activity of rescue dogs was recorded in combination with continuous heart rate measurement during various search operations, and the stress on the dogs during the search was determined.

The actometer ActiTrac™ by Somnomedics is a miniature motion monitor for the continuous registration of motoric activity and is based on acceleration sensors referred to as piezo sensors. Representation and analysis of the results of the ambulatory monitors ActiTrac™ and Polar™ can be synchronized. This reveals the regulation of the cardiovascular system upon physical exercise and permits a distinction between physical and psychological stress. In addition, exhaustion of the rescue dog can be demonstrated objectively based on a decreasing motion activity during the search operation.

Keywords

Actometer, continuous measurement of activity, heart rate, rescue dog, stress

1 Introduction

In human medicine actometers have been used for years in a variety of fields. The actometer ActiTrac™ by the German company Somnomedics was originally designed for the monitoring and documentation of training programs of athletes, for the documentation of rehabilitation programs, and for assessing the progress of rehabilitation on the basis of improved motoric activity. In the field of human medicine, it was for instance used successfully for motion analysis with Parkinson´s disease (Csoti et al. 2001). The objective of this study was to test and establish ActiTrac™ for recording motion in the field of working with rescue dogs. By combining ActiTrac™ with continuous heart rate recording using Accurex Plus™ by the Finnish company Polar conclusions could be drawn as to the stress on rescue dogs during search operations.

2 Material and method

The Somnomedics ActiTrac™ actometer is a miniature motion monitor for the continuous registration of motoric activity. It is based on acceleration sensors referred to as piezo sensors, which play an important role in the field of measuring technology as electromechanical energy converters. The sensor records motion in biaxial direction with a sensitivity of up to 0.312 g. The activity data are recorded with a scanning rate of 40/sec. The individually adjustable storage rate is used to determine the point in time to calculate and digitally integrate a mean value of the continuously registered acceleration signals. Depending on the epoch length, ranging from 2 sec to 2 min, the recording capacity is between 27 h and 68 d. The activity data are accumulated over the period of an epoch and stored in a permanent memory of 48 kb.

The ActiTrac™ has roughly the size and weight of a sports watch. Thanks to the integrated hardware and battery with a life of 2-3 years no patient cables are required. Thus, the instrument can be worn without impeding the range of motion in any way.

Polar´s Accurex Plus™ sports watch continually records the heart rate. The readings are sent to the transmitter via the electrodes in the chest belt. The transmitter will pass the signals on to the receiver, i.e. the wrist watch, which stores the heart rates in beats/min.

3 Implementation

19 rescue dogs of various age and from different rescue dogs units participated in this study. Each dog performed five operations (three rubble and two area searches) with identical test processes. Each search comprised 2 passes of 20 minutes each, with a 20-minute break in between. This was followed by a resting phase of 60 minutes. An individually adjustable collar which contained the actometer was placed around the neck of the dogs to be tested. The collar was tightened closely in order to protect the activity values from being compromised by the instrument´s own motion. The dogs wore these special designs just like normal collars; no impairment of the search activities by wearing the collars was observed. The beginning of the motion recording of ActiTrac™ was set using a setup program on the day preceding measurement. In view of the moisture sensitivity of the ActiTrac™ a rubber glove was pulled over it as a preventive measure to protect the instrument from any heavy precipitation.

What proved somewhat more difficult was the parallel heart rate measurement using Polar´s Accurex Plus™ sports watch. The older chest belt by Polar has rigid electrodes which fit tightly to the human chest. However, the rigid electrodes made it difficult to establish and maintain contact with the V-shaped chest of the dogs; contact with the body surface was continually interrupted due to the dog´s motion. Thanks to its flexible electrodes, the new elastic belt by Polar adapts much better to the dog´s chest. The rigid transmitter should come to rest between the two electrodes laterally to the *sternum*. However, even these new belts require placement of a second, flexibly adjustable neck belt in order to prevent the chest belt from slipping towards the *tail*. Rescue dogs are exposed to all kinds of weather conditions. Therefore, shearing off some fur for a better contact between the chest belt electrodes and the body surface is not desirable. Ultrasound transmission gel applied to the fur is well suited to establish contact between electrodes and skin. If the outdoor temperatures are very high, however, the gel will dry quickly, and transmission of the heart rate will be interrupted. The same holds true for heavy precipitation, which washes away the gel. Therefore regular control of the contact and re-application of gel as required is necessary.

The results of the ambulatory monitors ActiTrac™ and Accurex Plus™ can be synchronized and evaluated using a monitor coupling software (Version 2.3.0) by Somnomedics. The time recorded on the activity monitor and the heart rate watch must be synchronized with the PC used for analysis prior to the beginning of the test. Actigraphy provides a standardized activity output in gravitational acceleration (mG). The activity and heart rate measured are represented in parallel by the coupling software in the form of curves. It is possible to select time periods and allocate them to predefined descriptions. For the selected area mean value, standard deviations and extreme values of activity and heart rate are calculated. The resulting reports can be exported into other programs such as Excel for further statistical evaluation.

4 Results

ActiTrac™ can be used successfully for the recording of motion in the training of rescue dogs. Due to its lightness it does not impair the dog´s motion. The instrument is easy to handle. In addition, it can be programmed and evaluated independent of the time of testing. Transmission of the data saved to the PC is a matter of minutes. The ActiTrac™ permits the continuous registration of motion activity with a standardized activity output in mG. Analysis of the results is easy and takes relatively little time. Combination of the ActiTrac™ with Accurex Plus™ permits an illustration of the effects of motion activity on the cardiovascular system. In addition, it renders distinction between physical and psychological stress of the dogs possible to a certain extent. However, actigraphy alone does not suffice to draw any conclusions about the respective motion. High activity values, for instance, do not permit making a distinction between the dog jumping or shaking itself, or having its neck patted in praise by the dog trainer. Another major disadvantage is the high moisture sensitivity of ActiTrac™. Even a drizzle may be enough to damage the instrument.

5 Discussion

One major reason for employing rescue dogs is their quickness in searching spacious, difficult and dangerous terrain. This renders continuous video observation of the dogs during search operations impossible. Another difficulty involved in filming are poor weather conditions. The quality of the recordings decreases in the presence of precipitation, strong sunlight or low lux number, so that subsequent analysis is impeded.

One disadvantage of video monitoring as compared to ActiTrac™ is that it requires a cameraman who has to have the dog in his field of vision for the entire period of observation. Given the large areas to be searched and the high activity involved in the work of rescue dogs the fixed installation of a video system does not make sense. A cameraman, however, is an optical and olfactory disruptive factor for the rescue dog. ActiTrac™ on the other hand, due to its low weight and small size, will neither disturb the dog in its motion, nor will it distract the dog from its search operation. Another advantage of ActiTrac™ is that it is easy to handle. As programming and analysis can be performed independently of the test no further adjustment is required on site. This is a major advantage particularly in poor weather conditions as it is not necessary to have a laptop present on site. One huge disadvantage of ActiTrac™, however, it its moisture sensitivity. This means that ActiTrac™ has to be protected against moisture prior to each use in the open.

Another disadvantage of video recording as compared to ActiTrac™ is the requirement of video tapes and regular control of battery levels and remaining time on the tape during recording. Changing batteries or tapes is always associated with a loss of data. This is virtually ruled out with the ActiTrac™ due to the pre-programmable running time and the long battery life.

Another considerable advantage of ActiTrac™ has to do with data analysis. The rapid transmission of data to the PC via a serial interface yields instantly analyzable results. In addition, the standardized activity output in mG provides objective measuring values. By comparison, analysis of video material takes a lot of time and depends on the observer as well. This study alone would have involved analysis of 63 hours of video material for the search and 158 hours of video material for the resting phases of the rescue dogs. The coupling software also allows for the parallel calculation of specific sequences of the motion activity and the heart rate. This makes it possible to obtain an instant comparison of heart rate and activity, a quick overview of the curve progressions, and the interdependence of the two parameters. One disadvantage of ActiTrac™, however, is the fact that no conclusions can be drawn from the level of activity to the respective behaviour or the type of motion. It is necessary to record important pieces of behaviour by direct observation in order to be able to compare them to the respective activity and heart rate. Distinguishing between physical and psychological stress without direct observation is partially possible as a sudden increase in heart rate at a low activity level suggests emotional agitation. At a high activity level, on the other hand, psychological stress cannot be deduced from a high heart rate. However, ActiTrac™ is able to provide objective evidence of the dog´s exhaustion during a search operation. This study yielded a significant decrease in the level of activity between the first and the second pass of a search operation.

6 Summary

ActiTrac™ can be successfully used in the work with rescue dogs in order to obtain objective and continuous data on the motion activity of the dogs and analyze them within a relatively short period of time. However, ActiTrac™ can only help make a distinction in the level of activity; no conclusions can be drawn about the respective type of motion or the piece of behaviour. Combining the ActiTrac™ with continuous heart rate measurement makes it possible to observe the regulation of the cardiovascular system upon physical exercise, and to make a distinction between physical and psychological stress in certain situations.

Supported by Ernst Wippenbeck Foundation

References:

1. Csoti, I.; Werner, M.; Fornadi, F. (2001). ActiTrac-Bewegungsanalyse bei M. Parkinson. *Neurol Rehabil*, **7(3)**, 139-14

T-Pattern detection in physiological signals during sleep

G. K. Jonsson[1], H.G. Helgason[2], I. Hjálmarsson[2], M.S. Magnusson[3], S. Kristjánsson[2]
1 Human Behavior Laboratory, University of Iceland, Univerisy of Aberdeen & PatternVision
2 MedCare, Reykjavik, Iceland
3 Human Behavior Laboratory, University of Iceland & PatternVision

Abstract
Current analytical methods in sleep focus on counting specific events that are determined by predefined criteria and scoring rules. The main focus of the current study is to search for physiologically based sleep patterns that have not been defined in advance and to explore the clinical application of the findings. The investigation highlights the potential for T-pattern analysis to make a significant contribution to sleep research analysis. The data show that specific temporal patterns can be identified within scored sleep events. A high number of temporal patterns were detected in data from patient and healthy subject's polysomnography reports. Results indicate that the T-pattern detection algorithm discovered known sleep-patterns in manually scored apnea-patient records, sleep patterns that indicate some sort of sleep disorder. It also discovered unknown patterns that occurred exclusively in certain patient groups. Highly significant patterns were discovered in automatically scored data, based on new definitions, where certain pattern types also occurred exclusively in certain patient groups.

Keywords
T-patterns, Physiological Signals, Sleep research, Somnologica, Theme.

1 Introduction
Sleep recordings consist of various physiological signals, such as EEG, EMG, EKG, EOG, airflow, effort, SpO_2 and pulse. Current analytical methods in sleep focus on counting specific events that are determined by predefined criteria and scoring rules. In addition, specific events are counted, such as apneas, hypopneas, desaturations and snoring periods, as well as periodic limb movements to summarize the causes of sleep related problems. Indices are calculated from these events, diagnosis and treatment recommendations are given.
The main focus of our research is to search for physiologically based sleep patterns that have not been defined in advance and to explore the clinical application of the findings. We introduce a new approach, known as T-pattern detection, to the analysis of polysomnography data. Sleep research data was collected with Medcare Embla - Somnologica sleep diagnostic system.

Embla & Somnologica - These systems offer a high degree of flexibility and signal quality to meet the highest demands of clinical work and research. Somnologica Studio is the performance powerhouse behind the industry-leading Embla Recording Systems and is designed for both clinical and research applications. Somnologica Software provides solutions with a comprehensive array of tools that streamline workflow and improve patient data management.

Theme - Theme is built on a unique algorithm that searches for hidden repeated patterns in behaviour and interactions, based on a model of the temporal organization of behaviour. It considers both the order and the time distances between behavioural event types as well as hierarchical organization. The software behaves as an evolution program, in the sense that it detects complex patterns gradually as combinations of simpler ones and deals with combinatorial explosions through competition between patterns such that only the most complete patterns are retained (survive) while partial detections are discarded. Theme can detect complex repeated patterns that are hidden to observers and very hard or impossible to detect with other available methods. The software includes various tools for filtering and analysis of detected patterns on the basis of their frequency, complexity, structure, actor identity and behavioural content [1,2].

2 Method
250 healthy subjects and 10 subjects previously diagnosed with Obstructive Sleep Apnea (OSA) (all men, mean age 48.2 yrs (±6.2), who were on a waiting list for Continuous Positive Airway Pressure (CPAP) treatment were offered a full polysomnographic (PSG) overnight study at the National University Hospital. Subjects underwent one night of continuous polysomnographic recording in the sleep laboratory. Electrodes were placed according to the international 10-20 system for the electroencephalographic (EEG) montage as well as left and right electrooculogram (EOG), submental electromyogram (EMG) and electrocardiogram (ECG). Additional EMG leads were placed on left and right anterior tibialis to monitor movements and signs for Periodic Limb Movements (PLM). Respiratory parameters included nasal pressure transducer derived measurements with linearized airflow via a nasal cannula and an oronasal thermistor. Respiratory movements of chest and abdomen were measured with XactTrace™ [3] Respiratory Inductive Plethysmogram (RIP) belts. Oxygen saturation, pulse and pulse-waveform were recorded via a finger probe oximeter [4]. Snoring, body position and activity were recorded as well. Data was collected with the Embla N7000 amplifier and analyzed in Somnologica™ [3]. Thirty-second epochs from the EEG leads were used to visually score sleep stages according to established criteria [5]. Standard methods to score microarousals were used as well. Apnea was defined by absent inspiratory airflow for at least 10 seconds. Hypopnea was defined as a reduction in airflow signal by $\geq$ 50% from the level measured before the event lasting at least 10 seconds and subsequently followed by an oxygen desaturation of at least 3%.

Pattern Detection

Sleep records from manually scored data were exported for T-pattern detection with Theme. Scores from automatically scored data was also exported for further analysis.

3 Results and Discussion

A high number of temporal patterns were detected in the data from the patient's polysomnography reports. Results indicate that the T-pattern detection algorithm discovered known sleep-patterns in manually scored apnea-patient records. It also discovered pattern types that occurred exclusively in certain patient groups. Highly significant patterns were also discovered in automatically scored data, based on new definitions, where certain pattern types occurred also exclusively in certain patient groups. To enable the search for patterns occurring in certain groups, data files were sorted according to patient's score on several indexes. In Figure 1. a known sleep pattern, found in the group of patients scoring high on Body Mass Index, (BMI) is displayed as detected by the T-pattern algorithm. In this case, the patients are waking up due to a snoring period where the arousal (wake-up) is accompanied by increases in muscle tone and leg movements. These types of arousals are usually more forceful than those that do not have accompanying leg movements or increased muscle tone changes. This pattern also displays the known result of an hypopneic episode, where an hypopnea, accompanied by a desaturation of oxygen is terminated by an arousal (patient briefly wakes up). The pattern displayed in Figure 2., found exclusively in the group of participants under the mean age, accurately portrays a typical progression of sleep from an awake state to the lighter stages of sleep with the accompanying muscle tone relaxation. Figure 3. displays a large number of the same pattern found only for those scoring higher than 10 on the Apnea + Hypopnea Index. This pattern displays the regular progression of sleep from wake to a light sleep stage. The high frequency of this pattern indicates that these patients were repeatedly waking up and falling back to sleep into the lighter sleep stages. This corresponds to a higher degree of sleep fragmentation due to their high apnea/hypopnea indices.

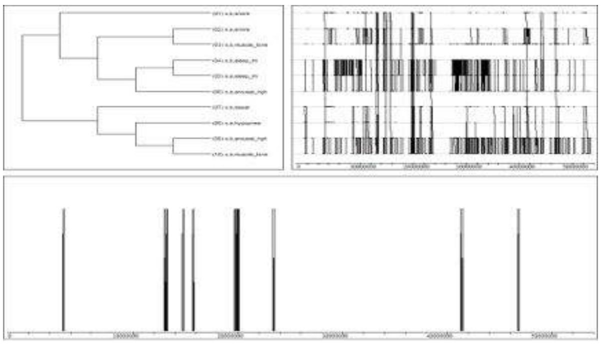


Figure 1. *Events and order: 1) x,b,snore 2) x,e,snore 3) x,b,muscle_tone 4)x,b,sleep_lm 5) x,e,sleep_lm 6) x,e,arousal_hgh 7) x,b,desat 8) x,e,hypopnea 9) x,b,arousal_hgh 10) x,e,muscle_tone.*

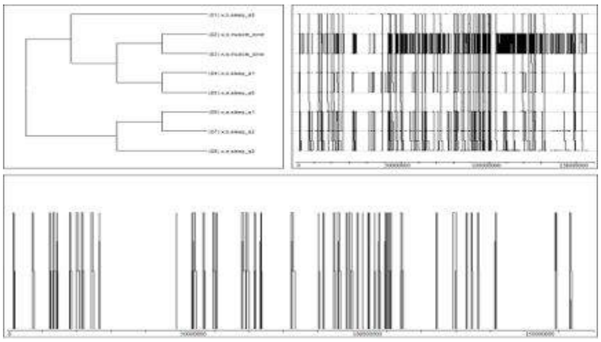


Figure 2. *Events and order: 1) x,b,sleep_s0 2) x,b,muscle_tone 3) x,e,muscle_tone 4) x,b,sleep_s1 5) x,e,sleep_s0 6) x,e,sleep_s1 7) x,b,sleep_s2 8) x,e,sleep_s2.*

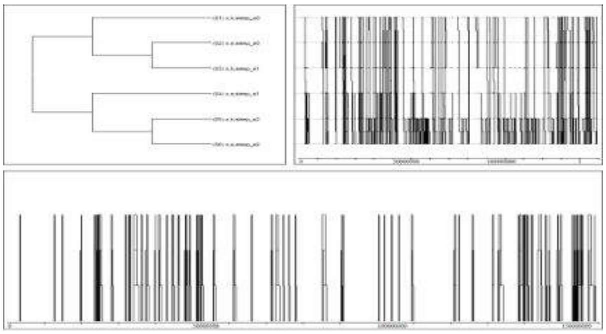


Figure 3. *Figure 3. Events and order: 1) x,b,sleep_s0 2) x,e,sleep_s0 3) x,b,sleep_s1 4) x,e,sleep_s1 5) x,b,sleep_s2 6) x,e,sleep_s2.*

4 Conclusions

The preliminary data shown in this paper highlights the potential for T-pattern analysis to make a significant contribution to sleep research analysis. Current analytical methods in sleep focus on counting specific events that are determined by predefined criteria and scoring rules that are based on the Rechtschaffen & Kales scoring system from 1968. In addition, specific events are counted, such as apneas, hypopneas, desaturations and snoring periods, as well as periodic limb movement (PLM) events to summarize the causes of sleep related problems.New analysis methods have recently been published in the field of sleep research. These methods are based on another set of predefined rules designed to analyze microstructural changes in sleep [6]. Cyclic Alternating Pattern (CAP) is an indication of unstable sleep and often gives a better indication of sleep fragmentation than the standard system from 1968. Preliminary investigations have reinforced the authors' belief that T-pattern identification has great potential as an effective research tool in sleep research. The primary reasons are that the T-pattern detection correctly identified predefined patterns that are used to analyze and score the sleep studies. In addition, it recognizes familiar sleep patterns that indicate some sort of sleep disorder. Most notably, that an arousal follows a severe desaturation, a pattern that repeatedly occurs with patients that suffer from severe sleep apnea syndrome.

This work was supported by a grant from The Icelandic Centre for Research (Grant N. 020360002)

References:

1. Magnusson, M.S. (2000) Discovering hidden time Patterns in Behavior: T-patterns and their detection. *Behavior Research Methods, Instruments & Computers*, **32**, 93-110.
2. Magnusson, M.S. (2004). Repeated Patterns in Behavior and Other Biological Phenomena. In Evolution of

Communication Systems: A Comparative Approach (Vienna Series in Theoretical Biology). D. Kimbrough Oller (Editor), Ulrike Griebel (Editor). The MIT Press. ISBN: 0262151111.
3. MedCare, Reykjavik, Iceland.
4. Nonin, Inc, Plymouth, Minnesota, USA.
5. Rechtschaffen, A & Kales, A (Eds.) (1968). A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. Los Angeles, UCLA Brain Information Service/Brain Research Institute.
6. Terzano, M.G. et.al. (2001). Atlas, Rules, and Recording Techniques for the Scoring of Cyclic Alternating Pattern (CAP) in Human Sleep. *Sleep Medicine*, **2**, 537-553.

Investigating self-hearing with psychoacoustic methods

K. Cole and L. Klein
Research Institute, Gallaudet University, Washington, D.C., USA

Abstract
A new method for measuring an individual's sensitivity to delay in speech feedback has been developed and evaluated. It involves automated sound manipulations and testing procedures, providing a reliable and efficient estimate of an individual's delay detection threshold during auditory self-monitoring. The approach is novel in the context of self-hearing although common in other psychoacoustic testing. It applies an adaptive up-down technique in a two-interval forced choice (2IFC) procedure. The procedure is implemented using the Capybara 320, a multi-processor sound computation engine which is controlled by the Kyma sound design language. The system's modularity allows simultaneous realization of complex spectral and temporal sound alterations and dynamic control of the testing procedure. This approach can be used for quantitative comparisons of the salience of speech feedback delay across different individuals and under different speaking/listening conditions. One such application of this approach is discussed.

Keywords
Self-hearing, speech feedback, delay detection, psychoacoustic methods, adaptive up-down procedure

1 Introduction
Self-hearing is essential for speech regulation and important to listening comfort while using hearing technologies, yet it is rarely considered in the design of these devices. While complex digital sound processing has improved auditory perception for hard of hearing individuals who rely on these technologies, it comes at a cost of increasingly longer processing time that may have significant consequences on self-hearing during speech.

There are two aspects of self-hearing that may set the limits on the acceptable amount of processing delay. One such aspect comes from the relationship between speech articulation and auditory feedback. When auditory feedback is delayed, it creates asychrony between motor commands and/or somatic feedback and the auditory feedback. Reseach has shown that asychrony exceeding 25-30 ms results in changes in speech production (Barac-Cikoja, 2004; Stone & Moore, 2002).

The other limitation on the acceptable amount of processing delay comes from the fact that the self-generated speech signal contains both air and bone-conducted components. When digital processing of the air component introduces a delay, asynchrony between the two components is created, potentially introducing listening discomfort (Stone & Moore, 1999). In addition to its effects on self-hearing, processing delays may also have consequences on speech comprehension. Processing delays introduce temporal discrepancies between the visual and auditory modalities which may make cross-modal integration of speech information difficult and consequently may reduce comprehension.

One key barrier to clinical consideration of listeners' sensitivity to delay in speech feedback is the lack of reliable and efficient assessment procedures. Investigations based on the ratings of subjective disturbance caused by feedback delay (Stone & Moore, 2002, 2003) are limited to a relatively small set of preselected delay values, and require substantial training of the subjects in order to achieve response consistency. Notably, critical delay values estimated using rating procedures represent an indirect assessment of the person's sensitivity to the delay in the speech feedback. While psychoacoustic methods have been used extensively to investigate listeners' sensitivity to temporal changes in *recorded* speech, there has been minimal effort to develop a similar approach in the domain of real-time self-hearing (Agnew & Thornton, 2000). The advantage of such an approach is that it allows efficient direct assessment of the limits on the person's ability to perceive experimentally-controlled changes in his/her speech feedback. Technological advances have made complex sound manipulations achievable in real-time. Specific temporal, spectral or intensity alterations of speech feedback can be implemented in a precise, reliable and efficient manner, enabling estimates of psychometric functions.

The method presented here is aimed at measuring an individual's sensitivity to delay in speech feedback based on his/her delay detection thresholds. The proposed method incorporates an adaptive up-down procedure that is simple to administer, has been proven highly efficient, relies only on the assumption of monoticity (i.e., that the probability of a correct response increases monotonically with stimulus level), and shows very good small-sample reliability (Levitt, 1971).

One of the goals in developing this approach is to enable assessment of how noticeable speech feedback delay is to different individuals in various noise backgrounds, using different hearing or telecommunication technologies. A study demonstrating one such application is presented here. In that study, sensitivity to feedback delay was investigated as a function of different signal processing strategies that may be found in some listening devices for hard of hearing individuals.

2 Method
The technique consists of adaptive up-down changes of the speech feedback delay in a two-interval forced choice (2IFC) procedure, until a minimal detectable delay is determined. Specifically, the subject is asked to repeat a syllable (e.g., "pa-pa-pa…") at a comfortable speaking rate for approximately 2 seconds and to listen for a delay (an echo) in his/her speech feedback. In each trial, the subject produces two such utterances, one of which is delayed. A delay is randomly assigned to one of the observation intervals (utterances).The subject's task is to identify which of the two utterances is delayed. Based on the subject's response accuracy, the size of the feedback delay is increased or decreased stepwise, making the task easier or harder, respectively. During the test, the step size

is progressively reduced as the delay approaches the threshold value. This results in increasingly fine adjustments in the magnitude of the delay. A delay detection threshold (DDT) is calculated from the delay values obtained on the last eight reversals (i.e., changes in the response accuracy, from correct to incorrect delay detection, or vice versa), using a step size of 1 ms. The rule applied to control changes in the delay (i.e., delay is decreased after two consecutive correct responses, and increased after a single incorrect response) renders threshold estimates that correspond to 70.7% point of the subject's psychometric function (Levitt, 1971).

To prevent the use of response strategies and to eliminate reliance on the acoustically most prominent first syllable, the initial 400-600 ms of the subject's utterance is muted. The feedback signal is then delivered for 1800 ms after which the remaining speech is muted again. Each interval (utterance) starts with a visual prompt. The signal processing is triggered by the subject's voice. Speaking at a normal rate, a subject typically hears 6-7 syllables in each interval. After completing both utterances, the subject is prompted to identify which one was delayed. No feedback on response accuracy is provided during testing. Testing time is less than 15 minutes.

2.1 Apparatus

A block diagram of the apparatus is shown in Figure 1. The subject speaks into a microphone (Sennheiser MKE2) and listens to his/her speech feedback via insert earphones (EAR Tone model 3A). The microphone is attached to the top of the subject's forehead, approximately 15 cm from his/her mouth. The microphone output is amplified (Mackie 1202 VLZPRO) and routed into the digital signal processor (Capybara 320, Symbolic Sound Co) for acoustic transformations. The output of the Capybara is amplified through an audiometer (GSI 10) to the subject's most comfortable level and presented binaurally. Microphone output and the audiometer output (i.e., the auditory feedback) are recorded on separate channels of a DAT recorder (Tascam DA-20 MKII) for later analyses. The signal processing time is fixed at 3 ms.

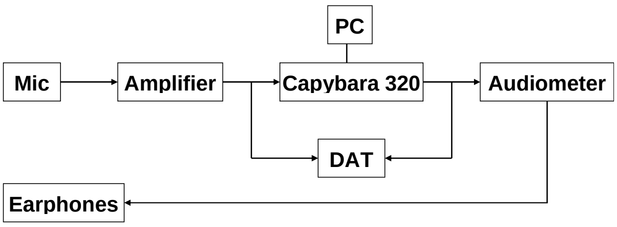


Figure 1. Block diagram of the apparatus.

The Capybara is interfaced with the computer (Dell Inspiron 5100 laptop) via a Firewire cable. The computer controls the Capybara using the Kyma X (6.18) software package. The experiment uses two components of Kyma: the flowchart and the script. The flowchart component acts as a virtual set of connected signal-processing elements (e.g. filters, delays). This set of elements is "compiled" into a single virtual device and used by the script component. The later component provides automated control for the virtual device. All programming is written in Capytalk, a language based on Smalltalk-80. The script component provides a graphic user interface which presents visual prompts to the subject, toggles the sound module on or off, evaluates the subject's response and sets the delay value in the virtual device accordingly. Finally, it stores the trial number, delay time, and subject's response in a comma-separated-values file. This file can be imported easily into several programs (e.g. Microsoft Excel, SPSS) for further analysis.

3 Study

This pilot study tested the feasibility of the modified psychoacoustic adaptive up-down procedure as a valid and reliable measure of DDT. The approach was used to quantify the relative salience of speech feedback delay as a function of low-frequency hearing (Barac-Cikoja & Adornetto, 2004). In self-hearing, the bone-conduction pathway acts as a low-pass filter contributing mostly low-frequency speech information. Consequently, when delaying the speech feedback, one can expect that a temporal asynchrony between the two self-generated components (air and bone conducted signals, respectively) is most prominent in the low-frequency region. The question addressed by this study was: How do different manipulations of the low-frequency information in the air-borne speech feedback affect the subject's ability to detect feedback delay?

The subject's speech signal was filtered in order to separate the low frequency (< 300 Hz) band (LFB) from the rest of the speech spectrum (HFB), and his/her DDTs were obtained under three feedback conditions: (1) both LFB and HFB were equally delayed (*overall delay*), (2) HFB was delayed while LFB remained unaltered (*spectral asynchrony*), and (3) HFB was delayed and LFB was filtered out and replaced by similarly filtered white noise (*masked LFB*). It was expected that DDTs would be the shortest (i.e., delay salience highest) for the overall delay condition, longest for the spectral asynchrony condition, and in between for the masked LFB condition. In addition, each feedback condition was repeated using different syllables (/pa/, /ta/ or /ka/). An additional goal was to evaluate how reliable the threshold estimates were over repeated testing (during and across sessions). Six normal hearing subjects participated in six testing sessions each. Three feedback conditions and three syllables were presented in each session. The presentation order of the experimental conditions was balanced within and across sessions and subjects.

Testing was conducted individually, in a double-walled, audiometric test booth. At that time, the procedure was semi-automated. The experimenter administered the conditions and collected subject responses, communicating visually from an adjoining booth through a window. Specifically, she presented visual prompts to the subject, controlled the sound module, and recorded subject responses. The experimenter adjusted the delay values in the sound module by applying the procedure described in the Method section. Each session lasted less than one hour. The experimenter's role is currently performed by the aforementioned script component (see Apparatus).

Repeated measures ANOVA on an individual subject's DDTs revealed no significant main effect of the two factors (feedback condition and speech content) or their interaction. Although the threshold estimates varied among subjects (1.9 to 7 ms), they were stable for each individual. No significant changes in thresholds were found as a function of practice (session number) or fatigue (presentation order within a session). Finally, the acoustic analyses of the recorded speech showed that the average speaking rate, speech level and voice register were remarkably stable during the testing. The stability of speech production ensured reasonably uniform stimulus materials to carry out the perceptual task.

The findings of this study suggest that the information in the frequency region below 300 Hz (which changed across the experimental conditions) was not necessary to detect feedback delay consistently. However, it is important to note that at the *suprathreshold* delay values, the subjects' experience with the speech feedback (as expressed in their comments after the test), was very different across the three experimental conditions. Specifically, the *masked LFB* condition was perceived as most difficult to do. Also, the quality of the speech feedback was perceived as considerably worse in the *spectral asynchrony* condition. Thus, although the threshold values failed to reveal a significant contribution of low frequency information to the perception of feedback delay, larger delays made such information perceptually significant.

4 Conclusions

Reliable estimates of a minimal detectable delay of the speech feedback can be obtained using an adaptive up-down psychoacoustic procedure. Results of an initial study showed that such estimates could be stable over repeated testing. Also, the use of different syllables might not affect these estimates.

This method not only can be used for determining various individuals' DDTs, but it also could be adjusted for testing sensitivity to other acoustic dimensions of the speech feedback. Furthermore, it could be used to obtain data on the effects of various audiological, technological, and environmental conditions on self-hearing in a way that is methodologically rigorous. For instance, this method may be applied to assess the relative salience of speech feedback delay under different listening conditions: listening in various noise backgrounds that reflect natural communication conditions, or listening binaurally with a combination of devices (e.g., a hearing aid and a cochlear implant) that may introduce an uneven delay across the ears. While the advantages of using this methodology to investigate the role of various factors in speech self-monitoring are apparent, limits on the interpretation of the psychoacoustic findings need to be considered. For instance, one's sense of disturbance by or effort to produce speech under delayed feedback may not be apparent at the level of the minimal detectable delay.

Supported by A Rehabilitation Engineering Research Center (RERC) on Hearing Enhancement to Gallaudet University.

References:

1. Agnew, J. & Thornton, J.M. (2000). Just noticeable and objectionable group delays in digital hearing aids. *Journal of the American Academy of Audiology*, **11**, 330-336.
2. Barac-Cikoja, D. (2004). Effects of Temporal and Spectral Alterations of Speech Feedback on Speech Production by Persons with Hearing Loss. *Arch Otolaringol Head Neck Surg*, **130**, *598-603*.
3. Barac-Cikoja, D., & Adornetto, C. (2004). The role of low-frequency hearing in speakers' sensitivity to delayed auditory feedback. *Journal of the Acoustical Society of America*, **115**, *2394.*
4. Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49(2)**, *467-477*.
5. Stone, M.A. & Moore, B.C.J. (1999). Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses. *Ear & Hearing*, **20**, 182-192.
6. Stone, M.A. & Moore, B.C.J. (2002). Tolerable hearing aid delays. II. Estimation of limits imposed during speech production. *Ear & Hearing*, **23**, 325-338.
7. Stone, M.A. & Moore, B.C.J. (2003). Tolerable hearing aid delays. III. Effects on speech production and perception of across-frequency variation in delay. *Ear & Hearing*, **24**, 175-183.

Graphical gesture analysis: a behavioral tool for virtual environment design

D. Fass
Augmented human project, ICN Business School and LORIA/INRIA Lorraine, Nancy, FRANCE

Abstract
Virtual environments design for training or operational purposes requires taking into account technical devices, artificial multimodal patterns of stimuli together, and their integration into the dynamics of the human sensorimotor behavior. It's a problem of coupling a biological system, the human participant, with a physical system, the less or more immersive artefactual interactive environment. Because sensorimotor integration is an essential element in control of posture, locomotion and gesture, as well as coordinated body activities as manipulation of object and use of tool, there is a critical need to assess the dynamical effects of the design and complexity of virtual environments on the integrative complex body activities. This paper focus on a gesture-based method for assessing both design and integration of the virtual environment structures, functional interactions and functions by their physiological effects. By analysis of three-dimensional hand movement (drawing of ellipses), we compare the dynamical sensory-motor integration and motor performance in real and virtual environments.

Keywords

Virtual environment, design, gesture analysis, human-artefact integration, integrative physiology.

1 Introduction
Augmented human capabilities and enhancing human performance by the use of virtual environment (VE) technologies involves a shift from metaphorical design (Stanney 1995), grounded on metaphysics, to a predictive engineering of interaction and immersion grounded on an integrative theory of human being biology. The question is how to couple and integrate a biological system with a physical and artefactual system in a behaviorally coherent way by an organizational design. This integrative design theory is based on Chauvet's theoretical integrative physiology (Chauvet 1993, a b and c). Within this framework it is necessary to research the technical and physiological principles and elements for virtual environments design. As a consequence, it requires dramatically an experimental techniques allowing and assessing integration of design principles for virtual environment in a dynamical approach. Thanks to its dynamical expression, graphical gesture in the three dimensional space is a behavioral tool for VE design.

1.1 Virtual environment: an artificial knowledge based environment
Virtual reality and augmented reality technologies, because they are multimodal and aesthetic, are obviously the tools for the design and development of the assistance action and multimodal knowledge based artefactual environments (Fass 2002a). Knowledge is gathered from interactions and dynamics of the individual-environment complex and motivations. It is an evolutionary, adaptative and integrative physiological process. It is fundamentally linked to emotions, mnesic process, perception and action. Thus, designing an artefactual or a virtual environment, a sensorimotor knowledge based environment, consists of making biological individual and artefactual physical system consistent. This needs an "eco-ethological" approach, both for the knowledge modelling and interactive system design.

1.2 Sensorimotor integration and motor control
Humans use multimodal sensorimotor stimuli and synergies for interacting with their environment, either natural or artificial (vision, vestibular stimulus, proprioception, hearing, touch, olfaction, taste..) (Sporns and Edelman 1998). When a subject is in a situation of immersive interaction, wearing head-mounted display and looking at a three-dimensional computer-generated environment, its sensorial system is submitted to an unusual pattern of stimuli. This dynamical pattern may largely influence the balance, the posture control, the spatial cognition and the spatial motor control of the individual. Moreover, the coherence between artificial stimulation and natural perceptual input is essential for the perception of the space and the action within.

1.3 Coherence and integrative design
If this coherence is absent, perceptual and motor disturbance appears, as well as illusions. These illusions are solutions built by the brain in response to the inconsistency between sensorial stimuli and internal processes. Therefore, the cognitive and sensorimotor abilities of the subject may be disturbed if the design of the virtual environment does not take into account the constraints imposed by human sensory and motor integrative physiology. The complexity of physiological phenomena arises from the fact that, unlike ordinary physiological system, the functioning of a biological system depends on the coordinated action of each the constitutive elements (Chauvet 2002). This is why the designing of a VE, as an augmented biotic system, calls for an integrative approach.
Integrative design strictly assumes that each function is a part of a continuum of integrated hierarchical levels of structural organization and functional organization. Thus, the geometrical organization of the virtual environment structure, the physical structure of interfaces and the generated patterns of artificial stimulations, conditioned the dynamics of hierarchical and functional integration. Functional interactions, which are products or signals emanating from a structural unit acting at a distance on another structural unit, are the fundamental elements of this dynamics.
Designing a virtual environment consist in organizing the linkage of multimodal biological structures, sensorimotor elements at the hierarchical level of the living body, with the artificial interactive elements of the system, devices and patterns of stimulation. It's also organizing the "transport" of functional interaction in the augmented

space of both physiological and artefactual units. Thus *function* can be viewed as the final result of a set of functional interactions that are hierarchically and functionally organized. Finally, the augmented physiological function may be identified with the global behavior of the hierarchical system in the dynamics of the human-artefactual environment complex (Fass 2002b).

Theoretical integrative physiology claims to develop experimental techniques to organize and to assess the behavioral coherence of the virtual environment design for augmented human performance. Because gesture is a high level integrated sensorimotor and cognitive physiological function, it is a primary expression of this global behavior.

2 Graphical gesture and virtual environment assement:

The gesture based method for virtual environment design and assessment is a behavioral tool inspired by the Chauvet's theoretical integrative biology and the neurophysiology. It's an integrated marker for the dynamical approach for VE integrative design, and the search out of interaction primitives and validation of organization principles It's an integrated marker for a dynamical approach for VE integrative design.

2.1 Traditional assement method

There are different traditional assessment methods to investigate user's behavior interacting with virtual environments (Barfield & Furness, 1995). Psychological method investigates cognitive strategies. Qualitative method uses questionnaire and semantic scale for assessing subjective feelings. Psychophysics experiments perceptual thresholds and reaction time according to stimuli. Classical physiological methods explore correlated variations of basic physiological parameters like blood pressure, heart and respiratory frequencies or EEG signal changing. Applied ergonomics (Wilson, 1999) investigates hardware and software usability in term of participant performance. For human centered design point of view, cognitive function analysis (Boy, 1998) analyzes human-machine interaction in term of function, assuming that a function includes both knowledge about an activity and the way this activity is processed.
If they are locally pertinent these analytical methods do not allow the assessment of human-VE integration and design.

2.2 Sensorimotor maïeutics

Following method developed for expert system engineering (knowledge based system) and to explore the knowledge nature, as a behavioral property of coupling generated in the dynamics of the individual-environment interaction, either natural or artificial, we use a sensorimotor maieutics.

2.3 Assessment method : graphical gesture

The goals of our research present in this paper is to search out the technical and sensorimotor primitives of sensorimotor knowledge for assisting gesture wearable virtual environment design, both for virtual reality and augmented reality.
Our behavioral assessment adapts to virtual environment a neurophysiological method used in motor control researches to study the role of the body in human spatial orientation (Gurfinkel and al., 1993) and the representation of the peri-personnal space in man (Ghafouri, 1996). This investigation are performed by analyzing the upper limb movement in the three dimensional space. This movement consisted in the production of ellipses drawing executed in the three anatomical references planes: sagittal, frontal and horizontal. In its comparative studies about mental representation of space and the influence of visual information Ghafouri showed that the motor expression of the anatomical reference planes was anisotropic (Ghafouri 1997)
By selecting this experimental paradigm, we consider the movement as the expression of a cognitive process *per se*: the integrated expression of the sensorimotor three-dimensional space. According to this paradigm, our experimental paradigm consist in drawing ellipses, whose eccentricity is 0.87 (major axis 40cm, minor axis 20cm), in the free 3-D space with or without wearing virtual environment.
Ellipses drawing without VE are the control experiment (figure1). It consists of two main situations: opened eyes and closed, touch or guided by a real wooden ellipse, and memorized without model.

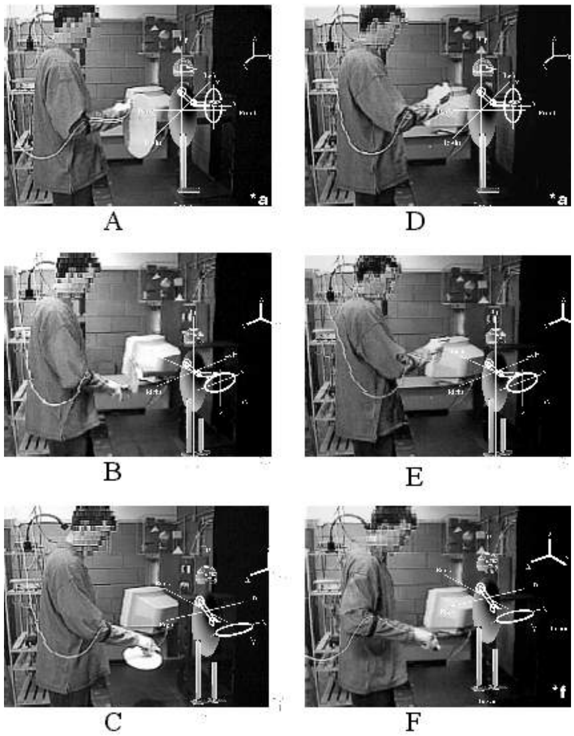


Figure1 . ***Learning and control situation of drawing ellipses, with model: A vertical sagital (SV), B transversal frontal (TF), C transversal horizontal (TH), and memorized: D SV, E TF and F TH***

To highlight structural and dynamical primitives of VE design for assisting gesture (figure2), we define the *a priori* main classes of VE elements of organization:

- Virtual reality
- Augmented reality (see-trough)
- Type of head mounted display (HMD) : immersive or not, low or high visual accuracy, narrow or large field of view, light or heavy, well balance or not
- Visual structure of space : allocentric or egocentric
- Background colour
- Model of ellipse and its plane of movement

- Movement visual feedback: anthropomorphic (hand) or abstract (ball), dynamical quality of feed back (remanence, persistence)
- Sensorimotor coupling: visuo-proprio-kinesthetics or visuo-haptic
- Spatial orientation of gesture: vertical sagittal, transversal frontal and transversal horizontal.

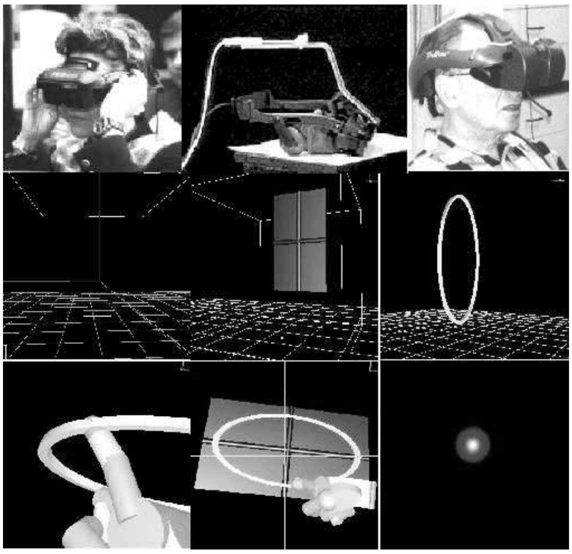


Figure2. *Examples of different structural and functional primitives for VE design.*

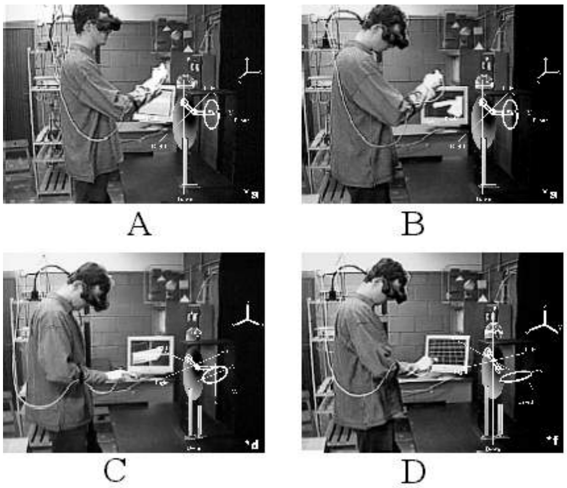


Figure3 . *Graphical gesture of ellipse drawing in the 3D space is performed and analyzed in different configurations, more or less complex, of immersive virtual environment assisted drawing ellipses: A- SV ellipses and neutral and colored background, B- SV ellipses and anthropomorphic visual feedback of movement (artificial hand), C- TF and model of ellipse insert in its plan of movement without visual feedback of movement, D- TH ellipses and abstract representation visual feedback of movement (ball).*

To highlight the dynamical principles of organization for assisting gesture, we set up a protocol according to a complex and incremental VE design, allowing intuitive learning of both task and use of VE. Ten volunteers (7 men and 3 women, 25 to 35 years old) were asked to performed graphical gesture in the three anatomical planes of reference for each step of design complexity (Figure3). The first step of the protocol consists of drawing ellipses wearing a turned off HMD to study the influence of HMD design and intrusiveness on sensorimotor integration and motor control. The last step of virtual reality artifact combined allocentric and egocentric prototypic structural elements of artificial visual space, model of ellipses and their planes of movement, and a visual feedback of movement.

2.4 Material

Material bring into play for our experiments are traditional hardware and software element of a wearable VE : head mounted display I-Glasses ® immersive or see trough, a Kayser Threde Proview® HMD a Frastrack Pohlemus® electromagnetic motion tracking system, a workstation and a specific software for managing and generating the visual virtual environment in real time. This software allow us to register data from sensor

2.5 Graphical gesture analysis

- Data analysis:

For gesture analysis, we calculate sixteen variables from raw data (xyz coordinates) produced and registered from the sensor wear on the tip of the index finger of the working hand during 15 second at 120hz of drawing ellipses. Four categories of variables characterize graphical gesture: kinematics (A, B, C, D), position (F, G, H, I), orientation (J, K, L, M) and shape (E, N, O, P).

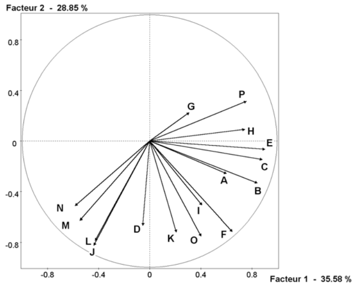
Index	Variables
A	Number of ellipse
B	Average velocity (cm/s)
C	Covariation Vt/Rt
D	Amplitude (cm)
E	Mean area (cm²)
F	Global position
G	Position / x axis (cm)
H	Position / y axis (cm)
I	Position / z axis (cm)
J	Global orientation
K	Orientation / sagittal plane(d°)
L	Orientation / frontal plane(d°)
M	Orientation / horizontal plane(d°)
N	Eccentricity
O	Major axis variation
P	Minor axis variation

- Statistical analysis:

Because each experimental step is described by a large number of parameters, we use a method of multidimensional statistical analysis, a principal component analysis (PCA) coupled with a hierarchical classification. These descriptive analyses are calculated with the Decisia SPAD 4.0® software from the mean values of the ten subject's sixteen computerized variables for each experimental situation.

PCA highlights the differences and similarities between each experimental situation and consequently for each element of VE design. The position of variables on the principal plane is interpreted in term of correlation. The property of hierarchical classification is to gather experimental situations in a limited number of homogeneous classes.

Thus we are able to show differential effects for each element of the incremental design of VE and to assess global design and dynamics of the human virtual environment integration. These experimental results will found VE design modelling according to the hierarchical organization of theoretical integrative physiology.



RANG	IND.	IDEN
1	0.38	19a
2	0.65	24f
3	24.18	22d
4	1.76	30f
5	7.41	201f
6	31.84	39f
7	0.54	25a
8	3.03	37a
9	1.41	163a
10	22.8	199a
11	1.13	28d
12	4.97	38d
13	0	164d
14	0	200d
15	-----	165f

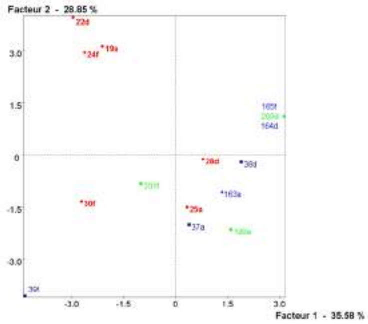


Figure 4. *PCA and hierarchical classification for ellipses drawing wearing three configurations of turn off HMD and compared to the control situation with closed eyes. It shows different intrusiveness of HMD physical designs and their influence on gesture according to anatomical reference planes.*

3 Feature developpement

Feature development concerns automatic gesture analysis link to on-line adaptation of multimodal artificial patterns generation within VE. We also look forward to fusion of data from different motion system analysis: gesture and sight to understand the dynamics of the visual-motor loop in VE to refine design principles. But, most important work ongoing concerns mathematical modelling of human-VE integration. It is required for a predictive integration design and engineering of VE.

4 Conclusion

Graphical gesture analysis is an objective behavioral tag of the sensorimotor integration within natural and virtual environment. Its afford designer to understand hierarchical structural and functional organization of VE for a predictive engineering of both interaction and immersion.
Knowledge about interaction primitives and their influences on high level motor performance as graphical gesture, its principles of organization and its assessment must be the basis for virtual environment design. Furthermore it must be fundamental to virtual environment assessments and their applications for training of technical gesture and expertise acquisition, or their variation for operational assistance like augmented reality, teleoperation e.g. in human space activities (Fass 2005).
Finally we would like to emphasize augmented human performance and enhancing human capabilities by virtual environments technologies need predictive integrative design using mathematical models ground on the Chauvet's theoretical integrative physiology. It is our future challenge.

Supported by CUGN, Région Lorraine, INRIA Lorraine.

References:

1. Stanney, K. (1995) Realizing the full potential of virtual reality : human factors issues that could stand in the way, *1995 IEEE Virtual reality annual international symposium*, Research triangle park, North Carolina, March 11-15 1995.
2. Fass, D. (2002) Base de connaisances multimodales et esthétique de la connaisance, aide à l'action et réalité virutelle: théorie et experiences, Phd thesis, Université Henri Poincaré, Nancy1.
3. Chauvet, G. (1993) a. Hierarchical functional organization of formal biological systems: a dynamical approach. I The increase of complexity by self association increases the domain of stability of a biological system. *Phil. Trans. R. Soc.Lond.*, **339**, 425-444.
4. Chauvet, G. (1993) b. Hierarchical functional organization of formal biological systems: a dynamical approach. II The concept of non-symetry leads to a criterion of evolution deduced from an optimum principle of the (O-FBS) subsystem. *Phil. Trans. R. Soc.Lond.*, **339**, 425-444.
5. Chauvet, G. (1993) c. Hierarchical functional organization of formal biological systems: a dynamical approach. III The concept of non locality leads to a field theory describing the dynamics at each level of organization of the (D-FBS) subsystem. *Phil. Trans. R. Soc.Lond.*, **339**, 425-444.
6. Sporns, O. and Edelman G. (1998), Bernstein's dynamic view of the brain: the current problems of modern neurophysiology (1945), *Motor Control*, **2**, 283-305.
7. Fass, D. (2002). Éléments d'organisation du complexe Individu – Environnement artefactuel : Évaluation d'un système de réalité virtuelle d'aide à l'action. *RIHM - Revue d'Interface Homme-Machine*, **3**, 2, 75-94.
8. Chauvet G. (2002) On the mathematical integration of the nervous tissue based on the S-Propagator formalism: I. Theory, *Journal of Integrative Neuroscience*, **1**, 1, 31-68.
9. Barfield, W & Furness, T.A. (1995) Virtual environments and advanced interface design. Oxford University Press, New York Oxford.
10. Wilson, J.R. (1999) Virtual environments applications and applied ergonomics, *Applied Ergonomics*, **30**, 3-9.
11. Boy, G. (1998) Cognitive function analysis, Ablex Publishing Corporation, Standford.
12. Gurfinkel, VS., Lestienne, F., Levik, YU.S., Popov K.E. et Lefort, L. (1993) Egocentric references and human spatial orientation in microgravity : II body-centred coordinates in the task of drawing ellipses with prescribed orientation. *Experimental Brain Research*, **95**, 343-348.
13. Ghafouri, M. & Lestienne, F.(1996) Drawing in space : orientation and figural aspect of movement, *Fourth Interantional Symposium on 3-D analysis of Human Movement*, Granoble, France, July 1996.
14. Ghafouri, M. (1997) Représentation sensori-motrices du geste graphique dans l'espace tridimensionnel. , Phd thesis, Université Henri Poincaré, Nancy1.
15. Fass, D. (2005) Virtual environment a behavioral and countermeasure tool for assisted gesture in weightlessness: experiments during parabolic flight, *ESA and ISGP Joint life conference "Life in space for Life on Earth"*, Cologne, Germany, 26 June- 1 July.

Acoustic monitoring of activity patterns in office, street and garden environments

A. Härmä, J. Skowronek, and M. F. McKinney
Digital Signal Processing Group, Philips Research Laboratories, Eindhoven, The Netherlands

Abstract
This paper is based on a simple experiment: we recorded all interesting sounds in an office, and a backyard garden over observation periods of several weeks. We then applied different types of signal processing and pattern recognition techniques to the data to produce representations that could reveal interesting structures of activity at the sites.

Keywords
Acoustic surveillance, office sounds, birds, dogs, people, unsupervised learning

1 Introduction
Many activities and behavior produce recognizable audible events. From the point of view of monitoring, sound has the advantage over light (vision) in that sound propagates through light walls and diffracts around obstacles. Therefore, acoustic monitoring of the activity of humans, animals, and machines in our environment is an interesting approach in many different applications.

Acoustic surveillance has a long tradition in military undersea applications, for example, the IUSS system of the U.S. Navy has been used for decades. The earliest form of techniques related to the automatic recognition of sounds were activity detectors that triggered alarms or started recording if the energy of the microphone signal exceeded a certain threshold. The activity detector also became a standard component in speech communication systems. The largest body of work in sound recognition is related to automatic speech recognition. However, recognition of the other sounds, including vehicles [1], machines [2], and sounds of closing doors, dropping or breaking objects [3-5] have been studied earlier. In addition, the new MPEG7 multimedia standard provides an interesting framework for sound recognition [6]. The automatic detection of the context of a mobile phone user is also an active topic of research [7].

Automatic recognition and classification of audio and music has recently gained much interest due to developments in Internet and broadcast technology. Users are enjoying larger amounts of multimedia content as a result of increased common storage capacities and data transmission rates and they need automatic methods to label and navigate their audio. A major challenge in this field is the automatic classification of audio and music [8-15].

Most audio classification systems combine two processing stages: feature extraction followed by classification. A variety of signal features have been used for this purpose, including low-level parameters such as the zero-crossing rate, signal bandwidth, spectral centroid, and signal energy. Another set of features used, inherited from automatic speech recognizers, is the set mel-frequency cepstral coefficients (MFCC) [16]. Typical performance of these feature sets in speech/music discrimination tasks is around 95% [17, 10, 18] but decreases as the number of audio classes increases [13]. There has also been some recent work on automatic music genre detection: Tzanetakis and Cook [15] combine standard features with representations of rhythm and pitch content and show classification performance in the range of 60%.

Several different classification strategies have been employed in these studies, including multivariate Gaussian models, Gaussian mixture models, self-organizing maps, neural networks, k-nearest neighbor schemes and hidden Markov models. In some cases, the classification scheme does not influence the classification accuracy [10, 19], suggesting that the topology of the feature space is relatively simple. An important implication of these findings is that, perhaps further advances could be made by developing more powerful features or at least understanding the feature space, rather than building new classification schemes.

In bioacoustics, acoustic monitoring pattern recognition techniques have been used to find specific predefined sounds from recordings. For example, in avian vocalization various types of sound-specific parameter sets have been used by many researchers to study geographical variation [20], social ranking in a population of one species of birds [21], or imitation [22]. Kogan *et al.* [23] introduced a bird sound recognition system based on template matching of spectrograms. The prototype sound was defined by hand and the recognition was then based on matching templates with spectrogram frames from a continuous recording. Methods for the automatic recognition of bird species based on more generic features have been developed, e.g., in [24-27].

Most of the audio classification techniques mentioned above are based on some predefined training set, that is, supervised classification. In speech recognition it is some predefined speech material, and in music genre and bird species recognition it is a set of known recordings. The need for a predefined training set is obviously a limitation for the applicability of automatic acoustic monitoring techniques in many applications. In acoustic activity and context detection in various *ambient intelligence* applications or in behavior research it would be desirable to have one system that could adapt to different environments, e.g., an office or a garden, without the pre-selection of training material, or consultation with an audio recognition specialist.

In the current paper we describe a straightforward experiment based on automatic unsupervised classification of short isolated acoustic events. A software system was designed that records all *interesting* acoustic events that deviate from a continuously updated estimate of the background noise spectrum. For classification, each sound event was represented by a small number of descriptive features. The set of features selected contains many typical signal measures used in the above-mentioned sound recognition applications. Next, an automatic clustering algorithm based on feature vectors was used to define an automatic classifier for the sound events at each recording site. Finally, the classifier was then used to construct different types of logs and statistical representations reflecting various types of behavior at the site.

The system was tested at three different sites: an office room, a street side, and a backyard garden. The office data reveals interesting details of the behavior of the people working at the site, and the backyard data draws a detailed picture of the daily activities of various bird species, and dogs, in the neighborhood. In this article we describe the method and show some specific cases of regular patterns of activity that we found from the data. Based on the results of the experiment, we try to characterize the usability of the proposed method in a generic case of measuring patterns of activity in an arbitrary acoustic environment.

2 Recording system and recordings

The recording system has been described in detail in [28]. The real-time recording system runs on a PC with a microphone. The microphone was an ordinary omnidirectional electret condenser microphone. The system analyzes the background noise spectrum from the microphone input signal. There are two different criteria for detection of the onset of an *interesting* event. The first onset detector is a *whistle* detector tuned to trigger recording when there is an increase of signal energy within a narrow frequency band. The second detector is a *slam* detector sensitive to sudden peaks in signal energy. When either of the two detectors activates the software starts recording the microphone signal into the recording database.

The system has an input buffer that makes it possible to start the actual recording from a signal position shortly before the time of the detected event. This feature is absolutely necessary in order to capture short sound events.

The recording system is designed such that once set up, it can run unattended as long as there is disk space available in the computer to store new entries. The maximum duration of an event is limited to four seconds in order to avoid filling the computer hard disk too quickly. In a typical busy environment the system stores approximately 1000-5000 sound files a day with an average duration of 2 seconds. That is, the recorded events typically cover only 1.5-8 % of the time. Note that at the sampling rate of 16 kHz an audio database of all interesting acoustic events of the year defined this way takes less than 40 Gigabytes. The recording time, site, and other meta information about the recording are written into a specific XML-based recording database which also provides a convenient display to the process.

In the current experiment the actual classification and log-keeping system runs off-line.

3 Automatic classification

The data contains hundreds of thousands of short audio files each associated with a time tag with one-second precision. There are several different ways to access the data. In this article we mainly focus on techniques based on automatic classification of sounds.

In order to use automatic classification and pattern recognition techniques it is necessary to represent a sound event by some small number of features. In the current article we use 12 features, which have been previously found useful in various applications. The features have been explained in detail in [14] and [29]. Some of the features characterize properties of the signal spectrum such as the spectrum centroid, or bandwidth. Other features are related to the temporal envelope of the signal and the duration of the acoustic event.

The classification of data was performed using the well-known k-mean clustering algorithm. The algorithm aims at finding a predefined number of prototype feature vectors representing clusters in the feature space. The number of prototype vectors was set to 30 in each case. This selection is arbitrary. The large number and variability of the data made it very difficult to utilize any of the known statistical techniques for the optimal selection of the number of classes and optimization of the parameters of the classification algorithm. In fact, it turned out that the k-means algorithm only rarely converged optimally. Nevertheless, at all sites it was common that a certain group of sounds was well presented in the classification.

The initial vectors were selected randomly from the data. Therefore, the classification led to different classes each time the clustering algorithm was run. Each result below represents one particular classification.

4 Experimental data

The feature vectors of each recorded sound event were computed and a classifier was created using the k-means algorithm. In order to know what the system actually learned it was necessary to perform some manual analyses. In each case we first determined the actual recorded entry that is closest to each of the class centers and named the event manually. Then we measured a large number of closest matches to the class center and ranked them by the similarity. Finally, the number of closest matches was chosen manually such that at least 90% of them represent the same sound event as the named closest sound. It is clear that this process can be made automatic.

4.1 The office data

The largest data set was collected in a small office room. The recording period started in September 2004 and continued 24 hours a day, seven days a week, until the beginning of January 2005. The total number of recordings was over 150 000.

Manual classification of a random selection of the office recordings showed that 65% of the recordings are sounds of the computer keyboard, 20% are speech sounds, 6% are high frequency whistling sounds from humans and the printer in the hallway. The rest of the sounds include sounds of doors (2%) and office furniture, coughing and sneezing, unidentified rumble, and ringing telephone.

The automatic classification method found several different classes for different types of keyboard sounds such as single hits and fast rolls of the keyboard. One event that was easily found was the loud slam of a door across the hallway from the office room. A histogram of door slams per hour in a single day in November is shown in Fig. 2. There is a clear a peak in activity right before lunchtime. Another event that was regularly found was a coughing sound of a person close to the microphone. A log of coughing sounds in the beginning of October 2004 is plotted in Fig. 2.

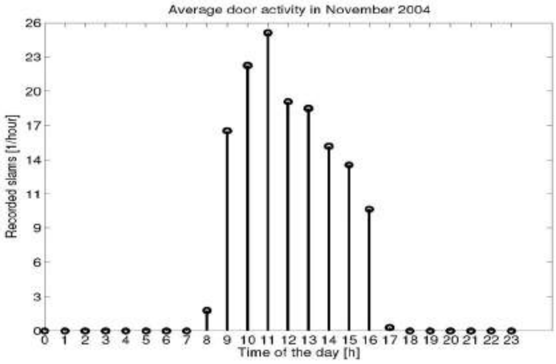


Figure 1 Door slams close to the office room as a function of the time of the day on week days in November 2004.

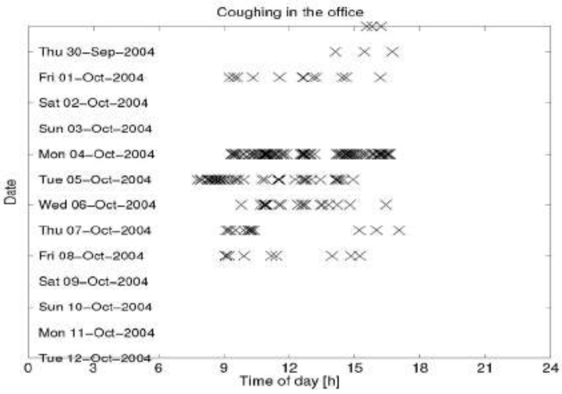


Figure 2 Log of coughing sounds in the office in the beginning of October 2004. Each cross indicates one recorded event.

4.2 Street side data

In January 2005 the microphone was moved outside of the office window with appropriate shielding from wind and rain. The system ran unattended for slightly more than one month collecting data from a relatively quiet side street in a closed industrial area. Unfortunately, a large number of recordings were corrupted by high levels of wind noise and electrical problems. There are some clear trends that can be revealed, for example, the low activity during weekends and voices of people going to the cafeteria at lunchtime.

4.3 Garden data

The recordings from a backyard garden of a typical Dutch townhouse were collected only during weekend days. Therefore the data is not as continuous as the two other sets. However, the data contains the sounds of almost all Saturdays and Sundays starting from December 2004 until the beginning of May 2005 and the final number of recordings was over 200 000. The manual classification of recordings of one afternoon in early February revealed that 30% of events were dog sounds, 23% human voices, 20% bird sounds, and the rest contained mainly sounds of motor vehicles.

Automatic classification of the garden data often found classes corresponding to dog sounds, and sounds of *Parus Major* (Great Tit), which is one of the most common bird species in the garden. One example from the data is shown in Fig. 3. The song activity of the birds is low in January but is significantly increased at the end of February. On the other hand, in most hours there is only a small difference in the barking of the dogs in the area. However, at the end of February the barking is more common at 1pm and late afternoon than in January.

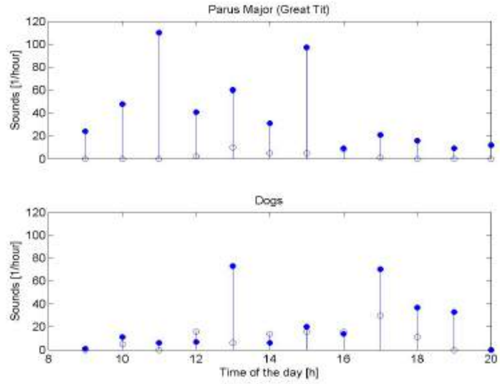


Figure 3 Daily statistics of sounds of Great Tit and Dogs in the garden on 14 January (circles) and 20 February (filled circles).

4.4 Analysis of the classifiers

To compare the performance of the proposed method at the three sites we trained 20 classifiers for each site from different random samplings of 10000 data points. The classes were sorted by their sizes and the percentages of samples in classes and the distances of individual events to the class centers were computed. The average results are plotted in Fig. 4. In the most common classes (low indices) the office and street data are similar. However, for smaller classes the distances (bottom panel) grow large in the street side data. The garden data is slightly more uniformly distributed to different classes but the distances are usually larger which indicates a larger variability in data.

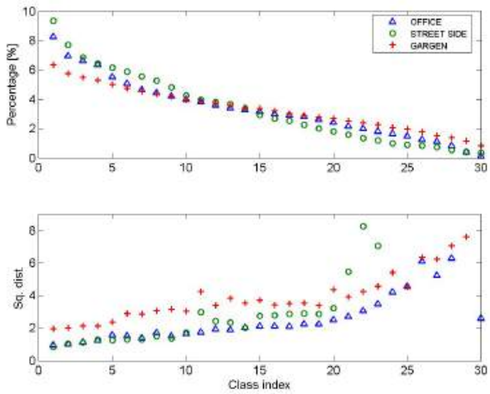


Figure 4 Class percentages and squared distances in automatic classification averaged over sorted classification results.

5 Conclusions

We have used some generic, robust, and well-known, yet relatively simple techniques from the field of automatic sound recognition and applied them to extensive data collections from three different natural environments. We have shown that with these standard techniques, information about general behavioral patterns can be gleaned from acoustic signals in common environments. The goal at this point is to investigate what can be learned from using these methods and where the biggest challenges lie in utilizing these techniques in a wider range of environments. The analysis of the results will be elaborated in a follow-up paper.

References:

1. Munich, M. E. "Bayesian Subspace Methods for Acoustic Signature Recognition in Vehicles." Proc. European Signal Processing Conf. (EUSIPCO-2004), Vienna, Austria: 2004.
2. Atlas, L. E., G. D. Bernard, and S. B. Narayanan. "Applications of Time-Frequency analysis to signals from manufacturing and machine monitoring sensors." Proc.IEEE 84 (/9): 1319-29
3. J.P.Woodard. "Modeling and classification of natural sounds by product code hidden {Markov} models." IEEE Trans.Signal Processing 40 (1992): 1833-35
4. Goldhor, R. S. "Recognition of Environmental Sounds", Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing (ICASSP'93). 1993. 149-52.
5. Watanabe, H. et al. "A new approach to acoustic signal monitoring based on the generalized probabilistic descent method." IEEE Trans.Signal Processing 47 (/9): 2615-18.
6. Casey, M. MPEG-7 Sound-Recognition Tools." IEEE Trans.Circ.Systems for Video Tech. 11 (/6): 737-47.
7. Eronen, A. et al. "Audio-Based Context Awareness - Acoustic Modeling and Perceptual Evaluation.", Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing (ICASSP'03). Hong Kong: 529-32.
8. Wold, E., Blum, T., Keislar, D., & Wheaton, J. (1996). Contentbasedclassification, search, and retrieval of audio. *IEEE Multimedia*, *Fall*, 27–36.
9. Spina, M. S., & Zue, V. W. (1997). Automatic transcription of general audio data: Preliminary analysis. In *Proceedings of the 4th International Conference on Spoken Language Processing.* Philadelphia, PA.
10. Scheirer, E., & Slaney, M. (1997). Construction and evaluation of a robust multifeature speech/music discriminator. In *Proceedings of the ieee 22nd International Conference on Acoustics, Speech and Signal processing* (pp. 1331–1334). Munich, Germany.
11. Scheirer, E. D. (1998). Tempo and beat analysis of acoustical musical signals. *Journal of the Acoustical Society of America*, *103*, 588–601.
12. Wang, Y., Liu, Z., & Huang, J. C. (2000). Multimedia content analysis using both audio and visual cues. *IEEE Signal Processing Magazine*, *17*, 12–36.
13. Zhang, T., & Kuo, C. (1998). Content-based classification and retrieval of audio. In *SPIE's 43rd Annual Meeting – Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations VIII.* San Diego.
14. Li, D., Sethi, I. K., Dimitrova, N., &McGee, T. (2001). Classification of general audio data for content-based retrieval. *Pattern Recognition Letters*, *22*, 533-544.
15. Tzanetakis, G., Essl, G., & Cook, P. (2001). Automatic musical genre classification of audio signals. In *Proceedings of the 2nd Annual International Symposium for Music Information Retrieval.* Princeton, NJ.
16. Davis, S. and P. Mermelstein. "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences." Proceedings of the Symposium on Time Series Analysis 28 (1980): 357-66
17. Toonen Dekkers, R. T. J., & Aarts, R. M. (1995). *On a very low-cost speech-music discriminator* (Tech. Rep. No. 124/95). Eindhoven, Netherlands: Philips Research Nat.Lab. Technical Note.
18. Lu, G., & Hankinson, T. (1998). A technique towards automatic audio classification and retrieval. In *4th International Conference on Signal Processing.* Beijing
19. Golub, S. (2000). *Classifying recorded music.* Unpublished master's thesis, University of Edinburgh. (Retrieved January 12, 2003, from `http://www.aigeek.com/aimsc/`)
20. Shieh, S. "Song Structure and Microgeographic Variation in a Population of the Grey-cheeked Fulvetta (Alcippe Morrisonia) at Shoushan Nature Park, Southern Taiwan." Zoological Studies 43 (2004): 132-41.
21. Christie, P. J., D. J. Mennill, and L. M. Ratcliffe. "Chickadee song structure is individually distinctive over long broadcast distances" Behaviour 141 (2004): 101-24.
22. Tchernichovski, O. et al. "A procedure for an automated measurement of song similarity." Animal Behaviour 59 (2000): 1167-76.
23. J.A.Kogan and D.Margoliash. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. J.Acoust.Soc.Am. 103[4]. 2005.
24. A.L.McIlraith and H.C.Card. Birdsong recognition using backpropagation and multivariate statistics. IEEE Trans.Signal Processing 45[11]. 1997.
25. A.Härmä and P. Somervuo. "Classification of the Harmonic Structure in Bird Vocalization." Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing (ICASSP'04), Montreal, Canada: 2004
26. Somervuo, P. and A. Härmä. "Bird Song Recognition Based on Syllable Pair Histograms.", Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing (ICASSP'04), Montreal, Canada: /5.
27. S.Fagerlund and A.Härmä. Parametrization of Non-Tonal Bird Sounds for Automatic Recognition. Proc. European Signal Processing Conf. (EUSIPCO-2005), Antalya, Turkey: 2005.
28. A.Härmä, M. F. McKinney, and J. Skowronek, 'Automatic surveillance of the acoustic activity in our living environment', Proc. IEEE Conf. Multimedia and Expo, (ICME'2005), Amsterdam, July 2005.
29. M. F. McKinney and J. Breebaart, "Features for audio and music classification," in *Proc. Int. Symp. Music Inf. Retrieval (ISMIR'2003)*, (Baltimore, USA), October 2003.

The use of digital video recorders (DVR) for capturing digital video files for use in both The Observer® and EthoVision®

J.S. Church [1] , D.G. Martz [1]
[1] *Livestock Welfare Unit, Alberta Agriculture, Food and Rural Development, Edmonton, Canada*

Abstract

Before switching your laboratory from analog to digital, for the recording of video files for use in Noldus software such as EthoVision® and The Observer®, it must be realized that not all digital video recorders (DVRs) are alike. There are obvious advantages in moving from analog to digital recording for behavioural work including: increased storage capacity; no requirement to purchase video tapes; immediate search by date, time or event; digital images tend to be of higher quality; ability to view study sites remotely by internet connection, and "smart" features such as motion detection. True DVR, generally manufactured and sold for security surveillance purposes are sophisticated systems not to be confused with set top cable boxes that are also referred to as "DVRs," they are not related. But before you throw away your time-lapse video recorders, time code generators and quad-splitters there are some important cautions to realize before switching to DVR. Some research groups have bought digital surveillance systems on the assumption that they work with EthoVision® and The Observer®, only to be disappointed. From a hardware perspective, the system's components appear simple; all that seems to be required are cameras and a DVR. There are many companies offering digital security systems, based on DVR technology, that appear suitable for behavioural observation research. In reality, the majority of systems work poorly, if at all, with EthoVision® and The Observer®. Many believe that building a DVR is as simple as placing a video capture card in a PC and purchasing some "off-the-shelf" software. To date we have found only one system that is appropriate for use by most behavioural researchers. The vast majority of systems depend on proprietary compression software that must then be converted in order for the digital files to work properly in EthoVision® or The Observer®.

Keywords

Recording, surveillance, file format, compression

1 Introduction

The purpose of this paper is to give some insight into digital video recording technology and systems to individuals who study behaviour. The purpose is to aid scientists in investing in the right technology for the job, particularly if contemplating a switch from analog to digital video recording of behaviour. The vast majorities of systems available are designed for security purposes, not for studying animal behaviour, and may not be compatible with EthoVision® and The Observer®. The technology is very new, and while many advantages are touted over traditional analog recording (Table 1), scientists who have invested substantially in these systems have reported difficulties, particularly regarding compatibility with behavioural analysis software like EthoVision® or The Observer®. The problem is that security systems have multiple viewing functions required for surveillance work, but they generally contain proprietary compression and limited video export capabilities. As a result, most systems export digital video to large files that are in a format that is difficult for behavioural software to use.

1.1 What is a Digital Video Recorder (DVR)

The first digital video recorders (DVR) that appeared on the market were mechanical devices that resembled conventional VCR's with similar function keys, making them very familiar and less intimidating, but the shortfall was that a price was paid in terms of functionality. The market demanded more and soon personal computer (PC) based DVR systems became available. PC based DVR systems are complete hardware/software computer-based systems incorporating video capture, video multiplexing, video/audio recording and playback, video display and remote video access via modem, Internet or Local Area Network (LAN). More recently, advances have been made with high-end digital non-PC based DVR's that can duplicate the functions of a PC-based DVR. These embedded DVRs do not use a Windows™ operating system and contain no software at all on the hard disk(s). All applications are contained or "embedded" in firmware (software encoded on chips). Mobile DVRs are embedded systems that were originally developed for use in vehicles. They incorporate anti-shock and anti-vibration mountings and generally have 12 V DC power inputs. At the heart of any DVR system is a video capture card or similar device. They are called "capture" cards because they are "capturing" and recording the video (encoding), but are also responsible for playback and display of the video on screen (decoding). The other component most often includes some kind of CCTV (closed circuit television) camera. Digital video recorders are sophisticated integrated systems composed of a combination of hardware components, software programs and subassemblies. These systems are most commonly used for security monitoring purposes via video surveillance and should not be confused with set-top consumer cable boxes such as TiVo™. While these are also referred to as "digital video recorders", the purpose and function of the two systems are different.

1.2 Problems with implementing "security" based systems for behavioural research

In purchasing a surveillance system, the first priority should be to assess and understand your needs. It is imperative that the system accomplishes exactly what you want it to do, otherwise the capital expenditure required to purchase the system will be wasted. The problem is that most of the companies and individuals selling these systems are unfamiliar with the unique application of the DVR for behavioural observation and simply assume that the system will work. From a hardware perspective, the systems seem simple enough; most are comprised of two main components: cameras and DVR. However, within each of these two components, there are a number of choices to consider, with prices varying significantly. In April 2005, the International Security Conference and Exhibition West was held in Las Vegas, Nevada, USA. The Conference was attended by 21,000 industry participants with 825 exhibitors. Of the exhibitors, one third (246) were promoting some combination of camera/DVR system for security monitoring. These are generally all custom-built systems, and consequently there is a lack of standardization between products. This is especially problematic for the scientist when trying to select the appropriate digital recording system to meet specific behavioural observation needs. Some problems encountered with implementing "security" based systems for behavioural research include:

- Specific systems often work with only certain cameras
- Quality of the recorded video is less than optimal
- Recorded and display speeds were not what was stated
- Numbers of days of storage is less than what was specified
- No one local to support the product
- File format when decoded not compatible with EthoVision® and The Observer®.

2 Understanding compression formats

Working with digital video requires lots of storage capacity, and digital video generally needs to be compressed before it can be stored. All companies provide some form of their own proprietary compression. Compression technology is based on mathematical algorithms. Compression is performed when an input video stream is analyzed and information that is indiscernible to the viewer is discarded. It is actually a case of compression and decompression. Compressed to travel down the network and then decompressed to transmit when it comes out the other side, hence the name "Codec". A video codec is software that can compress a video source (encoding) as well as play compressed video (decoding). Standards are set by The International Organization for Standardization – a non-governmental organization that works to promote the development of standardization to facilitate the international exchange of goods and services and encourage worldwide intellectual, scientific, technological and economic activity; and the International Electrotechnical Commission – the international standards and assessment body for the field of electrotechnology. These two industry bodies recognize a number of standard compression formats, including: MPEG-1, MPEG-2, MPEG-4, Morgan JPEG (MJPEG), Wavelet JPEG, H.264, etc. Manufacturers purchase a license to utilize one of these compression formats, but many attempt to "tweak" the format themselves, making the format unstable for use with either EthoVision® or The Observer®. AVI is especially problematic, but nevertheless it is a format that many digital surveillance companies use. It is important to note that EthoVision® does not support all formats of AVI (such as DivX) or MPEG-4. MPEG-4 has emerged as one of the new standards for both multimedia and Web compression. MPEG-4 is based on object-based compression, which is similar in nature to Virtual Reality Modeling Language. Individual objects within a scene are tracked separately and compressed together to create an MPEG-4 file. This results in efficient compression that is very scalable; from low bit rates to very high. It also allows developers to control objects independently in a scene, and therefore introduces interactivity. As a result it has become a popular format in the digital surveillance industry. However, it is important to reiterate that not all MPEG-4 formats are compatible with EthoVision®. In our experience, for the majority of researchers, it is best to stick with either MPEG-1 or MPEG-2 in order to maintain compatibility with behavioural software.

2.1 MPEG-1

MPEG is a file format for compressing full-motion digital video that was produced by the Moving Picture Expert Group of the International Standards Organization. It was originally designed for up to 1.5 Mbit/sec. For the last couple of years MPEG-1 was the standard for the compression of moving pictures and audio. This format was primarily used for CD-ROM video applications, and it is still a popular standard for video for use on the internet, transmitted as *.mpg files. MPEG-1 is the standard of compression for VideoCD, the most popular video distribution format throughout most of Asia. Of interest, level 3 of MPEG-1 is also the standard of digital compression for audio, better known as MP3. While MPEG-1 produces better results than QuickTime (.mov) and AVI, the MPEG-1 standard provides a video quality that is slightly below that of traditional VCR analog recording. The advantage of MPEG-1 is that the files are relatively small (about 10 Mb per minute) and therefore take up less storage space. Another further advantage is that MPEG-1 software decompression comes standard on the Windows operating systems so it will work on a variety of computers. If continuous observation is an important consideration, or if observation is going to take place over a long period of time, and quality may not be imperative, then this is a very useful format to use.

2.2 MPEG-2

The MPEG-2 format offers higher video quality than MPEG-1, but the downside is that the file size increases dramatically (about 4 times the size or approximately 40 MB per minute). MPEG- 2 is the standard for which most Digital Television set-top boxes and commercial DVD compression is based. While it is based on MPEG-1, it was designed for the compression and transmission of digital broadcast television. The advantage of MPEG-2 is the improved quality, sufficient for all the major television standards, including NTSC. The most significant improvement over MPEG-1 is its ability to compress interlaced video. MPEG-2 scales well to HDTV resolution and bit rates, which essentially eliminates the need for MPEG-3. In terms of quality for behavioural observations, MPEG-2 is adequate for the vast majority of behavioural researchers.

2.3 Considerations of storage

The primary advantage of digital video is that it is easier to manipulate with Noldus software than analog sources, also hardware prices have decreased rapidly in recent years. However, digital video still requires considerable storage capacity and needs to be compressed before it can be stored on a disk medium. Most DVR systems have a lot of flexibility in terms of file storage. Often files can be stored on large hard drives, or even stored on the departmental server over a network. Many systems included CD/DVD burner software, allowing the end-user to burn the files to a CD or DVD for archival storage and some miniaturized systems are designed to work with memory sticks.

As mentioned previously, digital video needs to be compressed before it is stored on a disk medium. With many DVR systems on the market, you can set the sound and image quality of the resulting media file. However, the higher the quality, the larger the file. Thus the highest quality of images and sound require a large amount of space on the disk. In practical terms, most users select their quality with the capacity of their long term storage media in mind. For example, an MPEG-1 of 10Mb per minute you can fit about one hour of video onto one CD (700 Mb).

2.4 Considerations about frame rate

An important consideration when analyzing specifications of a DVR system is to clarify "frame" or "images" per second. Many labs report that the frame rate promised is not the same as the frame rate realized under laboratory conditions. The frame rate issue is very tricky. The fact is the speeds that manufacturers quote are usually the "maximum" obtainable, meaning under ideal conditions, and does not take into account other functions that the PC, software, or video card might be processing simultaneously. There are an infinite number of ways of presenting these numbers, and the numbers can often be misleading. A standard in the marketplace is 30 frames per second which is considered to be real-time / real motion video, but that is for a single video stream. For example, if 4 cameras are recording simultaneously, all in real-time/real motion video, 120 frames per second with full unshared resources is required. It is very important that frame rate be clarified, for example the manufacturer may be discussing:

- The total number of frames/images per second for the entire card spread across all cameras (cumulative total),
- The total number of frames for each individual channel
- The maximum frame capacity of the hardware not taking into account software switching, simultaneous functions, etc. (rated hardware capacity)
- The display speed
- The recording speed
- Or a combination of all of the above

3 Advantages of DVR

One of the immediate advantages of digital versus analog recording for behavioural observation is that digital video no longer requires special hardware for reading time codes, controlling the video or displaying an image or multiple images on the monitor. Most DVR systems have software solutions that negate the need for many traditional hardware requirements in an analog based system, such as time code generators, quad splitters or even time lapse video recorders. In most cases, the DVR can provide all of those things. The advantages and disadvantages of DVR and analog video are outlined in table 3.1.

3.1 Advantages of DVR versus Analog recording

Characteristic	Digital	Analog
Search ability	Instant by date, time or event	Manual search
Storage capacity	Months of footage, low cost	Requirement to buy videotapes
Duplication	No signal or quality loss on recording	Video signal degradation on re-recording or duplication
Image quality	High quality	Poor image quality
Remote monitoring	View multiple locations over internet	Limited monitoring capability
Intelligent features	Motion detection, remote notification	No intelligence

3.2 New features of DVR

The reality is, DVR's are part of a very young technology and new features are continually emerging and the bar is being pushed higher everyday. With constant demand for new features, a powerful foundation is required. But new features mean new unknown variables and accompanying problems.

It is important to remember that the software that is provided with the DVR system is the foundation of the system and at least as important as the hardware. At the heart of any good DVR software program, is a solid database structure. Hardware components aside, the architecture of the software's database is the single most important part of the DVR. How the database stores the video files, what happens when multiple users simultaneously access the database, how does it prioritize? What happens when you have to search through thousands of files or conduct queries? How long does it take to convert or export the files from compressed files, to files that the user can actually use? What file format will ultimately be used with the behavioural software package? These are all important questions that the end user must be aware of.

Motion detection is an exciting new feature of many DVR systems. Motion detection essentially allows for motion detection triggered record function or alarm. Motion detection is a very useful alternative to simple continuous recording. Often continuous recording is used by behavioural scientists because they don't want to miss a behavioural activity or event that is difficult to detect with scan sampling. By only recording when activity is occurring, you can store significantly more video on the hard disk, but more importantly, the researcher does not waste time on playback watching when no activity is occurring. Often the system can be preset to capture a brief time segment prior to the motion occurring, which could be very useful data. Along with motion detection, many systems allow the user to specify an area within the field of view of a camera and perform a search for any motion that occurs within that zone, a useful feature that could be used to refine behavioural tracking with EthoVision.

Another useful advantage of DVR systems compared to other alternatives is the ability to monitor the system remotely over the internet. In order to connect the DVR to the internet from a remote location generally all that is required is an IP address of the location where the DVR is situated (Figure 1, Sentry SMD-04, Sentry Security Systems Inc. 1-1257 Midland Ave, Kingston, Ontario K7P 2Y1 Canada). In addition, many systems not only allow you to remotely view cameras simultaneously, but many systems allow you to manage the main functions of the DVR as well. And finally, cameras can be locally or remotely controlled for panning, tilting and zooming.

The inclusion of diagnostic software, which constantly monitors different functions of the hardware and software of both the DVR and server or other storage medium is also very useful. If the diagnostics detects any deviations from the established baseline criteria, the system can notify the system administrator it has detected a problem. The inclusion of remote diagnostic software which allows the research scientist or a technician to monitor system performance and diagnose and repair potential problems remotely is also very useful. This feature could save on unnecessary service calls to the DVR's physical location, which may be thousands of miles away from support personnel.

Figure 1. *Sentry SMD-04 4 Channel Embedded Mobile Digital Video Recorder being used for acquiring video files for behavioural tracking with EthoVision software. Two 250Gb hard drives are situated at the back to store the video files which can be viewed on one of two monitors. The computer monitor in the center is remotely displaying the arenas over the internet.*

Often the ability to time/date stamp the data file is built in with many systems. In the future, it is very likely that additional data will be readily superimposed on the video files generated from DVR systems. Already POS (Point of Sale) information is being included as imbedded text on the video files in some DVR systems used for security surveillance in the restaurant industry. The ability to include additional data such as physiological or environmental measures will likely become standard features with many DVR systems in the future.

3.3 Important questions to ask about DVR

The biggest obstacle within the DVR industry that researchers need to be aware of is that a number of entities that were already in the security business in the last generation of CCTV technology have attempted to crossover into a world that is far removed from their core competence. DVR systems are primarily an Information Technology (IT) product. Nothing within the last generation of CCTV resembles this, nor does it lend itself to a knowledge base that is readily transferable. The last generation of surveillance equipment depended on mechanical analog-based products, which are in essence fixed machines. As this paper has outlined, DVR's are sophisticated network appliances that are primarily software based. Many companies have entered this business with the best of intentions, but have become quickly overwhelmed. With each new company that enters into the marketplace, another one goes out of business. Unfortunately the number of assemblers and would-be-manufacturers are in the thousands and continues to grow. There are however a number of key questions that a researcher should ask the supplier if considering switching from an analog to a DVR based systems:

- What experience do they have with the product?
- How long have they been in business?
- If you have a problem who do you call?
- How long does it take to get service?
- Where are their service centers located?
- Are the end files that are generated compatible with Noldus behavioural software?

We have attempted to cover the basics in this paper on the use of DVRs for behavioural research, how they work, their components and software as well as related concerns. Based upon the facts presented in this paper, we are hopefull that behavioural researchers will formulate the right questions before acquiring a new DVR based system.

Developing a coding scheme for detecting usability and fun problems in computer games for young children

W. Barendregt , M.M. Bekker
Faculty of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract
This paper describes the development and assessment of a coding scheme to find both usability and fun problems in observations of young children playing computer games during user tests. The proposed coding scheme is based on an existing list of breakdown indication types of the DEtailed Video ANalysis method (DEVAN). This method was developed to detect usability problems in task-based products for adults. However, the new coding scheme for children's computer games should take into account that games are not task-based, that fun is an important factor besides usability, and that children behave differently than adults. Therefore the proposed coding scheme uses eight of the 14 original breakdown indications and has seven new indications. The paper first discusses the development of the new coding scheme. Subsequently, the paper describes the reliability assessment of the coding scheme. The any-two agreement measure of 38.5% shows that thresholds for when certain user behavior is worth coding are different for different evaluators. However, the Cohen's kappa measure of 0.87 for a fixed list of observation points shows that the distinction between the available codes is clear to most evaluators.

Keywords
Coding scheme, usability, fun, children, computer games

1 Introduction
Testing products with representative users is one of the core aspects of user-centered design. A common goal of a user test is to identify parts of a system that cause users trouble and need to be changed. When evaluating computer games for children both usability and fun problems can occur, and both are important to fix. However, no coding scheme of behavior that indicates these problems in computer games for children is available yet. The proposed coding scheme is based on a list of breakdown indication types of the DEtailed Video ANalysis method (DEVAN) [11]. This method was developed to detect usability problems in task-based products for adults. However, the new coding scheme for children's computer games should take into account that games are not task-based, that fun is an important factor besides usability, and that children behave differently than adults. Therefore, the definitions of existing breakdown indications probably need to be changed, new breakdown indications need to be added and some indications have to be removed. The paper starts from the list of DEVAN breakdown indications. Subsequently, the influence of the non task-based nature of games on the coding scheme is discussed. Furthermore, this paper describes new breakdown indications that reflect observed behavior of children indicating problems in games. Finally, the paper discusses how the reliability of the final coding scheme was assessed.

2 The coding scheme
As a starting point for the coding scheme the list of breakdown indication types of the DEVAN method [11] was used. This list is one of the most detailed lists of usability problem indicating behaviors. The original list is given in Table 1.

Table 1Breakdown indications of DEVAN

Breakdown Indication	Description
Breakdown indication types based on observed actions with the product	
Wrong action	An action does not belong in the correct sequence of actions, an action is omitted from the sequence, an action within a sequence is replaced by another action, actions within the sequence are performed in reversed order
Discontinues action	User points at function as if to start using it, but then does not, user stops executing action before it is finished
Execution problem	Execution of action not done correctly or optimally
Repeated action	An action is repeated with the same effect
Corrective action	An action is corrected with a subsequent action, an action is undone
Task stopped	Starts new task, before having successfully finished the current task
Breakdown indication types based on verbal utterances or non-verbal behavior	
Wrong goal	User formulates a goal that cannot be achieved with the product or that does not contribute to achieving the task goal
Puzzled	User indicates: Not to know how to perform the task or what function is needed for it, not to be sure whether a specific function is needed or not
Random actions	User indicates that the current action(s) are chosen randomly
Searches for function	User indicates: Not being able to locate a specific function, to be searching for a function of which the analyst knows it does not exist
Execution difficulty	User indicates: Having physical problems in executing an action, that executing the action is difficult or uncomfortable
Doubt, surprise,	User indicates: Not to be sure whether an action was executed properly, not to

frustration	understand an action's effect, to be surprised by an action's effect, the effect of an action was unsatisfactory or frustrating
Recognition of error or misunder-standing	User indicates: To recognize a preceding error, to understand something previously not understood
Quits task	User indicates to recognize that the current task was not finished successfully, but continues with a subsequent task

2.1 Non task-based nature of games

The list of breakdown indications of DEVAN is aimed at finding problems during user tests with task-based products and thus many breakdown indications on this list also relate to tasks. Since games only have internal goals and no external goals or tasks [9], it may seem that these breakdown indications would not be applicable for games. These internal goals can be considered tasks. By replacing the term 'task' with 'subgame' the indications that refer to tasks can still be used.

Because games are not task-based in the traditional sense it is also unclear what the expected actions are. Therefore the breakdown indications 'Discontinues action', 'Repeated action', and 'Corrective action' are very hard to determine, and they were removed from the list of indications. The breakdown indication 'Wrong action' was defined more clearly in terms of which types of actions could be considered wrong. Clicking on a part of the screen that cannot be manipulated is considered to be a wrong action. Furthermore, actions that are clearly not what the child wants to do, e.g. clicking a button to quit the game before the test is over, are also considered wrong actions.

2.2 Fun

Having pleasure and fun are key factors in a computer game [9], but fun problems are not explicitly covered by the DEVAN breakdown indications. Malone and Lepper's taxonomy for intrinsically motivating instructional environments was used as a starting point to detect fun problems [8]. This taxonomy contains four main heuristics: Challenge, Fantasy, Curiosity, and Control. Based on observations of children playing several computer games [4], we reasoned what verbal or non-verbal behavior children would display if these heuristics were violated.

- **Challenge:** When the provided challenge in a sub(game) is too high a child will want to quit the (sub)game or ask help from the facilitator. The first indication is already present in the original list, asking for 'Help from the researcher' has to be added to the list. When the provided challenge is too low the child may want to stop playing the (sub)game or become bored. The first indication is present in the list; the second indication 'Bored' needs to be added.
- **Fantasy**: When the child is not pleased with the provided fantasy he or she may express dislike. This indication 'Dislike' needs to be added.
- **Curiosity:** The child may signal to be frustrated by a lack of progress or new experiences. This behavior can be detected by the already existing indication 'Doubt, Surprise, Frustration'.
- **Control:** When children cannot control the game, even though they want to, they may show impatience. For example, when long introductions or feedback can sometimes not be interrupted, or when the game is so slow to respond to input that children think it is not reacting. The indication 'Impatience' needs to be added.

2.3 Behavior of children with games

Preliminary versions of the coding scheme were used to code behavior of children playing different adventure type games, for example 'Robbie Konijn' [2], 'Regenboog, de mooiste vis van de zee' [1] and 'Wereld in Getallen 3' [3]. While trying to code these user tests we discovered some behavior that could not be coded with the existing breakdown indications. Further breakdown indications for this behavior were added to the coding scheme. These breakdown indications are given below.

Perception problem*:* Children sometimes complained that they could not hear or see something properly. For example, the goal of a subgame is often explained verbally by one of the characters in the game. In some of the games another character would talk through this explanation of the goal, making it hard to hear. Because a similar situation could happen with visually unclear scenes or objects, 'Perception problem' was added to the list of breakdown indications.

Passive*:* Some children would stop interacting with the game when they did not know how to proceed. They would just sit and stare at the screen. Furthermore, games are often dialogues between the player and some of the characters; the child has to respond to questions and requests of the characters. However, it was not always clear to the children that an action was required of them. The children would remain passive while an action was necessary. Thus, passivity was added as a breakdown indication.

Wrong explanation*:* Sometimes children at first did not seem to have a problem playing a subgame, but later they gave an explanation of something that happened which was not correct and could lead to other problems. For example, in 'Regenboog, de mooiste vis van de zee' children can decorate pictures of fishes with stamps. When a child clicks outside the picture the chosen stamp is deactivated. However, one of the children in our tests clicked outside the picture without noticing it and then said: 'I've run out of stamps! I have to get new ones.' Because of this wrong explanation it was clear that the child did not understand the deactivation of the stamps outside the picture. Giving a wrong explanation of something that happens in the game was added as a breakdown indication.

2.4 Further adaptations

Some of the original breakdown indications have two similar versions, one as an observed action on the product and one as a verbal utterance or non-verbal behavior. With the original coding scheme evaluators usually coded only one of the two versions. Therefore, these breakdown indications were merged into one. This holds for the indications 'Execution problem' and 'Execution difficulty' and for the indications 'Stop' and 'Quit'. Furthermore, it appeared that the distinction between 'Searches for function' and 'Puzzled' was unclear. The indication

'Searches for function' was removed because the 'Puzzled' indication could usually cover these situations.

2.5 The final coding scheme

The final set of proposed breakdown indications to detect both usability and fun problems is the following:

- Indications from DEVAN (some with slightly adapted definitions): Wrong action, execution problem, stop, wrong goal, puzzled, random actions, doubt surprise frustration, recognition of error or misunderstanding.
- New indications: Impatience, wrong explanation, bored, dislike, help, passive, perception problem.

3 Measuring the reliability

To determine the reliability of a coding scheme two commonly used measures are Cohen's kappa, and the any-two agreement measure. Cohen's kappa [6] estimates the proportion of agreement between two evaluators after correcting for the proportion of chance agreement. However, Cohen's kappa is based on each evaluator classifying the same observation points. In the case of free detection of all breakdown indicating behavior not all evaluators may have noticed exactly the same behavior, resulting in different observation points (see Figure 1).

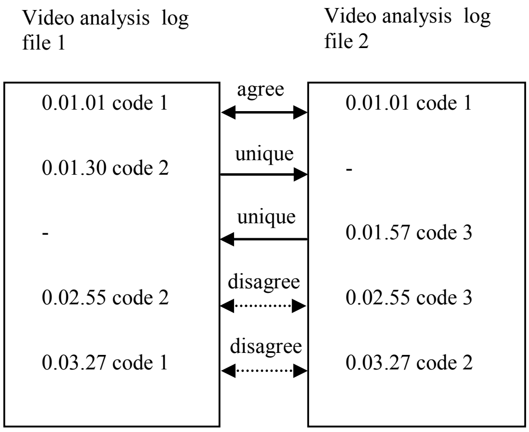


Figure 1 Different observation points from two different evaluators.

Furthermore, Cohen's kappa assumes that the total number of breakdowns that need to be coded is known, or can reliably be estimated. Since it is possible that all evaluators have failed to observe certain behavior this is probably not true. Therefore, we used the any-two agreement measure similar to Hertzum and Jacobsen [7] to determine the reliability of the coding scheme:

$$\frac{\left|P_i \cap P_j\right|}{\left|P_i \cup P_j\right|} (over\ all\ \tfrac{1}{2}\, n(n-1)\ pairs\ of\ evaluators$$

Here, P_i and P_j are the sets of problems detected by evaluator *i* and *j*, and *n* is the number of evaluators. By replacing the sets of problems with sets of breakdown indication type/time pairs, this measure can also be used to determine the agreement of coded observation points.

3.1 Procedure

To determine the any-two agreement for the proposed coding scheme, three evaluators and one of the authors coded a piece of ten minutes videotape of a child playing a computer game called 'Regenboog, de mooiste vis van de zee'. The child was asked to talk as much as possible about playing this game but was not reminded to think aloud. Furthermore, the child was not asked to perform tasks because this is not assumed a representative way of using a game [5].

Before the actual coding the evaluators attended a classroom meeting in which all breakdown indication types were explained. After the explanation the individual evaluators could get a laptop on which the Noldus Observer™ was installed along with the coding scheme. They could also play the game used in the user test to become familiar with it before the coding.

After the evaluators had completed their observations individually, all observations were compared to each other to determine the any-two agreement, number of agreements, disagreements and unique observation points. Observation points that were within four seconds of each other were counted as the same observation points. When two evaluators had the same observation point and the same code at this point, it was counted as an agreement. When one of the evaluators had an observation point and the other did not, it was counted as a unique observation for the evaluator that had coded it. When two evaluators had the same observation point but unequal codes, this was counted as a disagreement.

3.2 Results

Table 2 shows the results of the comparison for each pair of evaluators.

Table 2 Any-two agreement measure, number of agreements, numbers of unique observation points, and number of disagreements for each evaluator-pair.

eval.A x eval. B	Any-two	Agree	Unique A	Unique B	Disagree
1 x 2	50%	37	15	16	6
1 x 3	33%	21	28	9	8
1 x 4	47%	31	25	7	3
2 x 3	27%	21	39	10	8
2 x 4	45%	31	24	6	9
3 x 4	29%	16	17	16	7

The average any-two agreement is 38.5%. This is in the range of any-two agreement measures reported in an overview study by Hertzum and Jacobsen [7], although these numbers were based on problem detection instead of breakdown indication detection. This relatively low percentage is mainly due to the high numbers of unique observations and not to the number of disagreements. This shows that the ability of the evaluators to notice and log interesting behavior is low, while their ability to determine the right breakdown indication type once they agree that there is something going on is high. This is an indication that the codes of the coding scheme are clear to evaluators, but that they use different thresholds for when to code certain behavior as a breakdown indication.

A qualitative analysis was performed to determine causes for the unique observations. A major cause for unique observation points was that sometimes evaluators had not coded all indicating behavior but had made a decision about the severity or multiple occurrence of a breakdown. For example, when a child made the same error more than once, some evaluators had stopped coding this behavior because they reasoned that the problem would be reported anyway. Other unique observation points were caused by unintended additional interpretations of the breakdown indication types, e.g. one of the evaluators had coded 'Recognition of error or misunderstanding' also when a child said something like: 'I have been here before', which is not a recognition of an error or misunderstanding but a recognition in general. Thus, training the evaluators more intensively about how to apply the coding scheme could probably decrease the numbers of unique observations. Furthermore, automatic logging of actions could also reduce the number of unique observations because especially impatient clicking is hard to log manually.

Finally, it was discovered that part of the real disagreements was related to the codes 'Impatient' and 'Wrong action'. When a child clicked an object rapidly and frequently it could be coded as 'Impatient' because it showed impatience or it could be coded as 'Wrong action' because it usually involved an object that could not be clicked (and therefore did not respond, resulting in impatience). Two other codes that lead to disagreement were 'Puzzled' and 'Doubt Surprise Frustration' (DSF). 'Puzzled' is meant for confusion before an action is executed, DSF after an action is executed. However, sometimes it is difficult to determine whether the confusion is before or after an action. For example, incomprehensible feedback can lead to confusion about the performed action but also to confusion about what is expected next. In both cases it is probably not really important which code is used as long as all evaluators notice the behavior of interest.

4 Cohen's kappa for a fixed list of observation points

To determine the extent to which unclear breakdown indication contributed to the low any-two agreement another study was set up in which a Cohen's kappa measure could be calculated properly. From the lists of all four evaluators of the first study a fixed list of observation points was created for use in the second study. When at least three out of four evaluators agreed on an observation point (but not necessarily on the code) it was included in the list of observation points, resulting in a list of 29 fixed observation points.

Two experienced new evaluators received the latest list of breakdown indications with explanations, a list of observation points, the game, and the video data. Independently they had to code all 29 observation points by picking one of the breakdown indications.

Of the 29 given observation points, 26 were coded identically, resulting in a kappa of 0.87. According to the guidelines commonly used for interpreting Cohen's kappa [10], a kappa of 0.87 means that the evaluators showed excellent agreement. For the three points that were not coded identically one of the evaluators had actually given the code of the other evaluator as an alternative code.

These results give a clear indication that the low any-two agreement is not mainly caused by unclear breakdown indicator descriptions but rather by different thresholds when to indicate certain behavior as a breakdown indication.

5 Conclusions

This paper describes the development of a coding scheme for detecting usability and fun problems in computer games for young children. The coding scheme is based on the DEVAN method, and is adapted according to the theory of fun in computer games from Malone and Lepper [8], and observations of children playing games. Six breakdown indications were removed from the original list, and seven were added. The any-two agreement measure of 38.5% for four evaluators using this coding scheme shows that thresholds for when certain user behavior is worth coding are different for different evaluators. However, the Cohen's kappa measure of 0.87 for a fixed list of observation points shows that the distinction between the available codes is clear to most evaluators. Furthermore, in a pilot study not presented in this paper, it was shown that training the evaluators more intensively about how to apply the coding scheme decreases the numbers of unique observations and therefore increases the any-two agreement considerably.

Supported by a grant from the Innovation-Oriented Research Program Human-Machine Interaction of the Dutch government.

References:

1. Regenboog, de mooiste vis van de zee (Rainbow, the most beautiful fish in the ocean) [Computer software] MediaMix Benelux.
2. Robbie Konijn, Groep 3: Pret in de Wolken (Robbie Rabbit, Group 3: Fun in the Clouds) [Computer software] Mindscape.
3. Wereld in getallen, Groep: 3 Het Pretpark (World in numbers, group 3: The Fun-fair) [Computer software] Malmberg Uitgeverij.
4. Barendregt, W., Bekker, M. M. (2004). Towards a Framework for Design Guidelines for Young Children's Computer Games. *Proceedings of the 2004 ICEC Conference*, 1-9-2004, Eindhoven, The Netherlands.
5. Barendregt, W., Bekker, M. M., Speerstra, M. (2003). Empirical evaluation of usability and fun in computer games for children. *Proceedings of INTERACT-03'*, 3-9-2003, Zürich, Switzerland.
6. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37-46.
7. Hertzum, M., Jacobsen, N. E. (2001). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction: Special issue on Empirical Evaluation of Information Visualisations*, **13**, 421-443.
8. Malone, T. W., Lepper, M. R. (1987). Making learning fun: a taxonomy of intrinsic motivations for learning. *Aptitude, Learning and Interaction III Cognitive and Affective Process Analysis*, R.E.Snow and M. J. Farr (Eds.), Lawrence Erlbaum, Hillsdale, N.J.
9. Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R., Fuller, T. (2003). User-centered design in games. *Handbook for Human-Computer Interaction in Interactive Systems*, J.Jacko and A. Sears (Eds.), Lawrence Erlbaum, Mahwah, N.J. pp. 883-906.
10. Robson, C. (1993). Real World Research: A resource for social scientists and practitioner researchers, Blackwell Publishers, Malden, Mass.
11. Vermeeren, A. P. O. S., den Bouwmeester, K., Aasman, J., de Ridder, H. (2002). DEVAN: a detailed video analysis of user test data. *Behaviour & Information Technology*, **21**, 403-423.

Development of a protocol to measure team behavior in engineering education

C. Molano Naffah, S. G. Adams, K. Nagarajarao
Industrial and Management Systems Engineering, University of Nebraska-Lincoln, Lincoln, USA

Abstract
In engineering education, research on performance measurement has characteristically focused on the performance of individuals and has not addressed the measurement of team performance. There is growing research, which indicates that team behavior affects team performance and effectiveness. However, an extension of this research in engineering education could not be found. In an effort to assess and advance the science of measuring teamwork in engineering education, this study concentrates on the development of a protocol to measure team behavior. Results indicate that the protocol, which uses a structured checklist, is moderately reliable. Although some progress has been made, there is still much to be learned about team performance measurement in engineering education. This body of research and the results it produces will be useful to engineering faculty in their efforts to assist students in developing their teaming skills, which will lead to an effective team experience.

Keywords
Team performance and effectiveness, team behavior, Engineering education

Introduction
The use of teams in industry is becoming increasingly more common. There is every indication that it will continue to grow in the future. In many organizations, teamwork has become a critical element since modern task demands are likely to exceed the capabilities of single individual [7]. Such is the importance of teams that all organizations have teams and essentially all employees are members of at least one team [4]. In recent years, business publications and corporate recruiters alike have reported that businesses are increasingly looking for college graduates who can work effectively in teams and understand management processes. In response to this trend, institutions of higher education are incorporating teamwork more frequently in their curriculum in order to facilitate students' adaptation to the corporate environment.

In engineering education, graduates' readiness for teamwork is not meeting expectations. When hiring engineers, employers report that engineering graduates, astute and well prepared technically, lack in their ability to function effectively in teams. Felder, Woods, Stice and Rugarcia report that engineering programs in the United States have been asked by the Accreditation Board for Engineering and Technology (ABET) to modify their curriculum to teach more about teamwork skills [2]. Now, educators in engineering are faced with new challenges including measurement of team behaviors.

Despite the well-recognized significance of team behavior in team performance and effectiveness, few studies have examined behaviors specific to teams in engineering education. Instead, studies have focused on team behaviors in environments such as military [6] and business organizations [5,8]. Therefore, a fundamental gap remains in the understanding of team behaviors that affect team performance and effectiveness in engineering education. Attributes of these studies such as categories of behavior (i.e. dimensions) and the measures used to identify these behaviors may not be applicable for other studies involving teams such as academic and business teams.

The purpose of this research is to develop a protocol to observe team member behaviors in engineering education and to determine if members are behaving effectively. In this study, a structured checklist, derived from direct observation, was used to identify different team behaviors. The advantage of the direct observation technique is the ability to study an event, an institution, a facility or a process in its natural setting [3]. The model used for identifying these behaviors is from earlier research by one of the authors of this paper [1]. Based on this model, team behaviors are categorized into seven distinct constructs: common purpose, clearly defined goals, psychological safety, role clarity, mature communication, productive conflict resolution and accountable interdependence.

Methodology

The Sample
The sample consisted of engineering students from a large Midwestern university who were enrolled in an engineering management course. There were four women and fifteen men all aged nineteen years and older. Three students, (one woman and two men) were Hispanic, one man was African-American and the rest were Caucasian. The participants were grouped in five teams of three to four members. The instructor assigned teams making sure all women and Hispanic students belonged to different groups with the objective of making the teams as diverse as possible. The rest of the class was randomly assigned to a team. The teams were videotaped and observed by three raters twice during the course while working in teams. Training was given to the raters in behavioral observation techniques such as body language of participants and various other operations in the measurement process.

Effective Team Behavior Checklist
Thirty-five behaviors were identified, see Table 1, and divided among seven different checklists/forms corresponding to one of the constructs. The forms were designed to allow each behavior to be rated as it was exhibited by each of the members of a team. The results were given as a function of the behaviors exhibited by all the team members. This was achieved by having a matrix that includes behaviors in rows and a team m ember in every column.

Table 1. Construct and their corresponding checklist items.

Construct	Accountable independence
1. Honestly expressed abilities, capabilities and limitations. 2. Completed a task by the time agreed. 3. Reviewed work with other team members to ensure it satisfies the team's goals, assessed quality of work and proposed improvements if necessary. 4. Changed the way he/she performed the task if the team's needs required it that way. 5. Started working on a different task or helped another team member after completion of the one(s) assigned without waiting to be asked. 6. Assisted another team member accomplishing their task through suggestions, providing direction or sharing useful information.	
Construct	**Conflict Resolution**
1. Stated unconformity and willing to approach it providing clear explanations to others. 2. Listened and explored alternative positions 3. Deviated discussion from issues and addressed the task related conflict. 4. Defined common ground for assessment. 5. Proposed alternatives to try to satisfy both sides of the conflict relating solutions to the purpose of the team. 6. Ensured that all team members are convinced and satisfied with the solution and/or added to consensus.	
Construct	**Mature communication**
1. Listened actively without interrupting 2. Asked questions or clarified what others said to ensure understanding. 3. Articulated ideas clearly and concisely giving compelling reasons for ideas. 4. Provided constructive feedback and specific criticism 5. Spoke loudly and distinctly when communicating information.	
Construct	**Role clarity**
1. Recognized what his/her individual task was. 2. Defined the outcomes expected from an individual member's task. 3. Summarized or agreed upon group decisions on assignments to each member.	
Construct	**Clearly defined goals**
1. Defined goals to attain the common purpose. 2. Proposed a plan to achieve a specific goal. 3. Proposed a time frame to accomplish goals and/or kept team to its agreed time frame and deadlines. 4. Reiterated or asked fro explanation about a goal and confirmed that all other team members agree and understand. 5. Related new information or unfocussed tasks to a specific goal.	
Construct	**Psychological safety**
1. Used we not me or I when referring to the team. 2. Spoke out to propose an idea. 3. Encouraged another team member to voice their opinion. 4. Used positive language and attitude to discuss another team member's idea criticizing ideas not people. 5. Took the time to analyze and evaluate a teams member's idea and carried it out if possible. 6. Included all team members without being biased. 7. Sympathetic and respectful to another team members.	
Construct	**Common purpose**
1. Agreed on a main purpose for the team (proposed it or agreed with proposal of other team member). 2. Exchanged ideas to reiterate and make sure there is a clear understanding of the common purpose. 3. Stated how a task related to the common purpose of the team.	

Each of the forms allowed space for information as it was gathered by the observer. The first space is for the date, which should be recorded to distinguish between observations made on different days. The space below the date identifies the team members, in order to keep the information recorded for every team member in the same column (under the same team member number) throughout all observations. This was important if a researcher was interested in analyzing the behavior of single team members throughout the observations. Space was also provided to record the discipline of every team member, the name of the person performing the observation and the observation length, which are factors that can affect the results of an observation and should be recorded for analysis. A team code and observation number should be assigned to every team observed and to all observations performed, and recorded on the checklist in order to match the forms and the tape containing the group meeting observed. A synthesis of the instructions on how to fill out the form is also provided. The last column on the checklist was provided to calculate the results of every observation. Given the wide spectrum of behaviors that can be displayed it was possible that a behavior that describing a construct was not included in the corresponding form. Extra space was given in every form for these behaviors to be recorded if a rater finds any during an observation.

Scoring the Observation

Items on the checklist were scored using a Likert scale ranging from 0 to 2. This study used a nominal scale in which every behavior would be rated with a 0,1,2 or N/A, representing behaviors not exhibited, somehow exhibited, highly exhibited and not applicable behaviors respectively. This type of rating was used because the behavior exhibited is a variable that depends on the frequency of occurrence of the behavior. For example, if a behavior was present consistently throughout the observation the rater would score it as a 2. Whereas, if it was slightly present it was assigned a score of 1 and if it was not exhibited at all it received a 0. It was important to understand that the frequency itself was not a factor with a fixed scale but depends upon the kind of situation itself.

Under these same terms, there were behaviors that were not present in an observation because the situation did not allow for such behaviors to occur. In this case, using zero as a score would indicate that the team members lacked the ability to exhibit this behavior when applicable. This would not be a valid interpretation of the situation. For this reason N/A was been among the rating options. To obtain the results of a single observation the rates (0, 1 and 2) were totaled for every behavior. In this manner a sum was obtained for every behavior (row) on each of the seven forms. The score for every behavior was obtained by dividing the sums by the number of members in the team. Behaviors that were rated N/A were assigned a total score of N/A. This procedure was followed with every behavior for every form for all observations.

Calculating the Final Score

Once all observations were performed, it was necessary to summarize all the scores. The scores were assigned in the row that described the behavior and under the column of the observation for which the score was obtained. Make sure to record the team number in the space provided at the top of the form as well as the construct being scored. Once all the scores were summarized on the Score Calculation Sheets the rows were all totaled to obtain the total sum of scores for a behavior throughout observations, and averaged by dividing them by the number of observations. N/A behaviors in an observation were not be considered in the sum or average. In this case, only those with an applicable rating were added and divided by the number of applicable behaviors (the ones included in the sum). After calculating these averages for every behavior the final score was calculated. The averages were totaled and then divided by the number of behaviors that defined the construct. The result was used to compute the Overall Construct Score. This number was multiplied by 100% and divided by 2 (the maximum possible score). In this manner we obtained a percentile that was used to interpret the results. Table Two shows an example on how to perform the calculations detailed above.

Table 2. Final Score Calculation Example

Ques.	Ob 1	Ob 2	Ob 3	Sum	Score
1	2	2	n/a	4	2
2	.333	N/A	1	1.333	667
3	0	.333	0	.333	111
4	0	1.667	2	3.667	1.222
				Total	4
				Average	1
				Total Score	(1 X 100)/2
				Final Result	50%

A final result was computed for each construct using the same procedure. These percentages represented the extent to which each of the seven constructs was present in a team. For example, 70% scoring in mature communication indicates that there is still room for improvement in this area. The tool can be described as an assessment of the extent to which each of the seven constructs is present in the team based on what is exhibited by the team member's behavior.

Reliability was tested using inter-rater reliability. Inter-rater reliability defines the capacity of the protocol to provide similar results from different observers in the same observation. Three observers were used to complete the forms after videotapes were obtained from group working sessions in an Engineering Management course. The scores obtained by the three raters were compared and analyzed using Cohen's Kappa for statistical analysis to determine the inter-rater reliability of the tool.

Results and Discussions

A combined inter-rater reliability rate (by taking the average of both observation results) was computed and findings show a 62% rate of agreement between raters. As mentioned in the previous section, the results provided by this checklist are percentages relative to each of the seven constructs. While this is lower than the desired 70% it is encouraging and we can conclude that the protocol is moderately reliable but requires significant improvement.

The low reliability is attributed to a few checklist items that were problematic. According to feedback from the raters, these items were either: 1) related to tasks that were not complete enough for the team members to exhibit most of the behaviors or 2) related to tasks in which participants' characteristics were more related to collaborative groups than teams. Consequently, the raters found it difficult to identify behaviors specified by these items and also to interpret and score such behaviors. Hence, a significant inconsistency in rating prevailed for certain behaviors among the raters.

Conclusion

In this study a behaviorally based protocol was developed with the objective of measuring team behaviors specific to engineering education. The protocol provides an assessment of the extent to which a team exhibits each of the seven constructs in the team effectiveness model. The findings of the study conform moderately to the reliability expectations of the protocol. The data and the feedback from raters suggest that rater training in addition to more observations is necessary to achieve higher value of reliability.

The impact of this study is far-reaching for students and faculty members from a professional development perspective. Faculty members can use behavioral protocols to enhance their knowledge about issues to consider in managing student teams. By using this model of team effectiveness (specific to their field of study), faculty members can integrate teamwork into core curricula. Using the research findings, resources and appropriate training can be subsequently developed for students working in team projects.

References

1. Adams, S.G.; Simon Vena, L. C.; Ruiz Ulloa, B.C.; and Pereira, F. 2002. A Conceptual Model for the Development and Facilitation of Effective Teamwork, *2002 American Society for Engineering Education Proceedings*, Montreal, Canada, June 2002.
2. Felder, R M.; Woods, D.R .; Stice, J.E. and Rugarcia, A. 2000. The Future of Engineering Education Ii. Teaching Methods That Work. *Chemical Engineering Education* **34 (1)**: 26-39.
3. Kumar, K. 1991. Rapid Data Collection Methods for Field Assessment. In *Team Planning Notebook for Field-Based program assessment* USAID PPC/CDIE.
4. McCanus, K. 2000. Do You Have Teams? *IEEE Solutions,* **32 (4).**
5. Meister, D. 1985. *Behavioral Analysis and Measurement Methods*. New York, NY: Wiley.
6. Morgan, B B, A S Glickman, E A Woodard, A S Blaiwes and Eduardo Salas. 1986. Measurement of Team Behaviors in a Navy Environment. In *Naval Training Systems Center* Orlando, Florida: Naval Training Systems Center, Human Factors Division.
7. Salas, E.; Dickinson, T.L.; Converse, S.; and Tannenbaum; S.I. 1992. Teamwork From Team Training: New Directions. Paper read at the *Proceedings of the 7th Interservice/Industry Training Equipment Conference and Exhibition*, at Orlando, Florida.
8. Schifflet, S.; Eisner, E.J.; Price, S.J.; and Schemmer, M. 1985. The Definition and Measurement of Small Military Unit Team Functions (research Report 1397) Fort Benning, GA: Training research laboratory, ARI Field Unit.

Recurrence plot analyses suggest a novel reference system involved in spontaneous newborn movements

B. Aßmann[1], M. Thiel[2], M.C. Romano[2], and C. Niemitz[1]
[1]Institute for Human Biology and Anthropology, Free University Berlin, Germany
[2]Institute for Nonlinear Dynamics, University of Potsdam, Germany

Abstract

Newborn movements are well-studied in terms of reflexes, muscle synergies, leg coordination and also early forms of target-directed arm/hand movements. As former approaches concentrated mainly on separate accomplishments there remained, however, a clear need for more integrated investigations. Here we report an inquiry that explicitly was guided by such a perspective. Such was given by a methodological concept of Ilan Golani, which distinguishes three classes of behaviour: (a) staying at a homebase, (b) leaving this base for a so-called excursion, and (c) returning to the homebase. In newborn movements, the homebase was given by limb/body positions which alternated with special movements, but also occurred as both the origin and the target of these movements. Symbolic dynamics and recurrence plot analyses of kinematic data received from audiovisual newborn recordings yielded new insights into the temporal organization of limb movements. Thereby, our approach uncovered a novel reference system of spontaneous newborn movements.

Keywords

Spontaneous newborn movements, kinematics, reference system, symbolic dynamics, recurrence plot

1 Introduction

Compared to mature skilled actions of adults and older children, which are smooth and precise, the movements of newborn and young infants lack the smooth temporal and spatial integration of coordinated actions. Limbs, head and trunk seem to move as random elements, without reference to each other. Peiper (1963) described this overall impression of disorganization by saying that these ‘mass movements’ are ‘awkward and abrupt and follow each other without connection’.

However, there are many examples, in which newborn actions show some elements of coordination, meaning that various muscle groups work together rather than as independent elements. These include well-known newborn reflexes such as the Moro, grasp, palmar, and plantar reflexes [1,3], functional actions such as rooting, sucking, and swallowing [4,14], hand mouth synergism [11], reaching towards objects [10] and the coordination of leg movements in spontaneous kicking [16].

These results on coordinative features in the motor behavior of newborns motivated a more integrated investigation of the coordination of the overall system. New methods from non-linear dynamics combined with inspiration from models from animal exploratory behavior allow a new understanding of the intrinsic dynamics of the unintenional movements of human infants. To establish a commen language between biologists, psychologists and systems theory scientists it is necessary to obtain a notation system of behavior which corresponds to the natural morphology of behavior and is informative, parsimonious and of predictive value. Common movement patterns, that are masked by ‚ordinary language’ used in the terminology of Biology and Psychology, can be revealed by a suitable notation system [8]. According to the results of a recurrence plot analyses, we applied symbolic dynamics [9] to transform the kinematic data into a simple but informative form of notation. These methods from non-linear dynamics revealed a reference system involved in the behavioral motor organization that shows parallels to a methodological concept developed by Ilan Golani to discuss rat exploratory behavior as a natural manifestation of spatial learning [2,6,7]. When rats are placed in a novel environment, they typically establish a homebase. This is the place where they stay for the longest cumulative time and the number of visits is especially large [6].

In the case of spontaneous newborn movements, the homebase was given by limb/body positions which alternated with special movements, and seemed to be both the origin and the target of each movement. From this perspective, recurrence plot analyses of the newborn movements suggests a referencesystem that reminds - according to the above mentioned criteria - of homebase behavior and provides a basis to discuss the role of a reference system in newborn spontaneous movements in terms of concepts of spatial learning. Our study was designed to test whether and how new methods and concepts from different research fields can come together and shed light on old topics, and even suggests a reference system of static body/limb configurations indicating some kind of coordination in newborn spontaneous movements.

2 Methods

2.1 Data sampling and measurment

Subjects: Subjects were 6 apparently normal, full-term infants, 3 boys and 3 girls. Two subjects, children 2 and 3 were dizygotic twins. Subjects were recruited in a maternity clinic to participate in a ‘mini-longitudinal’ study of the development of newborn spontaneous movements over 3 to 10 days, depending on how long the mothers stayed at the clinic.

Procedure: The design of these observations was a frame-by-frame movement microanalysis using videotape recording. To videotape awake infants, parents were asked to bring their infants to the laboratory between two anticipated feedings. Kinematic data collection used videorecording with three cameras (50Hz) in a volume calibrated with a calibration frame. Infants were undressed and allowed to move spontaneously in the supine position for 20 min. There were no specific stimuli presented nor was the spontaneous posture of the infants controlled.

Videotape selection: For the kinematic data analyses, we chose videotapes from each of the six infants between 5 and 20 minutes from 2-3 different days. Restrictions were

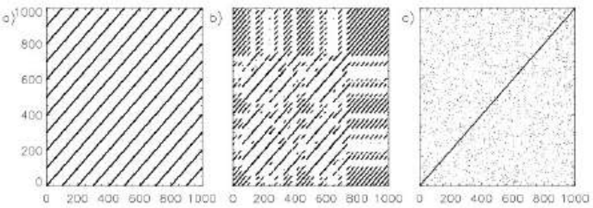


Figure 1. a) RP of a periodic function, b) RP of a chaotic System (the Rössler system with standard parameters), c) RP for random white noise

that the sequences showed continous motor activity without resting or crying intervalls longer than 20 seconds.

Kinematic data: Movement kinematics were analysed with APAS. Every 12 frames, the 2- dimensional positions of movement relevant joints were tracked by a mouseklick on the screen in each perspective. The APAS Software allows by triangulation to compute from multiple two-dimensional frames a 3-dimensional picture. Movement relevant joints were: nose, shoulders, elbows, hands, hips, knees and feet. Joint angles for shoulders, elbows, hips and knees were calculated from the coordinate data. Since the joint angle displacement of the elbow and shoulder of one arm and those of the hip and knee of the same leg displayed 80% correlation (data not shown), the movements of each limb can be captured by the displacement of one angle. The arms are defined by the angle of the elbow, the legs by the kneeangle.

2.2 Evaluation I: Symbolic dynamics

Symbolic dynamics is a natural way to describe data which appear as sequences of discrete states. This approach is based on a coarse-graining of the dynamics; i.e. the time series are transformed into symbolic sequences by using very few symbols. This way one loses some amount of detailed information, whereas some of the invariant, robust properties of the dynamics are kept [9]. To describe the bodyconfigurations of the newborns, we introduced a two-symbol encoding into the kinematic data. On the basis of the histogramms in figure 2, each limb was defined into two states: either the limb was streched, or it was considered to be angled. In terms of the arms, an elbowangle shorter than 60° degrees was regarded as angled, higher than 60° degrees as streched. Considering the legs, a kneeangle shorter than 120° degrees was reckoned as angled, higher than 120° degrees as streched. From the combination of four angles with two possible states respectively, there are 16 differently defined configurations.

2.3 Evaluation II: Recurrence plot analyses

This method was first introduced to visualize the time dependent behavior of a dynamical system, which can be represented as a trajectory $x_i \in \mathbf{R}^n$ ($i = 1, \ldots, N$) in a n-dimensional vector space [5]. It represents the recurrence of the trajectory in phase space to a certain state, which is a fundamental property of dynamical systems [13]. The main step of this visualization is the calculation of the NxN matrix, $\mathbf{R}_{i,j} := \Theta(\varepsilon - \| x_i - x_j \|)$, $i,j = 1, \ldots, N$, where ε is a cut off distance, $\| \cdot \|$ is the maximum norm, and $\Theta(\cdot)$ is the Heaviside function.The binary values in $\mathbf{R}_{i,j}$ can be

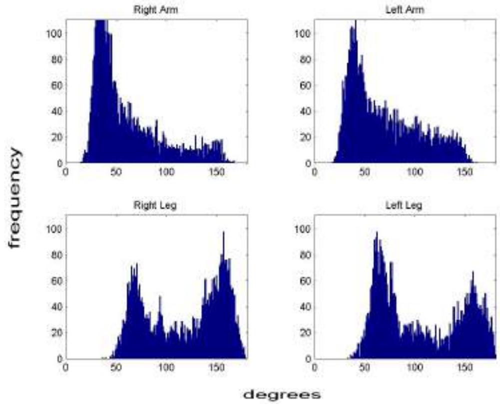


Figure 2. Frequency distribution of the joint angles of the single limbs.

simply visualized by a matrix with the colours black (1) and white (0). The recurrence plot (RP) exhibits characteristic large-scale and small-scale patterns that are related to typical dynamical behavior [17]. For a periodic signal of period T, the plot looks like Figure 1a for very small ε. This is a series of stripes at 45 degrees, with stripes separated by a distance of T in the vertical and horizontal directions. The RP of a chaotic system (Fig 1b) has a more complicated structure. Brief episodes of parallel stripes at 45 degrees are hints of almost periodic trajectories. For the RP of white noise (Fig. 1c), such a structure is not evident. It consists of mainly single points, indicating the randomness of the system. In all RPs, there is a stripe along the diagonal corresponding to i = j. While diagonals indicate similar evolution of different parts of the trajectory, horizontal and vertikal black lines show, that the state does not change for some time.

To capture the spontaneous movements of the newborns, the configurational state of their body is defined by the four joint angles of the limbs. Four values of the angles belong to each samplepoint of the time series. For a samplepoint i the vector $x_i = (w_1(i), w_2(i), w_3(i), w_4(i))^{\mathbf{T}}$ is generated, with $w_{1\text{-}4}(i)$ being the values of the four joint angles at samplepoint i. The same procedure is applied for the samplepoint j: $x_j = (w_1(j), w_2(j), w_3(j), w_4(j))^{\mathbf{T}}$. These two vectors depict the state of the system at samplepoint i and j with four angles respectively. In the RP the time series i is on the horizontal axis and j on the vertical axis. A black dot indicates, that the state of the configuration in x_i and x_j are the same (according to the value of ε). White dots represent different configurations.

3 Results

In our study, the spontaneous motor behavior of newborns was analysed on the bases of the movements of the four limbs, i.e. arms and legs. For this purpose, we first focussed on single limbs and then examined how this behavior assembled into overall movement patterns. The following report starts with results of symbolic dynamics and recurrence plot analyses referring to a 20 minute movement sequence of one child.

3.1 Motor behavior of the single limbs

The frequency distribution of the segmentangles in which the limbs reside, reveal one or two scopes in which the limbs stay for a longer cumulative time than in other domains. Figure 2 shows for the arms a unimodal shape

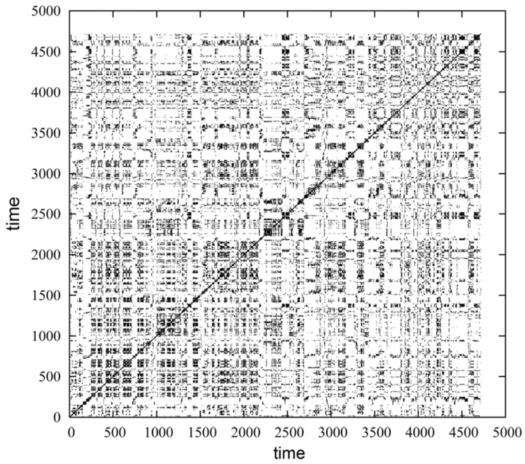


Figure 3. *Recurrence-plot of the integrated joint angle displacement timeseries with embedding dimension m=4 and* ε*=30.0. The horizontal and vertical axis display the timeseries* x_i *and* x_j*, respectively.*

with a major peak in the lower band between 20 and 60 degrees; the legs show a bimodal shape with equal peaks: one in the lower scope between 50 and 90 degree and a higher band between 130 and 170 degrees. This distribution indicates, that the arms are predominantly located in a angled position and relativly seldom in a streched position, while the legs are with a comparable frequency located in an angled or streched position, whereas the intermediate scope (90-130 degrees) is comparatively seldomly found. This means that the legs mostly move between the angled and the streched position and rest in these scopes and rarely on the way in between. The unimodal distribution of the arms implicates that the arms mostly move out of the angled position back to the angled position and exhibit only one resting position.

4.2 Recurrence plot analysis

From the combination of the angle positions of the four limbs at each samplepoint arise certain configurations, in which the body resides. The dynamic of the distribution of these configurations can be visualized by a recurrence plot (RP). Figure 3 presents the RP of the motor behavior of a newborn child defined by the described configurations. The RP is characterized by rectangles on a time scale of 10-15 samplepoints (3 seconds), which shows, that the system is determined by static states.

Such structure results from the black rectangles being piled upon each other and interrupted by white rectangles. This indicates that configurations which are taken very often (black) are interrupted by configurations that are taken rarely. Some configurations reveal in parts of the plot predominently black, in other areas white sections. That means, that the relevant configurations in given sections are taken very frequently (black) while in other sections they are hardly taken (white).

The superior structure is demonstrated by the rectangular structure on a higher timescale of 400-500 points of measurement. On this timescale alternations between different configurations become apparent and show that 2-3 configurations can serve as a reference system. Absent diagonal structures reveal, that there are no recurring movementsequences or sequential recurring successions of configurations.

Another measure of RPs is the length of vertical structures. The length of black vertical lines shows how long a particular configuration is taken, the length of the white vertical lines specifies the interval between the recurrence of a configuration. Both the length of the black and the white lines does not display predominant values if calculated over the overall time series (data not shown).

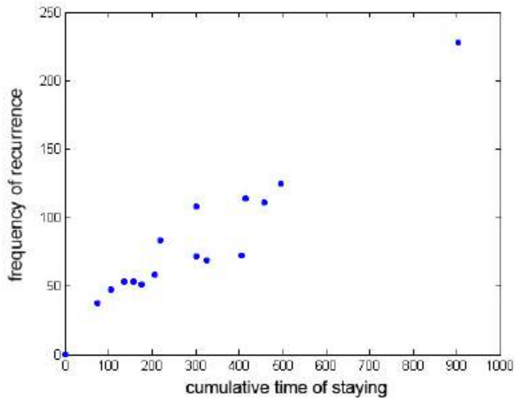


Figure 4*. Correlation of the cumulative time of staying in a given configuration versus the frequency of recurrence.*

3.3 Frequency distribution of the configurations

The RP demonstrates that decisive configurations are taken very long and often, while others are taken much less frequently and with shorter duration. For further quantification we introduced on the bases of Figure 2 symbolic dynamics, so that each limb is defined by a either angled or streched state. From this two-symbol coding of four limbs result 4^2 defined configurations. By the use of this technique, the studied movements can be described as follows: limbs are moved singularily or simultanously together whereby they take one of the 16 configurations, mostly for a varying span of time. The cumulative time of staying in a given configuration plotted against the frequency of recurring to the same configur-ation show a distribution of datapoints along a diagonal, which presents a high correlation between these two parameters (Fig. 4). This correlation shows that those configurations which are taken the longest cumulative time are the same ones that are taken with the highest frequency. Furthermore, the plot demonstrates that there is one configuration that clearly shows the highest values.

4 Discussion

Our analyses of motor behavior uncovered a preferred configurational position of single limbs displaying confined scopes of joint angles. Whereas the unimodal distribution of the arm positions in a lower scope of angles is biomechanically plausible, the bimodal shape of the distribution of the angles of the legs in the low and high scope of angles appeares somehow remarkable. Here, the peaks in the distributions of leg- and arm-related data can be understood as reference points in the movement patterns of limbs.

In addition, our analyses have shown that all configurations with the longest cumulative time of staying are at the same time those with the highest frequency of recurrence. This phenomenon reminds of the concept of homebase behavior, developed by Ilan Golani (1989) for studies of the exploratory behavior of rats placed in a novel environment. Crucial characteristics of the home-base are the longest cumulative time of staying in this place and the highest number of visits. In spite of the clear differences between the two accomplishments, i.e. the locomotion of a rat and the spontaneous movements of a newborn, there are parallels that suggest considering the special static states or configurations of newborn movements as analogous bases in which the system can settle and start new movements.

The RP of the kinematic data shows, that (i) the dynamics of the system can be characterized by static states that (ii) certain configurations are taken especially often in defined time windows, characterizing these time windows by certain combinations of configurations and (iii) that the durations of staying and the intervals of recurrence of decisive configurations are variable. Referring to these features, the RP indicates, that combinations of configurations in given time windows act as homebase configurations. This reveals another parallel to the homebase behavior of rats, which can establish several transient homebases [15]. The self-similar aspect of the RP on different time-scales suggests an existence of principles of organisation that could play a role at different hierarchical levels. On a lower level, single limbs seem to indicate reference points to which they return. On a higher level, however, the organization of configurations seems to be more complex and related to association groups that require a special reference system. We tend to presume that the combinations of reference configurations in the spontaneous newborn movements can be considered as groups of temporally associated configurations, and also that the spontaneous movements can be characterized by a structural hierarchy. Such presumption requires further inquiries into the temporal diversity of intervals within and between the various association groups, which can be tested e.g. by analysing the length of vertical structures in the RP.

Since all individuals show one or several, but different reference configurations which even can change over time, the tendency to establish these reference points appears quite notable. We assume that these features do not simply reflect physiological properties or constraints of the body, and thus tend to explain them as an endogenous component of the motor behavior. Such a view would, however, require further inquiries into the reference system of static configurations and the dynamic part of the spontaneous movements, as well. Future investigations could be done by analysing the temporal parameters represented by the length of the vertical structures in the RP in reference to the homebase configurations.

Our finding that the spontaneous movements are structured by static configurations and that the newborns return to specific configurations could be explained also by a mechanism that supplements the model cited above. That is, there could be some kind of memory system which captures the spatial relationship of the limbs and provides a reference frame to return into this configuration. The mechanism that allows the animal to return to its homebase is called path integration system [2,7], which permits position and direction to be updated solely on the basis of ideothetic information. McNaughton (1996) proposed an intrinsic, two-dimensional manifold within a high-dimensional neuronal representation space in which locations are defined by stable patterns of neural activity and in which there are orderly proximity relationships among the locations. The preconfiguration of a state space topology of a two-dimensionally organized system of stable attractors would provide a robust mechanism for the spontaneous, off-line reactivation of recent experience that is thought to be necessary for memory consolidation [12]. It seems rewarding to assume that also newborns may use their physical body to examine or develop such a two-dimensionally organized system of attractors. Support for this assumption can be derived from our evidence that multiple reference configurations can be coded on the bases of angular (two-dimensional) motion signals. With this as a reference, we like to advertize the methodological concept introduced here also for an application in other studies on newborn spontaneous movements.

Supported by Berliner Programm zur Förderung der Chancengleichheit für Frauen in Forschung und Lehre.

References:

1. Barnes, M.R.; Crutchfield, C.A.; Heriza, C.B. (1978). The neurophysiological basis of patient treatment. *Volume 11: Reflexes in motor development*. Morgantown, WV: Stokesville.
2. Biegler, R., & Morris, R.G.M. (1996). Landmark stability: studies exploring whether the perceived stability of the environment influences spatial representation. ***J Exp Ecol*, 199**, 187-193.
3. Capute, A.J.; Accardo, P.J.; Vining, E.P.G.; Rubenstein, J. E.; Harryman, S. (1978). *Primitive reflex profile*. Baltimore: University Park Press.
4. Crook, C.K. (1979). The organization and control of infant sucking. *Adv Child Dev Behav*, **14**, 209-252.
5. Eckmann, J.P.; Kamphorst, S.O.; Ruelle, D. (1987). Recurrence plot of dynamical system. *Eur Phys Let.* **4**, 973-77.
6. Eilam, D.; Golani, I. (1989) Homebase behavior of rats (R. norvegicus) exploring a novel environment. *Behav Brain Res*, **34**, 199-211.
7. Etienne, A.S.; Maurer, R.; Séguinot, V. (1996). Path integration in mammals and its interaction with visual landmarks. *J Exp Biol,* **199**, 201-209.
8. Golani I. (1992) A mobiliy gradient in the organization of vertebrate movement: The perception of movement through symbolic language. *Behav Brain Sci*, **15**, 249-308.
9. Graben, P. beim; Kurths, J. (2003) Detecting subtreshold events in noisy data by symbolic dynamics. *Phys Rev Lett* **90** (10), 100602.
10. Hofsten, C. von (1982). Eye-hand coordination in the newborn. *Dev Psychol*, **18**, 450-461.
11. Korner, A.F.; Beason, L.M. (1972). Association of two congenital organized behavior patterns in the newborn: Hand-mouth coordination and looking. *Perceptual and Motor Skills,* **35**, 115-118.
12. McNaughton, B.L.; Barnes, C.A.; Gerrard, J.L.; Gothard, K.; Jung, M.W.; Knierim, J.J.; Kudrimoti, H.; Qin, Y.; Skaggs, W.E.; Weaver, K.L. (1996). Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *J Exp Biol*, **199**, 173-185.
13. Ott, E. (1993). Chaos in dynamical systems. Cambridge University Press, Cambridge.
14. Peiper, A. (1963). *Cerebral function in infancy and childhood*. New York: Consultants Bureau.
15. Tchernikowsky, O.; Benjamini, Y.; Golani, I. (1995). Constraints and the Emergence of 'free' exploratory behavior in rat ontogeny. *Behav*, **133**, 519-539.
16. Thelen, E.; Fisher, M. (1983). The organization of spontaneous leg movements in newborn infants. *J Motor Behav*, **15**, 353-377.
17. Webber, C.L. Jr; Zbilut, J.P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *J Appl Physiol*, **76** (2), 965-73

Comparing activity scores by Actiwatch® tied to leg and arm in preschool children

R. Juyal[1,3], C.P Khokhar[3], S. Sazawal[1, 2], U. Dhingra[2], P. Verma [1], A. Sarkar [1], P. Dhingra [1], M. Osmany[1], V.P. Menon[1], M. Black[4], R.M. Angulo-Kinzler[5] and R.E. Black [2]
[1]*Center for Micronutrient Research, Annamalai University, India*
[2] *Johns Hopkins Bloomberg School of Public Health, Baltimore, USA*
[3] *Departments of Psychology, Gurukul Kangri University, Haridwar, India*
[4] *University of Maryland School of Medicine, Baltimore, Maryland, USA*
[5] *Division of Kinesiology, Center for Human Motor Research, University of Michigan, USA*

Abstract
Present study evaluated qualitative and quantitative differences in measurement of activity by actiwatch applied to arm or leg. In 126 children aged 13-69 months, two actiwatch one tied on arm and the second on leg recorded for one hour, for the same time period, activity of the child was recorded by direct observation. Time spent with activities was estimated in 1 min intervals to coincide with activity scores of 1 min epochs estimated from Actiwatch. Overall leg scores were found to be nearly twice as high as arm scores. Results of regression analysis indicated that leg actiwatch score was associated with 6 out of 18 gross motor activities, while arm actiwatch score was associated with 12 of 18 gross motor activities. For leg related movements score in arm actiwatch is half that of leg actiwatch but for arm related movements score is higher in arm monitor. Using a total score a sum of arm and leg provide a better estimation of overall physical activity.

Keywords
Actiwatch, Young children, Leg activity, Arm activity, The Observer.

Introduction
Physical activity and active exploration play an important intermediary role in a child's developmental process Quantification of levels of physical activity of young children is important and apart from direct observation, accelerometers have recently been used in measuring voluntary physical activities. Zhang et al[1] reported that accelerometer using sensors on the thighs and chest (trunk) could be used to assess the duration of gross physical activities. Studies on site of placement of accelerometer have been conducted on adults or older children using either hip or leg as two sites of recording activity. To our best knowledge no data is available for young children and for comparing arm and leg, the two logical places to tie the new smaller accelerometer devices. In this study we compared the leg and arm activity scores quantitatively as well as qualitatively against actual observation of types of activities which were recorded using observer software.

Methods
Subjects and Setting
Study was conducted in Sangam Vihar, a peri urban locality of New Delhi. The study was performed as an add on to an ongoing clinical trial in which physical activity of children was being assessed. In a sub sample of 126 children aged 13-69 months one hour actiwatch record and observation for activity was undertaken at home of the child.

Actiwatch
Actiwatch (Mini-Mitter co. Inc) is an accelerometer that is capable of sensing any motion with a minimal resultant force of 0.01g/rad/s. Actiwatch stores this information as an activity counts. The weight and dimensions of this device are minimal (27×26×9 mm), weight (17g) and durable casing looking more like a watch. However, it is an omni directional, resulting in sensitivity to motion in all direction.
Epoch length is the period of time actiwatch will accumulate activity counts before saving the sample and setting the counter to zero. In this study 1 min epoch length were set in all the actiwatches. There is an event recorder that was used to indicate the start and end of the observation.

The Observer Software
The Observer 5.0 (Noldus Information Technology B.V), an automated manual event recorded installed in a laptop for collecting, collating and analyzing the data. We configured the software as per our activity coding scheme which consisted of five broad areas Gross Motor activity, Fine motor activity, Behavioral states, Contact with care giver and Location. **Gross Motor activity consisted of 16 activities** (Being carried, Lying, sitting squatting, hanging, crawling, running, climbing, rough-tumble, dancing, kneeling, stand with support, walk alone, stand alone, swinging and walk with support**), Fine motor consist 12 activities** (eating, drinking, self care, dressing, game with little movement, game with active movement, mild movement with object, vigorous movement, repetitive movement, No movement, out of view and other fine movement) **Behavioral states consisted of 3 states**(Sleeping, alert and fussy), **Contact with care giver consisted of 5 states**(Beyond arm's length, within arm's length, Holding hands, Clinging and touching with hand) and **Location consisted of 4 states**(Home restricted, home non-restricted outside restricted and outside non-restricted) respectively. The codes within each of the 5 groups were mutually exclusive such that start of one activity automatically marked termination of the previous activity. One hour observation synchronized with activity recording using two watches was performed.

The Observer training
Observer scientists were trained to record the activities of the children. Videotape and group discussion were used for the training session which continued till inter observer agreement of more than 80% was established.

Procedure

Prior appointments were taken before visiting the child home for activity assessment. Activity assessments were done in the natural environment of the child i.e. at the home of the child. A trained scientist gave necessary instruction to the caregiver and loaded the identification information (user Identity information and numbers, age, gender), ensure data and time which automatically changes was right and set epoch length. Actiwatch was tied on arm and leg of the child for 1hour. Simultaneously the same identification information was loaded in the observer software. As soon as the marker was pressed on the actiwatch the observer scientist also started the coding. Time of actiwatch and laptop was synchronized. Activities of the child were coded using Observer 5.0 (Noldus Information Technology) with the help of laptop for one hour.

Analysis

Actiwatch data was downloaded with the help of actimeter. Using a in-house developed software in visual basic the activity score, start time and end time of each epoch was appended to a oracle data base which had child ID information already. From direct observation, we selected an interval of 1 minute to match the epoch of acti watch and within that estimated time spent in each activity codes. For each interval time spent in 16 gross motor activities and 12 fine activities was estimated these data were also appended to Oracle data base. Data were analyzed as means ± SD of the scores. Pearson correlation and multiple regression analyses were performed using STATA, version 8 (Stata Corp, Union Station, TX).

Results

Out of 16 gross motor activities (i.e. being carried, lying, sitting, squatting, hanging, crawling, running, climbing, rough & tumble, dancing, kneeling, stand with support, walking alone, standing alone, swinging and walk with support.) 6 were (i.e. crawling , running , climbing , walking alone , standing alone and walk with support) was significantly associated with leg activity scores (Table-3) while arm actiwatch score was significantly associated with squatting , hanging , crawling , running , climbing , rough & tumble, dancing, kneeling , walk alone , standing alone , swinging , walk with support (Table-1).

In another model, 12 fine motor activities (eating, drinking, self care, dressing, game with little movement, game with active movement, mild movement, no movement, other fine movement, out of view, repetitive movement and vigorous movement) as independent variable were regressed with leg and arm actiwatch score. Out of 12 fine motor activities 6 were significantly associated with leg actiwatch score i.e. game with little movement , game with active movement , mild movement , out of view , repetitive movement , vigorous movement (Table-4), while 10 fine activity scores are significantly associated with arm actiwatch score i.e. self-care, dressing, game with little movement, game with active movement, mild movement with objective, no movement, other fine movement, out of view, repetitive movement, vigorous movement (Table-2). Among epochs with more 50% time spent in activities associated with 6 gross motor activities associated with leg monitors, mean activity score in leg monitor was 896.29±1070.99 compared to arm monitor 419.82±460.56. Among the epochs with more than 50% time spent in arm related activities , leg monitor scores were 190.93±381.89 compared to 278.97±364.84 for arm .

Table 1. *Association between Gross Motor activities and Arm actiwatch score.*

Activites	Beta values	Sig. levels
D_beingcar	-0.010	0.587
D_lying	-0.002	0.919
D_sitting	0.046	0.355
D_squatting	0.073	0.001
D_Hanging	0.028	0.025
D_Crawling	0.160	0.000
D_Running	0.307	0.000
D_Climbing	0.029	0.023
D_Roug_tum	0.068	0.000
D_Dancing	0.056	0.000
D_Kneeling	0.031	0.050
D_Stanwsup	0.018	0.663
D_Walkalon	0.224	0.000
D_Stanalon	0.079	0.038
D_Swinging	0.022	0.069
D_Walk_wsu	0.045	0.002

Table 2. *Association between Fine Motor activities and Arm actiwatch score.*

Activites	Beta Values	Sing. Levels
D_eating	0.004	0.894
D_drinking	-0.026	0.174
D_self_car	0.024	0.098
D_dressing	0.034	0.044
D_gamw_lit	0.236	0.000
D_gamw_act	0.108	0.000
D_mmovw_o	0.132	0.022
D_nomovmen	-0.025	0.084
D_othefinm	0.043	0.084
D_outview	0.153	0.000
D_repmov	0.092	0.000
D-vigmov	0.144	0.000

Table 3. *Association between Gross Motor activities and leg actiwatch score.*

Activites	Beta Values	Sing.levels
D_beingcar	-0.010	0.587
D_lying	-0.002	0.919
D_sitting	0.046	0.355
D_squatting	0.073	0.001
D_Hanging	0.028	0.025
D_Crawling	0.160	0.000
D_Running	0.307	0.000
D_Climbing	0.029	0.023
D_Roug_tum	0.068	0.000
D_Dancing	0.056	0.000
D_Kneeling	0.031	0.050
D_Stanwsup	0.018	0.663
D_Walkalon	0.224	0.000
D_Stanalon	0.079	0.038
D_Swinging	0.022	0.069
D_Walk_wsu	0.045	0.002

Table 4. *Association between Fine Motor activities and Leg actiwatch score.*

Activites	Beta Values	Sign. Levels
D_eating	0.046	0.137
D_drinking	-0.026	0.161
D_self_car	0.005	0.745
D_dressing	0.025	0.121
D_gamw_lit	0.122	0.013
D_gamw_act	0.082	0.000
D_mmovw_o	0.124	0.022
D_nomovmen	-0.018	0.208
D_othefinm	0.003	0.888
D_outview	0.129	0.000
D_repmov	0.047	0.000
D-vigmov	0.114	0.000

Discussion

The results of this study indicated that the Actiwatch tied to leg records more of activities associated with leg movements, while Actiwatch tied to arm records more broader range of movements including leg movements. Overall movement scores for leg monitor is significantly higher then arm monitor. Study conducted by Kumahara et al[2] found that the acceleration of the upper limbs that includes fidgeting is more elevated than that of the whole body for sitting/lying down activities. Acceleration data of the wrist and arm are known to improve recognition rates of upper body activities[3,4]During tasks involving movement from sitting to standing the upper body movements still had a higher intensity than the whole body movements whereas their contribution to energy expenditure is lower than whole body trunk movements. This was reversed for walking activities. Similarly Tamura et al[5] reported that wrist accelerations were observed when the subject was in sitting position, with waist acceleration close to zero. In another study[6] it showed that arm movements during sitting –like activities had larger interindividual variability than locomotion activities such as walking or jogging. Levine et al[7] reported that even very slow walking can increase energy expenditure almost two-fold compared to resting level. In contrast, mixed sedentary activities with little gross body movements involved increasing energy expenditure above resting by less than 50% on average. This study indicated that recording of various activities leads to a different score in leg monitor as compared to arm monitor. If only one monitor is to be placed seems arm would be a better place to do it, but ideally using two monitors and estimating total score provides a better assessment of activity.

The study was approved by the human research review committees at the Center for Micronutrient Research, Annamalai University, and the Johns Hopkins Bloomberg School of Public health, and the WHO.

References:

1. Zhang K, Werner P, Sun M, Pi-Sunyer FX, Boozer CN.(2003) Measurement of human daily physical activity. Obes Res; 11: 33-40.
2. Kumahara H, Tanaka H, Schutz Y.(2004 September) Daily physical activity assessment: what is the importance of upper limb movements vs whole body movements? Int J Obes Relat Metab Disord;28(9):1105-10.
3. G.S. Chambers, S. Venkatesh, G.A.W. West, and H.H. Bui.(2002) Hierarchical recognition of intentional human gestures for sports video annotation. In Proceedings of the 16th International Conference on Pattern Recognition, volume 2, pages 1082–1085. IEEE Press.
4. F. Foerster, M. Smeja, and J. Fahrenberg. (1999)Detection of posture and motion by accelerometry: a validation in ambulatory monitoring. *Computers in Human Behavior*,15:571–583.
5. Tamura T, Fujimoto T, Sakaki H, Higashi Y, Yoshida T, Togawa T.(1997 May) A solid-state ambulatory physical activity monitor and its application to measuring daily activity of the elderly. J Med Eng Technol;21(3-4):96-105
6. Sugimoto A, Hara Y, Findley TW, Yoncmoto K. (1997 March)A useful method for measuring daily physical activity by a three-direction monitor. Scand J Rehabil Med;29(1):37-42
7. Levine JA, Schleusner SJ, Jensen MD.(2000 December) Energy expenditure of nonexercise activity. Am J Clin Nutr;72(6):1451-4.

Objective and continuous measurement of limited mobility and / or limited upper limb activity

F.C. Schasfoort, J.B.J. Bussmann, H.J. Stam
Department of Rehabilitation Medicine, ErasmusMC, University Medical Center, Rotterdam, the Netherlands

Abstract
Our research aims at the development, validation and application of ambulatory monitoring devices that allow objective, continuous, and long-term measurement of activity of healthy and disabled subjects. A recently developed instrument is the Upper Limb-Activity Monitor (ULAM), which is based on ambulatory accelerometry and consists of body-mounted sensors connected to a portable waist-worn data recorder. Acceleration data are stored digitally and downloaded onto a computer for automated kinematic analysis after the 24h measurements, and the ULAM allows objective measurement of the mobility-related activities lying sitting, standing, walking, cycling, and general (non-cyclic) movement, and, more importantly, it measures activity of both upper limbs *during* performance of the mobility-related activities. The ULAM can be used to objectively determine activity limitations in subjects with upper limb disorders, which will be illustrated by the results of several studies with the ULAM in patients with unilateral upper limb Complex Regional Pain Syndrome type I.

Keywords
Ambulatory accelerometry, upper-limb activity limitations

Introduction
The goal of Rehabilitation Medicine is regaining and/or maintaining functionality by decreasing the consequences of a disease or disorder. With that aim, it's primary focus is the patients' activity (limitations) and human behaviour during normal daily life. This makes monitoring devices for continuous, long-term and (preferably) objective measurement fundamentally important. Feasible, reliable and valid instruments that objectively measure during everyday behaviour are essential to provide insight into activity limitations of patients.
One instrument that has been developed in our department is the Upper Limb-Activity Monitor (ULAM). It consists of several body-mounted uni-axial piezo-resistive acceleration sensors that are connected to a portable waist-worn data recorder (see Figure 1). Data are stored digitally and downloaded onto a computer for automated kinematic analysis, based on Signal Processing and Inferencing Language (SPIL) routines. The ULAM allows objective measurement of the mobility-related activities, the body postures lying sitting, and standing, and the body motions walking, cycling, and general (non-cyclic) movement. More importantly, the ULAM allows objective measurement of activity of both upper limbs *during* performance of these mobility-related activities, and can therefore also be used to determine (limitations of) mobility and activity limitations in upper limb disorders. Its primary outcome measures are the intensity, percentage, and proportion (= ratio) of upper limb activity during sitting and standing.
The ULAM was initially developed for patients with upper limb Complex Regional Pain Syndrome type I (CRPSI). CRPSI is a disorder with controversial pathophysiology. When it occurs, it usually follows surgery or trauma and is generally expressed in the limbs (about 50% upper limb). CRPSI comprises a combination of sensory, trophic, autonomic and motor impairments which traditionally are the main focus in CRPSI research and practice. Activity limitations accompanying CRPSI were generally considered important but hardly quantified, and if quantified, merely scales or questionnaires were used [1].
The aim of this paper is to describe the development and characteristics of the ULAM, and second, to illustrate the applicability of this measuring technique by discussing the results of 4 studies with the ULAM in upper limb CRPSI.

Figure 1. *Two subjects wearing the ULAM configuration.*

Development and characteristics
The ULAM is an extended version of its ´older brother´ the Activity Monitor (AM) [2]. Because the AM was insufficient to detect activity limitations that are characteristic of upper limb disorders, we validated the use of additional accelerometers on the forearm which resulted in the ULAM. The ULAM consists of five (or six) ADXL202 accelerometers (Analog Devices, Breda, the Netherlands): one sensor lateral on the thigh halfway between the spina iliaca anterior superior and upper side of the patella (sensitive direction sagital plane), two sensors on the sternum (sensitive direction in sagital and longitudinal plane), and two sensors on each forearm just proximal from the wrist joint (sensitive direction in sagital plane being in the anatomical position).
Raw acceleration signals are expressed in g (9.81 ms $^{-2}$) and are a combination of two components: gravitational acceleration and accelerations due to activity. The magnitude of these components depends on the extent and direction of the accelerations with regard to the sensitive direction. The raw signals (32 Hz) were stored digitally on a PCMCIA flash card, and downloaded onto a PC for analysis after the measurements.
For detection of mobility-related activities, three feature signals are derived from each raw acceleration signal: the angular, motility and frequency feature (time resolution 1s). For each activity and for each feature signal, a maximum and minimum value is pre-set. Each second, the

‘distance’ from the actual feature signal value to the pre-set range is calculated for each feature signal from each sensor. The mobility-related activity with the lowest total distance is detected.

To detect upper limb activity, the motility feature was used. This feature is created after zero-phase finite impulse response high pass filtering (0.3-16 Hz) of the raw acceleration signal, rectifying and averaging over 1 second. The value of this signal depends on the variability of the raw signal around the mean and is also expressed in g (9.81 ms^{-1}). The variability of the raw upper limb acceleration signal can be regarded as a measure for upper limb activity; the more the value is varied, the higher the motility value, the more the upper limb activity. To determine whether an upper limb was active, it was automatically determined whether the motility values of an upper limb sensor exceeded a preset threshold assigned to the mobility-related activity that was performed during that second. Each second the motility values exceeded the threshold, the ULAM signal for the upper limb forearm was positive, indicating upper limb activity. Each second the motility value did not exceed the threshold, the ULAM signal for the upper limb forearm was zero, indicating upper limb inactivity.

In a feasibility study, the ULAM’s ability to discriminate between upper limb usage and non-usage in healthy and disabled subjects during normal daily life was assessed [3]. Based on our definition of upper limb usage (i.e. active movement of (parts of) the upper limb(s) in relation to proximal parts, holding and leaning) and a framework of different forms of upper limb usage, an activity protocol was compiled that represented normal daily life upper limb usage or non-usage. Video recordings were used as a reference method and agreement scores between ULAM data and videotape recordings were calculated. The ULAM data of special interest for rehabilitation medicine were detected satisfactorily (overall agreement 83.9%), and there were no systematic differences between healthy and disabled subjects for different forms of upper limb usage or non-usage. Although the ULAM did not allow valid measurement of every aspect of upper limb (non-)usage, its use was considered feasible for future application studies on upper limb activity in upper limb CRPSI.

Clinical application studies

For each clinical application study ethics approval was obtained. Measurements of each 24 hours were performed in the natural environment of patients with unilateral upper limb CRPSI and no other co-morbidity. The primary ULAM outcome measures were the intensity, percentage, and proportion (= ratio) of upper limb activity during sitting and standing, respectively.

The first study determined the long-term impact of CRPSI on everyday activities by comparing objectively measured '24-h ULAM activity profiles' of ten chronic CRPSI patients with ten healthy comparison subjects [4]. It appeared that the general mobility of subjects with CRPSI in their non-dominant upper limb was not affected by CRPSI. CRPSI in the dominant upper limb had modest impact on general mobility, however. For the ULAM outcome measures related to upper limb activity there were marked differences between CRPSI patients and control subjects, although less obvious during standing than during sitting. Especially patients with dominant side involvement clearly significantly less activity of their involved limb during sitting. It was concluded that chronic CRPSI patients still have objectively measurable limitations in upper limb usage during everyday life.

The second study analyzed the relationship between impairments and activity limitations to address questions as: ´does an impairment always lead to activity limitations?´ and ´which impairment particularly affects everyday activity?´ [5]. Six impairments were measured in thirty chronic CRPSI subjects: 1) temperature differences between both hands, 2) a visual analogue scale to measure pain resulting from effort, 3) the number of pain words from the McGill Pain Questionnaire to assess pain during the previous days, 4) a goniometer to determine differences in maximum active range of motion (AROM) within pain threshold of the wrist and fingers between both hands, 5) volumeter fluid overflow as a means to determine volume differences between both hands, and 6) a portable hand-held dynamometer to assess differences in grip strength between both upper limbs. These 30 CRPSI patients also wore the ULAM for 24 hours. It was found that impaired active range of motion, grip strength, and to a lesser extent pain resulting from effort were the most important impairments explaining variance in activity limitations. It was concluded that the more impairments a subject had, and especially motor impairments, the more objectively measured activity limitations were present.

Because the ULAM measurement technique clearly differs from what is commonly used in CRPSI research and clinic with respect to several methodological and practical criteria, the third clinical application study determined how the ULAM outcome measures were related to four questionnaires that also aim to assess the functional / behavioural consequences of diseases [6]. In a cross sectional comparison study, thirty patients with chronic CRPSI were measured with the ULAM and they also completed four questionnaires. The questionnaires used were two generic questionnaires, the 68-item Sickness Impact Profile (SIP68) and the RAND 36-item Health Survey (RAND36), and two body-part specific questionnaires, the Disabilities of Arm Shoulder and Hand questionnaire (DASH) and the Radboud Skills Questionnaire (RASQ). Spearman rank correlations were calculated between the outcome measures. It appeared that 83% of the inter-questionnaire correlations were significant, whereas 46% of the correlations calculated between the ULAM and the questionnaires were significant. The number and strength of the correlations between the ULAM and questionnaires was dependent on the degree to which the same aspects of functioning were measured. The relationships between ULAM and questionnaire outcome measures generally appeared to be non-significant or weak, whereas more often significant and stronger relationships were found between questionnaire scores. It was concluded that the ULAM measures similar aspects of functioning / behaviour only to a certain extent, and it measures – at least partly – different areas of functioning and health. The more characteristics the ULAM has in common with other instruments, the stronger and more often significant relationships were found.

In the fourth clinical application study, we explored upper limb activity over time in four subjects with acute CRPSI [7]. In this study, we compared the upper limb activity time course as measured with the ULAM to the time course of other outcome measures for activity (limitations) and impairments (i.e. three questionnaires: RASQ, DASH, RAND36, and six impairment outcome indicators: VAS-momentary pain, VAS-pain resulting from effort, volume, temperature, AROM, strength). The subjects were measured at four moments in time during a treatment protocol. Objectively measured upper limb activity frequently improved; improvements of >5% were found for 63% of the ULAM outcome measures at final assessment. The ULAM outcome measures had a time course more similar to the body-part and CRPS1 specific questionnaire RASQ than the other questionnaires. The time course of impaired temperature was most often in accordance with the ULAM, and both VAS scores showed least accordance. The relationships between the time courses of the ULAM outcome measures and the time courses of other outcome measures for activity limitations and impairments appeared logical and for a great deal explainable. These results, in our opinion, demonstrated that the current ULAM has potential to objectively and validly assess changes in upper limb activity over time in future longitudinal (treatment efficacy) studies.

Conclusion

We have shown that the ULAM has added value for research and clinical practice in Rehabilitation Medicine. It also has potential as an outcome measure for a large number of other disciplines and research fields when measuring behaviour is important. Of course, the technique has some limitations; there are some practical and methodological issues that need careful consideration and further study. Development and validation of ambulatory accelerometry devices like the ULAM is an ongoing process. If unobtrusive, objective and valid measurement of mobility-related activities and/or upper limb activity during everyday life in a subject's personal environment is required, instruments like the ULAM should be considered.

References:

1. Schasfoort F, Bussmann J, Stam H. (2000) Outcome measures for complex regional pain syndrome type I: an overview in the context of the international classification of impairments, disabilities and handicaps. *Disabil Rehabil* **15, 22(9)**, 387-98.
2. Bussmann J, Martens W, Tulen J, Schasfoort F, van den Berg-Emons H, Stam H. (2001) Measuring daily behavior using ambulatory accelerometry: the Activity Monitor. *Behav Res Methods Instrum Comput* **33(3)**, 349-56.
3. Schasfoort F, Bussmann J, Stam H. (2002) Ambulatory measurement of upper limb usage and mobility-related activities during normal daily life with an upper limb-activity monitor: a feasibility study. *Med Biol Eng Comput* **40(2)**, 173-82.
4. Schasfoort F, Bussmann J, Zandbergen A, Stam H. (2003) Impact of upper limb complex regional pain syndrome type 1 on everyday life measured with a novel upper limb-activity monitor. *Pain* **101(1-2)**, 79-88.
5. Schasfoort F, Bussmann J, Stam H. (2004) Impairments and activity limitations in subjects with chronic upper-limb complex regional pain syndrome type I. *Arch Phys Med Rehabil.* **85(4)**, 557-66.
6. Schasfoort F, Bussmann J, Stam H. (2005) Correlation between a novel upper limb activity monitor and four other instruments to determine functioning in upper limb complex regional pain syndrome type I. *J Rehabil Med* **37(2)**, 108-14.
7. Schasfoort F, Bussmann J, Stam H. (2005) Upper limb activity over time in complex regional pain syndrome type 1 as objectively measured with an upper limb-activity monitor: An explorative multiple case study. *Eur J Pain* In press

The use of Eshkol-Wachman Movement Notation (EWMN) reveals disturbances in movement in infancy in autism but not in schizophrenia

O. Teitelbaum, P. Teitelbaum, A. Prince
Psychology Dept., University of Florida, Gainesville, Florida, U.S.

Abstract
Using Eshkol-Wachman Movement Notation (EWMN) [2,3] to analyze movement patterns, we compared home videos of 17 Autistic-to-be-infants and 16 Asperger's-to-be infants with those of 2 Schizophrenic-to-be infants. All Autistic and Asperger's infants showed movement disturbances, whereas Schizophrenic-to-be infants showed no movement disturbances.

Keywords
EWMN, Autism, Schizophrenia, infancy.

1 Introduction
Autism was long considered to be a juvenile form of Schizophrenia. Recently our work using EWMN has shown in all cases studied so far (n=33) that quite profound movement disturbances exist in infancy in Autism [5] and Asperger's [4]. We have applied this methodology to the study of Schizophrenic-to-be infants. So far (n=2) Schizophrenic-to-be infants were found to be perfectly normal in their movements. This emphasizes a possible profound difference in infancy between Autism and Schizophrenia.

2 Methods
We obtained videos of 33 Autistic and Asperger's infants which were taken during the first year of life. These videos were analyzed frame by frame using EWMN to reveal possible movement disturbances.
EWMN is a movement language. Whereas the use of other measurement devices (such as computerized movement measurement systems) permits a high standard of *recording* of movement, EW "creates a system of concepts that serves as a guide in the bewildering chaos of perceptions so that we learn to grasp general truths from particular observations". [1]. The basic concepts of the EWMN system are: 1) the relation of a spherical System of Reference (SoR) to the limb segments of an abstraction of the body; 2).the classification of circular paths of movement into three types; 3).the representation of the human body on a manuscript page.

3 The Spherical System of Reference (SoR)
The movements of a single axis of constant length free to move about one fixed end will be enclosed by a sphere. The free end will always describe a curved path on the surface of this sphere. *Every limb in the body can be regarded as such an axis*. Typically, the curves described on the surface of the sphere will be circles or parts of circles. In order to define these data, coordinates dividing the surface of the sphere – like the geographical globe – form an intersecting network of 'lines of longitude' (vertical circles) and 'lines of latitude' (horizontal circles). One direction on the horizontal plane is selected as the starting position for all measurements. This direction is labeled **absolute zero**. By measuring off intervals of 45 degrees, eight positions are obtained. Vertical planes are perpendicular to the horizontal plane. The line in which all vertical planes intersect is the vertical axis of the sphere. By stating the horizontal component and the vertical component any *position* in the SoR can now be defined. Numbers are written in parentheses, the vertical component above and the horizontal component below. (See fig. 1)

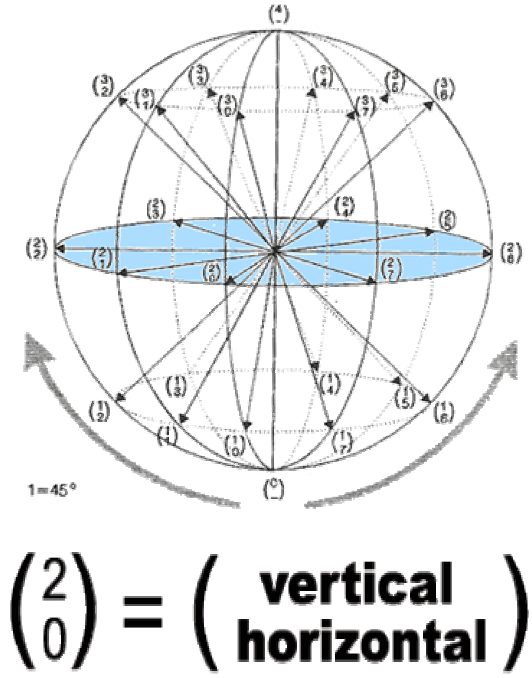


Figure 1. *The spherical System of Reference (SoR) with an example of how to write a position..*

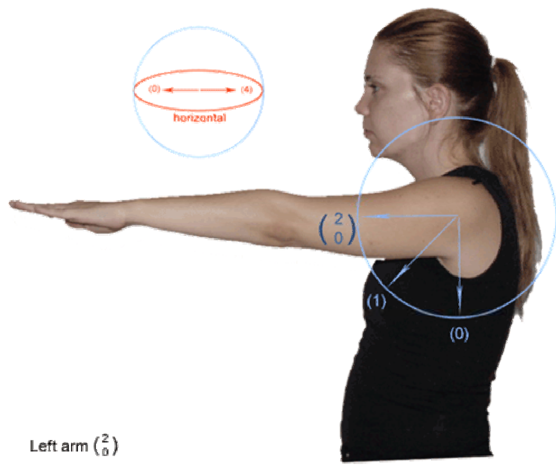


Figure 2. *Integration of the abstract SoR with the physical body.*

4 The Body

To establish one general form that will stand conceptually for all bodies, an entirely abstract body image is proposed: a 'man without qualities', in which the individual shapes of the parts of the body are *temporarily* disregarded. Each limb is reduced to its longitudinal axis (the axis of the limb) – an imaginary straight line of unchanging length.
In EW a limb is considered to be any part of the body which lies between two adjacent joints, or a joint and a free extremity. Therefore, a limb is not solely considered to be an arm or a leg.
The axes are articulated with one another in a manner analogous to the skeletal structure of the body. (See fig. 3)

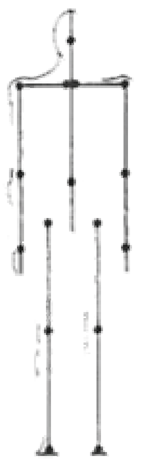


Figure 3. *An abstract image of the body in which each limb is reduced to its longitudinal axis (the axis of the limb).*

5 The Law of 'Light' and 'Heavy" limbs

The structure of the body is dealt with as a branching linkage. The base is conceived as the heaviest segment of the body; the anatomically adjacent segment is lighter, moving in relation to the heavy limb, and carried by it.
The terms 'light' and 'heavy' do not refer to the actual weight of the limbs, but are figurative. However, the phenomenon itself is real and is an inevitable property of the human body in movement. (Fig. 4)

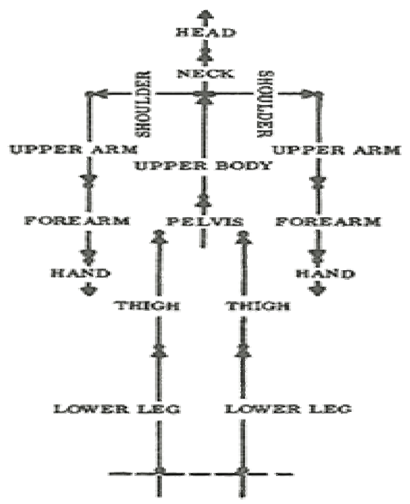


Figure 4. *The base of the body is considered to be its heaviest segment.*

6 The Tilt Test

When babies, who were later diagnosed with Autism or Asperger's lost their balance, either when they were sitting, or when they were walking, they fell 'head on' to the floor. The protective reflexes which should have prevented injury to the head didn't appear when they should have. In terms of pure movement description, the head was passively carried by the movement of the heavy limb (in this example – the torso). In contrast, when normal babies lost their balance, the head maintained its verticality, i.e. it was not passively carried along by the movement of the torso.
Based on those observations, we devised the "Tilt Test", a simple, quick and non-invasive test for possible damage in the vestibular system. It is a clear and simple example of the use of EW law of 'light' and 'heavy' limbs.: A baby at the age of six to eight months is held around his/her waist in an upright position. The baby is facing outward (his back is towards the caregiver) (see fig. 5 top). From that position, when tilted slowly 45 degrees to each side, the head of a normal healthy baby will maintain its verticality, i.e. it *should not* be carried passively along by the movement of the torso, its adjacent heavy limb (see fig. 5 bottom). When, as in autistic and Asperger's infants, the head *is carried passively* by the torso, it doesn't maintain its vertical orientation with respect to gravity. This signals possible damage to the vestibular system. Previously we have shown [6] that the Tilt Test can be used as an early marker for the possible detection of Autism.

Figure 5. *Top: A normal baby being held upright.*
Bottom: The Tilting Test in a normal baby. Note that the head remains vertical when the body is tilted.

7 Classification of Circular Paths of Movement into Three Paths

Positions in EW are defined by identifying them with coordinates of the SoR. Movements of limbs are also defined, oriented and measured in relation to this SoR.
The basic ‘unit’ of movement for a single limb is the circle. When the limb describes a circle, an *imaginary* line passing through the center of any such circle and perpendicular to its plane is said to be the *axis of movement*. The angular relationship between the *axis of movement* and the moving limb results in 3 types of movement.

1. When the relationship is 90 degrees, a *plane movement* is produced. Plane movement is the shortest distance traveled between two positions on the SoR..

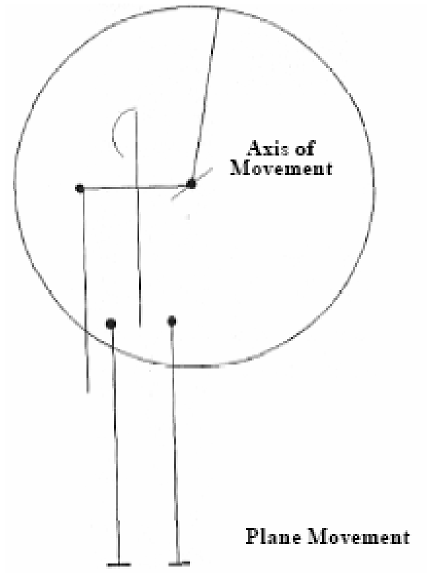


Figure 6a. *Plane Movement.*

2. When the relationship is greater than 1 degree and smaller then 90 degrees, the resulting path has a *conical shape*.

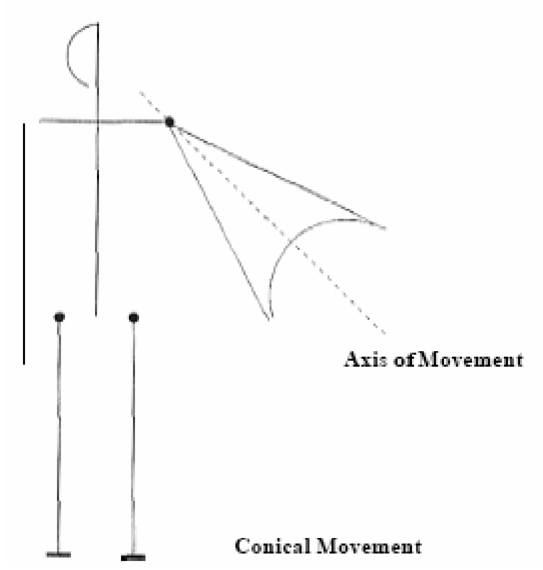


Figure 6b. *Conical Movement.*

3. When the relationship between the moving limb and the axis of movement is 0 degrees the limb doesn’t change its place in space; it rotates around its own axis. This type of movement is called *rotation.*

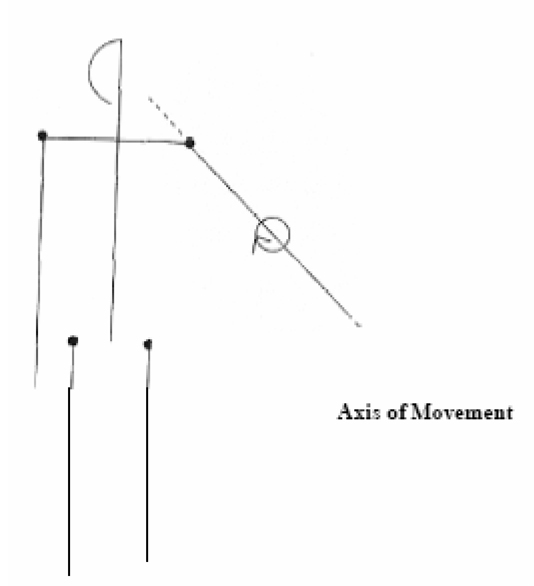


Figure 6c. *Rotation.*

The underlying pattern of all movements of the human body is made up of these three types of movement – whether they appear alone or in pairs (Plane movement with Rotation etc.) For a more complete presentation of the principles of EW, see [2,3].

8 Righting

Righting is the process in which a baby, from around the age of 4 to 6 months, is able to turn by himself from lying supine to lying prone. In a normal healthy baby this is accomplished by a corkscrew rotation of the trunk. In Autistic-to-be infants such rotation is absent. In contrast they lift the head and the pelvis sideways upward in a plane movement. Therefore it appears that Autistic babies have profound deficits in such rotation. [5].

9 The representation of the human body on a manuscript page

Movement Notation is designed to express the relations and changes of relation between the parts of the body, and information which can be derived from these. In the analysis of movement, we treat the body as a system of articulated axes. A horizontally ruled page represents the body. The spaces between the lines are assigned to the parts of the body whose movements are to be represented. Vertical lines divide the manuscript system into columns denoting units of time, and the symbols for movements are written in order, from left to right.

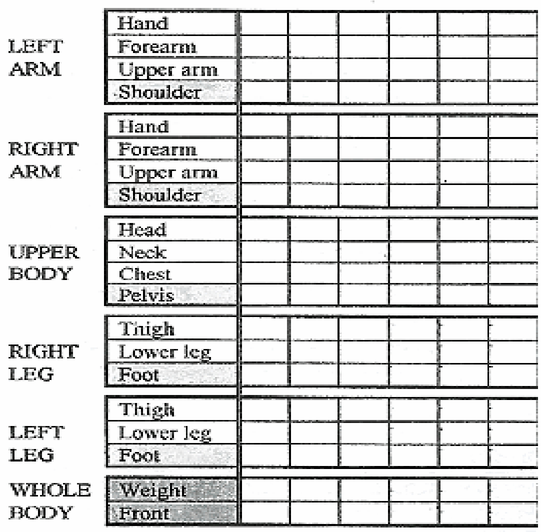


Figure 7. *Scheme of the EW Manuscript Page.*

10 Schizophrenic Babies

In the videos we have seen so far, none of the above symptoms were present in babies who were later eventually diagnosed with Schizophrenia. There is a puzzle here. In the present paper, we have mentioned only two early symptoms of movement disturbances in Autism. Other symptoms such as Moebius mouth, persistent asymmetry, idiosyncratic patterns of crawling and walking also appear in Autism [4, 5] but were completely absent in the Schizophrenic-to-be babies we have observed. One possibility is that if more Schizophrenic-to-be babies are studied such movement disturbances may appear in some of them. Alternatively, our findings suggest that the locus of damage in the brain is different in Schizophrenia from that in Autism.

References:

1. Einstein, A. (1956). *Out of my later years*. pp. 113.
2. Eshkol, N.; Harris, J. (2001). *EWMN Part I.*
3. Eshkol, N.; Harris, J. (2004). *EWMN Part II.*
4. Teitelbaum, O.; Benton, T.; Shah, P.; Prince, A.; Kelly, J.; Teitelbaum, P. (2004). *Proc. Natl. Acad. Sci. USA*. **101**, 11909-11914.
5. Teitelbaum, P.; Teitelbaum, O.; Nye, J.; Fryman, J.; Maurer R. (1998). *Proc. Natl. Acad. Sci. USA*. **95**, 13982-13987.
6. Teitelbaum, P.; Teitelbaum, O.; Fryman, J.; Maurer, R. (2002). *The Journal of Developmental and Learning Disorders*. **6**, 15-22.

Studying usage of complex applications

J. Kort [1], M. G. D. Steen [2], H. de Poot [2], H. ter Hofte [2], I. Mulder [2]
[1] *Mobile: User Centered Innovation, TNO Information and Communication Technology, Groningen, The Netherlands*
[2] *Mobile: User Centered Innovation, TNO Information and Communication Technology, Delft, The Netherlands*
[3] *Telematica Insituut, Enschede, The Netherlands*

Abstract

Innovative products and services are characterized by an increased complexity in technology and functionality (see e.g. product development and research in areas like personal networks, intelligent home environments, ubicom and CSCW). This Increased complexity also poses new problems and challenges for usage studies. Firstly because current methods and tools are not always applicable or too resource intensive to apply in the context in which the product or service is used. Secondly, current methods and tools don not provide the desired insight. They are too focused on task performance and usability issues, while research interests have changed and broadened to include context and user experience, leading to implementations of for example personalization, context awareness, adaptive and pleasurable services. In this paper we present the conceptualization and design of new tools for usage research which address and might solve current problems in usage research in complex application.

Keywords

Usage, usability, user experience, quantitative measurements, qualitative measurements.

1 Introduction

New developments in technology create a very complex design environment [1]. “Always on” is gradually becoming reality with mesh or ad-hoc, hybrid, personal (area) and sensor networks. Standardized transmission techniques and data transmission protocols enhance data flow between different networks and applications, resulting in service bundles or distributed services (e.g. Live Contacts [2]). Last but not least, developments in hardware technology enable devices supporting more and more functionality.

These technological advancements lead to a very complex design environment, offering many different choices to develop and implement increasingly complex products or services for users. Technical delimiters appear less frequent and important and the question “what should we design?” has gained more attention [3]. The increased product or service complexity in the absence of technical delimiters has many effects on User Centered Design (UCD) and evaluation. Trade-offs between sample size, data transparency, real world validity, and effort, create complex relationships between product development goals, evaluation methodologies and issues of resources and scope [4].

These changes in, and challenges posed on UCD and their possible solutions were discussed during the CHI2005 workshop “Usage Analysis: Combining logging and qualitative methods” [1]. The main conclusion being:

“The most fruitful direction to develop solutions for dealing with usage research and evaluation in UCD of complex system is to be a combination of quantitative and qualitative methods, that can be applied more unobtrusive and on a larger scale, in situ”.

Freeband User Experience (FRUX), strives to develop and implement these different qualitative and quantitative solutions in a toolkit, the ‘SocioXensor’. This year’s focus of SocioXensor is on qualitative and quantitative data gathering. Qualitative data gathering is hereby defined as data gathering in which a participant under study takes an active role (e.g. providing answers to questions). Quantitative data gathering is defined as data gathering in which neither the researcher nor the participant takes an active role.

In section 2, the changes and challenges in UCD research and evaluation are explained in detail, according to the UCD cycle used within FRUX. Section 3 addresses solutions to the challenges for data gathering, currently developed and implemented in SocioXensor. Section 4 concludes this paper with a short summary and topics for future research in the coming years.

2 Challenges in UCD

FRUX’s design cycle has many common elements with other UCD design cycles described in literature, especially with the one defined by Cooper [5]. It recognizes the following phases: idea generation; concept design; design; implementation; evaluation & pilot; and the launch & thereafter (this last phase is currently not part of the project). Each of these phases has its own specific research interests, methods, activities and results which a researcher or designer can use according to good iterative practice. Challenges met in the different UCD phases can be categorized in three distinct groups: Those before the implementation phase, during the implementation itself and those during the evaluation. The challenges are briefly described according to these groups, in terms of research interests, methods and encountered problems.

2.1 Before the implementation

Every UCD project starts with ‘understanding the user in his or her context’ (idea generation) [5]. From this ‘understanding’ different solutions to user problems or enhancements in work or everyday life can be formulated (concept design), resulting in a final product, application or service design. In ‘understanding the user’, insight in different variables plays an important role (e.g. user characteristics, user behavior, contextual variables and means used. For an overview see [5,6]). Good product solutions are characterized by the right identification of and insight in these variables. Once insight in relevant variables is obtained, directions in which one can find solutions are further explored, by developing personas and scenarios [5]. Through iteration personas and scenarios become more refined and one gradually enters the design phase.

Methods often applied to explore and identify the relevant variables are field studies (e.g. ethnographic studies, contextual inquiry, probe studies, diary studies, observations and literature studies); surveys (questionnaires, interviews, workshops and focus groups) and in later stages modeling techniques like task analysis/modeling and mental model assessment. Methods that are often applied in the concept design phase are participative design sessions (e.g. workshop, focus groups). With the help of design patterns, user interface guidelines, checklists and interaction design principles the functionality, as given in scenarios, is translated to an interaction design and finally a product [5]. By iterations through rapid prototyping, mock-up and demonstrator testing a final design for implementation is created. Most of these methods have their own specific challenges.

Challenges in field studies

Insight gained with field studies is incredibly helpful in understanding the user and in formulating possible design solutions. However, they do have some disadvantages. Field studies tend to be very resource intensive. Results are often unstructured and subjective in nature and take time to analyze. Relevant variables are often sparse and sometimes difficult to capture [3]. Observations in context cost a lot of time and therefore typically use small sample sizes which gives no or very limited insight in the quantitative nature of variables. A limited number of observations over a small sample furthermore lead to biased insight (not mirroring reality). Last but not least, most field study methods are quite obtrusive (e.g. observation) which can lead to changed user behavior. Some field studies, like probe and diary studies don't have all the problems just mentioned [3]. They tend to be less resource intensive because participants capture relevant variables themselves. These types of studies can therefore be used on larger sample sizes and are less obtrusive. Still, other problems remain (e.g. resource intensive analysis, no quantification, partial or biased insight, etc) and some new are introduced. Part of the control over the study might be lost since participants do the capturing [3]. Furthermore one is asking participants to make an effort to capture relevant variables [3].

Challenges in surveys, participate & evaluative session

Surveys, participative and evaluative sessions provide the means to obtain detailed insight in relevant variable identified in field studies and to elaborate on functional and interaction details of different scenarios and designs. Their main disadvantage is that they often can not be used in situ. Some detailed research interests can not be or are very difficult to answer out of context (e.g. heuristic or implicit knowledge, feelings like trust and needs like privacy, frequency of usage, time assessments, and other variables or research interests in which context factors play an important role).

Coping with these challenges

Some of the challenges just addressed can be solved by performing triangular studies (using data from multiple resources to establish similar patterns) [4]. Triangular studies however can become resource intensive when combining different methods mentioned in the last paragraphs and some problems in gathering data will still remain. Different researcher have coped with these problems by developing and using less resource intensive methods for data gathering, sometimes combining these with (parts of) data gathering methods commonly used in UCD before the implementation. Some examples of these methods are the Experience Sampling Method (ESM) [7,8]. Letting participants capture variables (by means of photographs and audio recordings) which provide a context reference for follow up interviews [3]. Capture usage variables (web page visits, search tasks, browser functionality use through client logging) and ask users to classify these variables [9, 10] or use follow up interviews to add subjective information and user interests to usage patterns [11]. Some of these studies can easily be performed over a longer period of time, in situ, with larger sample sizes, are less obtrusive, tend to be less biased and provide qualitative as well as quantitative information.

2.2 Implementation

Close cooperation between researchers, designers and software engineers results in better products and services due to common knowledge and shared understanding of the user and technology [5]. Currently, engineers are seldom involved in other UCD phases than the implementation. On the other hand researchers and designers are seldom involved in the implementation. Often, the result of this separation of involvement is misunderstanding each others objectives and an inability to help each other with design or implementation problems in a constructive way. This holds true for the design and implementation of product and services as well as for the implementation of evaluation methods [12].

Creating shared understanding

To support shared understanding researcher, designers and software engineers need to share insights on a regular basis and speak a common language. Different efforts have been made in this area [5]. Besides 'getting involved' in each others fields of work during different phases of UCD and software engineering process, design patterns can help in the transition from design to implementation. In software engineering design patterns are common practice. In UCD design patterns have started to emerge. Design pattern can enhance shared understanding about product functionality, interaction design and implementation architecture.

2.3 Evaluation & pilot

In the evaluation & pilot phase a product or service is tested with users. Interests during evaluation are the usability of a product or service (formative as well as summative), the user experience and the usage. Recent research in these three interests now also includes the influence of contextual variables and context awareness, adaptability of applications and personalization [13,14]. Usability assessments are mostly performed in usability lab studies, model based or heuristic evaluations (expert evaluations, user interface guidelines and checklists, cognitive walkthrough, task analysis, etc.) and questionnaires (e.g. SUMI). Some of the methods or parts of them have been automated (for an overview of automated usability testing see [15]). Assessing the user experience and usage of products and services is commonly done during or after in situ pilots (ranging from a few weeks to months). User experience is still a vaguely defined concept [13], though some means to assess the user experience are: PrEmo [16] with which participants select 'puppets that enact an emotion', different physiological measurements [17], Jordan's questionnaire for qualification of product pleasurability [18] and the ESM [7,8]. Usage is often studied with methods mentioned in section 2.1 (field studies and surveys). Log files (server, client or service logs e.g. call detail records from a telecom provider or web server, IM and error logs, etc) are sometimes used to provide more detailed quantitative usage data (e.g. feature usage, frequencies of use, action sequences, etc).

Results of the evaluation & pilot phase typically consist of usability problems, realization of usability, user experience and usage performance goals and redesign proposals [4].

Challenges in usability assessments

Usability studies in labs can provide valuable insight in different aspect of the usability and issues or parts of the product or services for which redesign should be considered. However, in testing complex services some problems do arise. Lab test are performed with a limited number of participants and scope (due to needed resources), are often task based and not performed in situ, leading to a sometimes unrealistic estimation of certain usability aspects of a product or service [17]. Furthermore, formative evaluations tend to be sensitive to interpretation bias [15]. Summative evaluations are valuable for gathering quantitative information to prioritize problems and compare design to re-design. However, since sample sizes are limited summative results can be questionable [15].

Challenges in model based & heuristic evaluations

With model based and heuristic evaluations insight is obtained in design flows, design obstacles, and inconsistencies [15]. They provide relatively quick means for design evaluation with fewer resources than needed for usability studies. Disadvantages are that heuristic and expert evaluation can be biased (consistency over evaluators is sometimes low), though they often do cover the most important design flaws [15]. Furthermore, in situ influences on the usability are not accounted for.

Challenges in usage studies

Field studies and surveys are commonly used to study usage. For an overview of challenges and coping strategies of these types of studies see section 2.1). Measurements in usage assessment with e.g. log files have great value in providing quantitative information about usage in situ, with large sample sizes over a longer period of time. However, most commonly used log files are system or service logs, intended to support software engineers and technicians to trace problems or provide service owners with e.g. billing or management information. Though these log files have value for usage research, specific research interests concerning usage sometimes remain unanswered. Logging methods specifically developed for usage studies have been implemented [4,9,10,11,12,19] though these methods are often situation or application specific (no standards) and problems during the implementation of these methods as mentioned in 2.2 are common.

Challenges in user experience

To utilize standard user experience measurements, we first need a commonly accepted definition. Since user experience is still a vaguely defined concept, there are no standardized methods to measure it. User experience is within FRUX defined as the 'total experience of a user before during and after product usage'. Measurements for these three stages of experience need further development and standardization, see also [13].

Coping with these challenges

Automated usability measurements and heuristic evaluation can realize in situ measurements and partially solve problems like needed resources, sample sizes and study scope [1]. However, the problems concerning standardization of methods and their model- or task based nature still persist. Developments in predictive models, pattern recognition algorithms, and learning algorithms however might provide good solutions [4]. In usage studies implementations of logging methods combined with (automated) qualitative data gathering techniques (e.g. ESM, interviews, triggered video recordings) have been used [7,8,11,12] and provide good solutions to study specific aspects of usage a researcher is interested in.

3 FRUX SocioXensor

FRUX SocioXensor has the long-term goal to better support UCD research and evaluation throughout the design cycle, by solving current problems experienced in UCD. This year FRUX is aiming to create the data gathering infrastructure and the design patterns to support usage of this infrastructure, as a first step towards the long-term goal. Key concepts are automated data gathering methods for qualitative and quantitative measurements. Special attention will be paid to resources needed to perform studies, the scope of sample sizes and studies, obtrusiveness of measurements and support for in situ data gathering. Though FRUX is mainly striving to support data gathering concerning its own UCD research and evaluation interest, methods currently developed will apply in research of other interests as well.

3.1 Data gathering support

To support qualitative and quantitative data gathering SocioXensor provides two different types of measurements: sampling and logging.

Sampling

Sampling provides the possibility to capture subjective information from participants in situ. During sampling a participant is either given an assignment beforehand to capture specific data, or is confronted with a question or assignment during the sampling study (randomly, at specified time intervals or based on the occurrence of a specific event). A sampling tool that is currently being developed for SocioXensor is the Experience Sampler (see for more details on this tool [8]).

Logging

Logging is a method that provides (in situ) data gathering without the researcher or the participant being actively involved in the data capturing process itself.

SocioXensor will provide design patterns and standardized logging tools to support quantitative data gathering for usability, usage and user experience measurements by supporting gathering data via:

- Devices (e.g. smartphones, PDA's, PC's [2])
- Network (e.g. node or cell a client is connected to)
- Servers (e.g. dynamic information send to or received from clients, active clients, communication)
- Applications running on devices, networks and/or servers (e.g., active applications, information presented, actions/events occurring within applications)

By combining one logging method with another logging or sampling method a researcher can perform powerful quantitative analyses supported by qualitative insights, in situ. SocioXensor will support the integration of these combinations of methods as well.

4 Future research

Triangulation and combining partially automated qualitative and quantitative measurements are very powerful methods for UCD. Data gathering with SocioXensor is the first step towards realizing such studies, but there are many more challenges to cope with before a perfect solution will be available. Some examples of these challenges are coping with data transformation (filtering and combining data from different sources), data analyses (pattern detection, data interpretation) and data visualization (making complex data structures insightful). In the coming years different stages following the data gathering will be important topics to study within FRUX.

5 Acknowlegdements

This paper has been written in the project "Freeband User Experience" (FRUX), which is part of the research programme Freeband Communication (www.freeband.nl), which aims at the generation of public knowledge in advanced telecommunication technology and applications in the Netherlands. FRUX is a joint effort of Ericsson, ISC (Dutch Police ICT service organization), Telematica Instituut, TNO ICT, Delft University of Technology, VU University medical center, Waag Society, and Web Integration..

References:

1. Kort, J, Poot, H. de (2005). Usage analysis: Combining logging and qualitative methods. CHI '05 extended abstracts on Human factors in computing systems, pp. 2121-2122.
2. Poot, H. de, Kort, J. and Langley, D. J. (2004). Enhancing Presence and Context awareness in Collaborative Settings. *P. & M. Cunningham (eds.) eAdoption and the Knowledge Economy, Issues, Applications, Case Studies. IOS Press. 2004, pp. 1325-1332*.
3. Mankoff, J., Carter, S. (2005). Crossing Qualitative and Quantitative Evaluation in the domain of Ubiquitous Computing. *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods"*, Portland, OR, USA. April 2, 2005.
4. Edmonds, K. A. (2005). Every Move You Make: Re-engineering UCD. *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods"*, Portland, OR, USA. April 2, 2005.
5. Cooper, A., Reimann, R. (2003), About Face 2.0, The Essentials of Interaction Design. Wiley Publishing, Inc. Indianapolis, USA.
6. Holtzblatt, K. (2001). Contextual Design: Experience in Real Life. Mensch & Computer 2001. March 8, 2001. Bad Honnef, Germany.
7. Hofte, G.H. ter, Mulder, I. (2003). Dynamic Personal Social Networks: a new perspective for CSCW research and design. SIGGROUP Bulletin, 24(3), December 2003, pp. 139-142.
8. Mulder, I., Hofte, G. H. ter, Kort, J. (2005). SocioXensor: Measuring user behavior and user eXperience in conteXt with mobile devices. *Measuring Bahvior 2005*. Wageningen, the Netherlands. August 30, 2005.
9. Hawkey, K. (2005). Privacy Management of Incidental Information During Collaboration: Data analysis and Evaluation Challenges. *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods"*, Portland, OR, USA. April 2, 2005.
10. Kellar, M., Watters, C. (2005). Studying User Behavior on the Web: Methods and Challenges. *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods"*, Portland, OR, USA. April 2, 2005.
11. Rodden, K., Blackwell, A., Jones, R., Milic-Frayling, N. (2005). Combining Logging with Interviews to investigate Web Browser Usage in the Workplace. *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods"*, Portland, OR, USA. April 2, 2005.
12. Hartman, G., S., Bass, L. (2005). Interpreting Audio/Video Recordings and Logs. *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods"*, Portland, OR, USA. April 2, 2005.
13. Mulder, I., Steen, M., Hofte, H. ter, Kort, J. (2004), Mixed emotions, mixed methods. *Freeband User Experience*. Available online: https://doc.freeband.nl/dscgi/ds.py/Get/File-48176/
14. Eijk, R. van, Mulder, I., Hofte, H. ter, Steen, M. (2004). We-centric, context-aware, adaptive mobile service bundles. *Freeband User Experience*. Available online: https://doc.freeband.nl/dscgi/ds.py/Get/File-48178/
15. Yvory, M. Y., Hearst, M. A. (2001), The State of the Art in Automating Usability Evaluation of User Interfaces. *ACM Computing Surveys*. Vol. 33, No. 4, pages 470 – 516. December 2001.
16. Desmet, P. (2002). Designing Emotions. *Ph.D. Thesis*. Delft University of Technology. Delft, the Netherlands.
17. Esch – Bussemakers, M. P. van, Neerincx, M. A., Lindenberg, J., Streefkerk, J. W. (2005). User Experience Testing in Context: Attentive Services for the Police. *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods"*, Portland, OR, USA. April 2, 2005.
18. Jordan, P. (2000). Designing Pleasurable Products: An Introduction to the New Human Factors. *Taylor & Francis*, London.
19. Campbell, J. D. (2005). Log and Qualitative Analysis of Collaborative Systems. *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods'*, Portland, USA. April 2, 2005.

Mobile measurements of mobile users

V Lantz[1], J Marila[1], T Nyyssönen[2], H Summala[3]
[1] *Nokia Research Center, Helsinki, Finland*
[2] *Nokia Technology Platforms, Helsinki, Finland*
[3] *Department of Psychology, University of Helsinki, Helsinki, Finland*

Abstract
This paper describes development of a portable measurement setup and experimental work conducted in co-operation between Nokia and University of Helsinki. The focus of the project was on mobile Human-Computer Interaction research.

The main motivation for the project was to gain a deeper understanding of everyday mobile device use in context of walking in a city environment. The ultimate goal is to better be able to design devices that enable fluent, convenient and truly mobile user-device interaction, and uninterrupted and safe operation in various situations and environments.

Keywords
Mobile usability research, gait analysis, attention, miniature cameras, movement sensors

1 Introduction
To some extent, the usability of mobile devices can be evaluated in laboratory conditions, e.g. with heuristic walk-through techniques. That way, we can identify a significant part of the design problems. But, with an overly simplistic test condition, we will get overly optimistic results. The usage situations where the user needs to divide her attention between the environment and the device has to date been mostly neglected, though such probably occur to every user every day. To be able to carry out accurate usability testing in mobile environment, we need equipment that provides accurate and realistic information on the device usage and human behavior in common mobile situations.

1.1 Walking as an object of study
Walking is the common form of human movement. For instance, in a study on physical activity patterns on African American men and women (age 30 to 60, medium income) 60% informed that they spent at least half of their workday walking 1. That is being mobile for at least 4 hrs on 5 days a week. Mobile phones are a class of devices that are commonly used in this environment. The design of mobile phone user interfaces should take into account this demanding usage context.

Walking includes several sub-tasks: obstacle avoidance, path selection, and choice of speed, limb movements and balance and posture control. Performing secondary tasks that are not related to walking, like using a mobile phone, can affect the subtasks of walking in different degrees. Posture control is mildly affected by visual and verbal secondary tasks [2, 3]. Verbal memory load affects walking speed selection[4] and (mathematical) problem solving affects the transition from walking to running [5]. Decision-making, in form of sorting coins, has been observed to slow down pedestrians' walking speed [6]. Text reading speed on a mobile phone display has been found to decrease when the subjects are walking, as compared to sitting or standing still [7].

1.2 Mobile usability
When testing user interface efficiency and usability, the testing is usually carried out in closed, controlled settings. This way the effect of external factors to the test results can be minimized. But consumer electronics is not used just in closed, controlled environments. For example, mobile phones are used in more or less any environment that the user happens to be. Being in locomotion just adds to the difference between the controlled, immobile test settings and the actual usage situations.

In two tests comparing controlled laboratory settings with different walking conditions, in and outdoors, Kjeldskov & Stage found that more qualitatively classified usability problems were discovered in controlled, immobile indoor settings than in any other, but the surplus problems were less crucial, cosmetic, to usage. Authors conclude that the difference is due to the subjects' decrease ability to notify the test facilitators about the usability problem they come across [8]. Device usability thus seems to have some relationship with the users' movement conditions.

If mobile phone usage taxes user's cognition, it should reflect itself in the walking characteristics – supposedly as a decrease in walking speed, and possibly walking rhythm. We set out to study this possible effect, and also whether it could be used as a usability metric: walking speed decrease, or increase of variation walking speed, could reflect the amount of load that device user interface puts on the cognitive skills such as decision-making, memory retrieval and visual search.

1.3 Requirements for measurement setup
In order for us to study the behavior of our test users while they walk in mobile use contexts, we needed to be able to track down their individual steps and velocity, location and device usage. *Stride* length (same-foot ground contact interval) would be a good control for *cadence* (steps per minute) since the *walk ratio* (stride length divided by cadence) is good rough measure over time on the pedestrian speed choice: when walking is slowed down, cadence drops but stride length remains roughly constant.

Required from our measurement setup was:

- Video footage of user's surroundings, her actions on the usability-tested device, her gaze direction
- User step frequency, absolute or relative position and orientation
- Recording of all sounds occurring near to the user, coming outside of the device and from the user
- Reliable, fast and spacious storage for the video and walking data

- Compact pack where the user can carry all the measurement equipment and the required power sources

2 Measurement setup for Mobile HCI reaseach

The measurement equipment consists of devices that are commonly available in consumer markets. They are slightly modified if required, attached to a plywood rack with Velcro, connected with custom or standard cables, and put together to fit into the backpack. The plywood rack is illustrated in Figure 1 .
The mobile measuring backpack combines video signals from a set of four miniature cameras, records them to a notebook laptop, and onto external memory card as a backup. Data from an advanced movement and orientation measurement system are stored on the same PC. In addition, various physiological or environmental measurements can be stored via data channels. Movement sensor system connects to PC via Bluetooth or serial port. There are many simultaneous measurement and recording processes running on the PC. They are asynchronous and controlled with different software tools. Synchronization of various signals is based on time stamps in every input data line, and synchronization pulses and clapper signals.

The measurement system contains two independently operating and recording peripheral sensors. One is an independent pedometer AMP311 by Dynastream Innovations Inc. – used as reference – and another one is a versatile GPS-sports logger by FRWD Technologies Ltd. with sensors measuring heart beat rate, geographical position, temperature and air pressure. These devices record on their internal memories.

Peripheral equipments have one sample per second time resolution. Other sensors have higher sampling rates and these measurements can be synchronized and resampled to match the video frame rate (1 frames per 25 seconds).

2.1 Batteries

A big challenge is to provide all the different voltages the components of the system require – and make it with a single central battery that is not too big or heavy, and which is easy to change and easy to charge. As the lead acid batteries and chargers are relatively cheap, we decided not to so much optimize the weight or capacity, but make the system work with batteries that can be easily changed between test sessions. Our backpack is powered with a traditional miniature 12 V lead acid battery pack. There are two or three voltage regulators in the backpack to provide different voltages: 6V, 7.5V, and 5V.

2.2 Cameras

Perhaps biggest single effort in building the measurement system was the design the camera system.

There is one camera for monitoring what is going on with the test user's phone UI. This camera is attached to the phone with a special grip and rod system. The angle and distance of the camera can be adjusted so that the camera captures the display, and possibly also the keys of the phone.

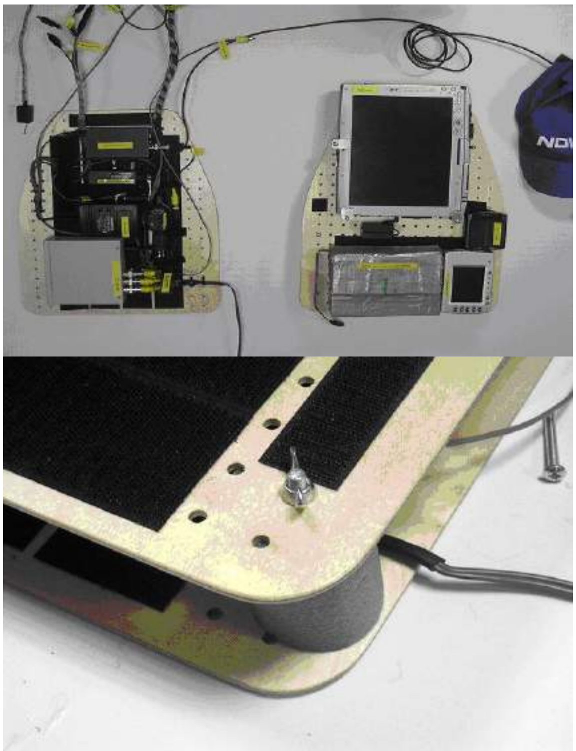


Figure 1. *The plywood rack for organizing the equipment inside the backpack. Above, opened; below, closing mechanism is shown.*

Figure 2. *The device UI / user gaze camera combination attached on the device.*

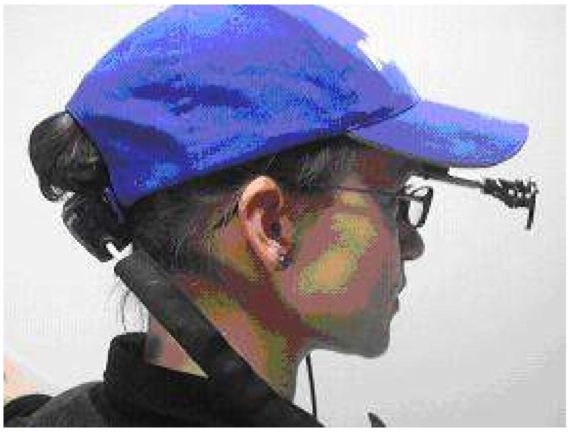


Figure 3. *Alternative gaze camera, attached to the peak of the cap movement sensor is attached to the adjustable strap in the back of the cap.*

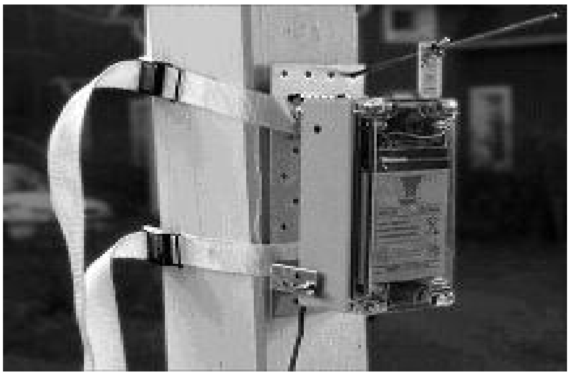


Figure 4. The wireless satellite camera attached on a pole.

In addition, the camera can be positioned so that it is directed to the user face when the phone held in a typical use posture. With such a setup is easy to detect when the user is glancing at the phone. Naturally, two cameras can be attached to the same rod, upper camera facing the user and lower camera monitoring the UI. Such a setup is shown in figure 2.

.

Second camera is attached to a peak of baseball cap. The camera is attached to and aligned with a telescope stalk having an adjustable mirror in the end. This camera provides a view of the user's eye and helps to keep track of the directions where the user is looking at. The baseball cap limits the user's angle of view to upward directions, but at the same time, it makes the protruding camera less obvious for outsiders. The head-mounted camera is shown in figure 3. Third camera is attached to the shoulder strap of the backpack with Velcro and it shows what the test user is seeing in her front view.

Fourth camera channel was reserved for wirelessly connected cameras. A wireless camera can be either carried and operated by the test moderator or positioned like a surveillance camera to some strategic places along the test route, as in figure 4. The wireless cameras are connected to the rest of the measurement system with a radio link.

If several wireless cameras are installed along the test route and they all use the same radio frequency – and have some distance between each other – the strongest camera signal from the camera closest to the test user is selected automatically and recorded via the wireless camera channel. With such a setup, the relevant video sequence showing test user's behavior in mobile context is recoded directly into backpack, in synchronization with other video streams and sensor data. This simplifies the video analysis process significantly as there is no need for going through several separate video recordings and manually selecting the parts where the user appears. In theory, the maximal distance between the user, or radio receiver, and the wireless camera is 300 meters.

The ambulatory miniature cameras carried by the user and the wireless cameras are manufactured by Watec. The cameras get 6 VDC power from the backpack battery (12 VDC), through a DC-DC regulator or they have their own integrated 6 VDC batteries.

The four camera channels are combined into a single video stream with a quad, see Figure 5. The combined video data is directed thru DV converter and recorded with a laptop. A DVCR is used as a backup method.

Figure 5. The stored quad-camera video (satellite camera corner being blank at the moment).

Recording to the laptop is made with a compact USB 2.0 video mpeg converter Video Grabbee X. Video signal is imported to the device from the quad, in composite PAL video format. DVCR uses Mpeg4 format. The miniature cameras and lenses used in our measurement system are manufactured by Watec.

2.3 Sensors

The core of sensor setup was a motion sensor system that combines 3D acceleration, 3D compass and 3D gyro data from multiple sensors. This data can be used to measure human moving and provides versatile information about gait parameters and head movements, for example. In our measurement system there are three movements sensors attached to the user's foot, lower back and back of the head. The movement sensors are connected with wires to a hub unit that coordinated the measurements and sends the data to PC using Bluetooth or serial port connection. The sensors (MT9-B) and the hub unit (Xbus Master) are manufactures by Xsens Technologies. Movement data can be sampled with frequency up to 400Hz.

A movement sensor is attached to the foot with a Velcro strap and anti-skid shoe strap, with the ice-pins removed. The head sensor is attached to the back strap of baseball cap (see Figure 3) and the back sensor to an adjustable, wide elastic band (originally designed for carrying Dexal sport drinks and energy gels) with duct tape.

Unfortunately our movement sensor setup did not provide the accurate measurement of step length. It is very difficult to accurately extract the gravity component of the acceleration measurements and the double integration of acceleration into position will cumulate all the errors. Just one stride, usually taking less than one second, is a long enough time interval to make integration-based position estimates uselessly inaccurate. On the other hand, step detection, and thus cadence estimation, can be performed very accurately and robustly using the movement sensor data and simple algorithm looking for acceleration peaks, points of low variation in acceleration, and zero-crossing of the Euler angles of the sensor orientation.

Other sensors can be connected to PC via a portable 8-channel USB data logger or an integrated data logger with parallel port connectivity. Some potential parameters to measure are background noise level (dB-level), lighting conditions (lux-level), temperature, for instance. Large variety of physiological or other environmental sensors can be connected as peripherals to data logger. USB logger sample accuracy is 12 bits.

3 Annotation and analysis software

For viewing and visually inspecting the video and collected data, a piece of Java software was created in the University of Helsinki Traffic Research Unit.

The software allows importing all the data and video material into one view and to view them synchronously. Time-based textual annotations can be inserted and the route points and other interesting locations can be marked on the video screen co-ordinates by a mouse click. The annotation software is illustrated in Figure 6.

For purposes of analysis and data preprocessing, e.g. step detection and cadence estimation, we use Matlab, SPSS and Excel software for Windows PC.

4 Experimental work

So far, we have used the equipment in three separate testing sessions. We have evaluated the effect of a memory-loading phone conversation to the walking performance. Sensor data about cadence and head turning as well as video on gaze direction and route progress were used to assess the subject's walk while audio recording was used to assess the verbal memorizing and retrieval efficiency.

Two new user interface solutions were tried on test users who performed tasks on them while walking on the city streets and riding on subway train.

At the University of Helsinki Traffic Research Unit, experiments have been carried out on turn anticipation. The first test considered anticipatory head movements when turning 90 deg around a corner. People typically look where they go, and it is also critical to pedestrians to prepare for potential collisions with other pedestrians, bicyclists, and even with cars in certain environments. In these tests, carried out in a corridor, we manipulated task load and walking speed. The task load conditions included a visuomotor task (continuous writing of a text message) and simple auditive arithmetic task. The participants had to produce three different walking speeds: slow, normal, and fast. Turns were made at both ends of the corridor, one with plenty of lateral space and another with a narrower opening that forced participants to a more uniform path. Both time and space scaled analyses of data were performed, as based on remote photocell information. The data showed both task load and speed effects on anticipatory head movements, however dependent on the lateral space in front of the corner. In a wide corridor the anticipation was more marked, also reflecting the path selection. The results also indicated substantial individual variation in anticipatory reactions among participants.

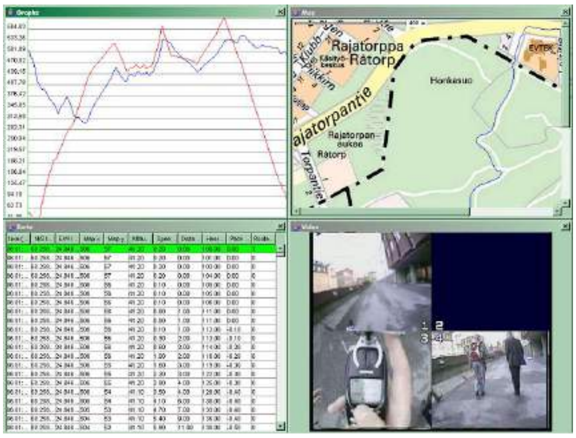


Figure 6. *The video/data annotation and inspection software, main view.*

These results emphasize the need of exact measurement of pedestrians' movements and steps in space, to be able to consider higher level navigation and path selection, lower level path control, and even managing with uneven ground such as curbs and holes in pavement.

References:

1. Young D.R., Miller K. W., Wilder L. B., Yanek L.R., Becker D.M. (1998). Physical Activity Patterns Of Urban African Americans. *Journal of Community Health,* **23**, 99-111.
2. Dault M. C., Frank J. S., and Allard F. (2001). Influence of visuo-spatial, verbal and central executive working memory task on postural control. *Gait and Posture*, **14**, 110
3. Shumway-Cook A., Woollacott M., Kerns K. A. and. Baldwin M. (1997). The effects of two types of cognitive tasks on postural stability in older adults with and without a history of falls. *Journal of Gerontology*, **52A**, 232.
4. Li K. Z. H., Lindenberger U., Freund A. M. and Baltes P. B. (2001). Walking while memorizing: Age-related differences in compensatory behavior. *Psychological Science*, **12**, 230.
5. Daniels G. L. and Newell K. M. (2003). Attentional focus influences the walk-run transition in human locomotion. *Biological Psychology*, **63**, 163.
6. Shkuratova N., Morris M. E. and Huxham F. (2004). Effects of Age on Balance Control During Walking. *Archives of physical medicine and rehabilitation*, **85**, 582.
7. Mustonen T., Olkkonen M. and Häkkinen J. (2004). Examining Mobile Phone Text Legibility while Walking. *CHI2004*, Vienna, Austria, April 24-29.
8. Kjeldskov J., Stage J. (2004). New techniques for usability evaluation of mobile systems. *Int. J. Human-Computer Studies*, **60**, 599–620

Measuring Affect with Plutchik's Model in Human Computer Interaction and Computer Mediated Communication

Laszlo Laufer[1], Gábor Tatai[2], Bottyan Nemeth[3], Janos Loke[4]
[1] *Department of Psychology, Budapest University of Technology and Economics, Budapest, Hungary*
[2, 3] *AITIA Inc., Budapest, Hungary*
[4] *Institute of Behavioral Sciences, Semmelweis University of Medicine*, Budapest, Hungary

Abstract
We adapted Robert Plutchik's model of emotions to Internet chat, and chatter-robot systems. First we introduce an application which helps to label phrases and expressions in continuous texts. We use this application to build our emotion-text database which is one of the foundations of the emotion module of our chatter robot system. We also created an emotional chat application in witch the two conversational partners had to label their sentences with basic emotions. In the paper we try to approach empathy as altering one's emotion in response to the partner's change of her label. We also developed a visualization tool based on Plutchik's circumplex model, which displays the two conversational partners' emotions on a series of circles placed one above the other as they followed each other in the discourse. With the help of these tools we can visualize the changes of emotional load in a discussion, and calculate several statistical parameters which can describe empathy in the discussion.

Keywords
Affective Computing, HCI, CMC, Empathy

1 Introduction

Affective Computing is a flourishing research area at the meeting point of artificial intelligence and human computer interaction (HCI) research. One of the most prominent advocates of the use of affective modalities in computer systems is Rosalind Picard from MIT. According to her definition this field is dealing with problems where "computing is related to, arises from or deliberately influences emotions" [2]. It is a significant challenge to model emotions and effectively use them in computer systems.
According to recent results in the use of emotions HCI we adopted a "basic emotion" type of model by Robert Plutchik. Models of this kind are preferred in HCI, because on one hand they give a well defined set of emotions, which the system can be prepared to react on consistently and it is easy for both the system and the user to discriminate between them, on the other hand basic emotions have the advantage of being similar across cultures, thus designs that attempt to act upon or manipulate these dimensions may be the simplest to implement [1].

The emotional model defines what to detect from the users' input, and what type of emotions can be incorporated in the system's response, but it does not give an answer to the "how"-s:

- How to detect these emotions?
- After detecting an emotion how to react on it?

The answers to these questions are very much dependent on the modalities of the interaction between the computer system and the user.
Our major goal was to develop a chatter robot system, an animated character carrying out conversations with the users, and able to detect and to react on the emotional load of the users utterance adequately. In the case of a chatter robot system, during the interaction the user only utilizes written verbal modality: types in text messages. The output modality is also mainly written verbal: the chatter robot replies by displaying a message. Sometimes an animation is accompanying these messages expressing an emotional state, so this other modality can be used as well.

In this paper we are introducing two applications, giving an answer to each of the above mentioned "how" questions.
The first tool is used to label text segments with basic emotions. This tool is used to establish an emotion text database, which can be used by the chatterbot to extract the emotional load of the user messages.
The second tool, which we are going to emphasize more in this paper, is emotional chat interface, which makes it possible for two (human) chatting partners to label their text messages with basic emotions during their conversation. As a result of this second application we get basic emotion label sequences, which we can analyze afterwards. This analysis can lay the foundation of how to plan the emotional aspect of the interaction, and thus giving an answer to the second "how" question.

1.1 The theoretical model: the Plutchik cone

In Plutchik's theoretical model of emotions there are eight basic emotions. In addition he suggested another dimension, intensity with 3 levels, thus altogether 24 emotional states. A novelty of his proposal is that the emotional space is mapped to an upturned 3D cone, where the positioning of a particular emotion reflects psychological distances and intensity differences between states [3, 4].The origo is in the apex of the cone expressing neutrality ("no emotion"). Subsequently each of the 24 emotional states can be identified to a vector starting from the apex and pointing to the surface segment assigned to the emotion consequently.

1.2 Measuring change in emotions

The distance between two basic emotions out of the set of 25 (24 + no emotion) can be calculated as the distance between the two center points of the surface segments assigned to these emotions. In other words the subtraction of the two emotion vectors starting from the apex.

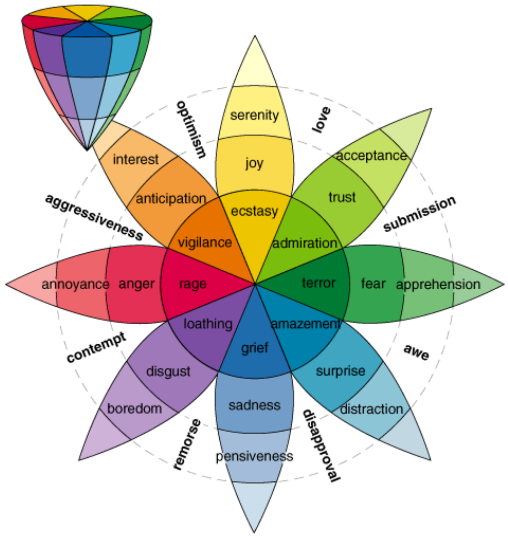


Figure 1. *The 3D cone representation of emotions.*

We can also calculate the intensity difference between the two emotions, if we decompose Plutchik's 24 basic emotions as a "primary" basic emotion from a set of 8 basic emotions found in the middle stripe of the cone (sadness, surprise, fear, etc.) and a number between 0 and 3, showing the level of intensity. Only zero emotion has 0 intensity. This second way of calculating the distance between two emotions might be a bit more complicated, but it corresponds more to other widely accepted theories of emotion. So in this second way of calculation the distance between two emotions can be min.0, max 3 far away from each other in their intensity, and min 0, max 4 away from each other regarding their "emotional" distance. E.g. the distance between annoyance and surprise is 1 regarding their intensity and 3 regarding their primary basic emotion. The detailed emotional model has been published and evaluated in our earlier publications [5, 6].

2 Emotion labeling and visualization in texts

As we mentioned in the introduction in order to measure the affective content of a text message, we have to establish an emotion text database. For this purpose we created an application with graphical interface. The interface can load continuous texts, and visualize the emotional labels that are already stored in the emotional expression store assigned to the expressions in the content and can also be used to assign emotion labels to segments of the text. This way through the emotion editor we can read out the text, and follow the changes of emotional states in time.

This interface allows the carrying-out of reproducible experiments on the emotional model, while the various visualization tools allow us to trace the process of emotion generation. Due to the lack of space we can not go into details how our emotion modeler works, so in the following passages we are just we just briefly describe its components.

The three rotateable and zoomable camera views (see Figure 2, panels 1.1, 1.2, 1.3), displaying the cone with the emotion vectors. These vectors are representing the different emotional influences caused by the previous passages, they are constantly changing, and fading away (fading is caused by a separately adjustable function, which manipulates the length of the vector in time). The resulting emotional vector (sum of the emotion vector) is also being displayed on all three panels.

The basic emotion monitor shows the intensities of each basic emotion at a given moment (see Figure 2, panel 2). It provides a great help to refine the basic emotions assigned to each phrase in the text.
The cursor on the Text panel (see Figure 2, panel 6) shows us which part of the text is being currently read. On the Basic emotion monitor it is possible to readjust the intensity of the 24 basic emotions participating in the composition of the complex emotional load of the current phrase. At the button part of the Text panel (see Figure 2, panel 6) the user can start, stop and choose a position in the reading or adjust the speed for better following the changes in the resulting emotion vector. The dominant emotion (or active emotion) is calculated from the projection of the resulting emotion vector to the surface segments of the cone and being displayed at the button left corner of the panel. This emotional state is the output of the system; this emotion is incorporated in the response of

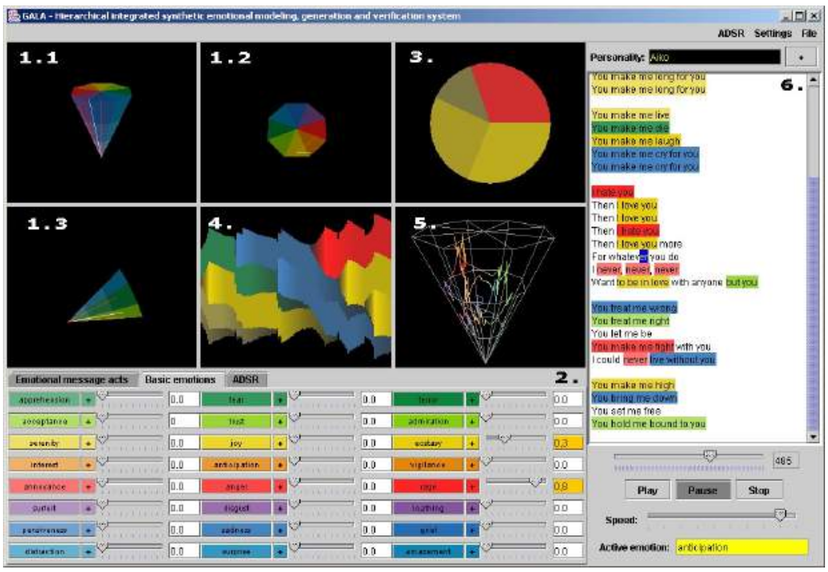


Figure 2. *The interface of the emotion editor*

the chatterbot.

Besides the camera views, there is another possible way to track and display active emotions (see Figure 2, panel 3) The sectors of the circle symbolize the active emotion components at any given time and their colors correspond to that of the appropriate emotion. The areas of the sectors are directly proportional to the percentage shares of the basic emotions from the current resulting emotion vector, the present composite-emotional state. The length of this vector, the intensity of the composite emotion is represented by the radius of the circle.

Panel 4 has an important role in displaying the emotions synthesized throughout the whole conversation. When "reading" the text, in this case a poem, the system generates a colorful pattern that gives us a picture of the poem and enables us to get an overview of the emotional characteristics of the analyzed poem (see Figure 2, panel 4).

The X axis of the pattern stands for the time dimension: a vertical slice shows the active emotions (represented by the appropriate colors) at a given time, proportionate to their actual intensity. The height of a vertical slice represents the length of the resulting emotion vector. As time goes by, new slices ("emotion ensembles") are added to the picture from right, resulting in the continuous condensation of the existing piece.

Panel 5 allows us to follow and depict the route of the resulting emotion (the sequence of emotional change) of the conversation, in the function of time. For better visualization, we used only the wire-frame view of the Plutchik cone (see Figure 2, panel 5).

3 Measuring emotional change in chat

We developed a user interface which enables the chatting person to send emotional states assigned to his/her messages in the same way as our chatterbot is revealing her emotional states while being engaged in a conversation with the user.

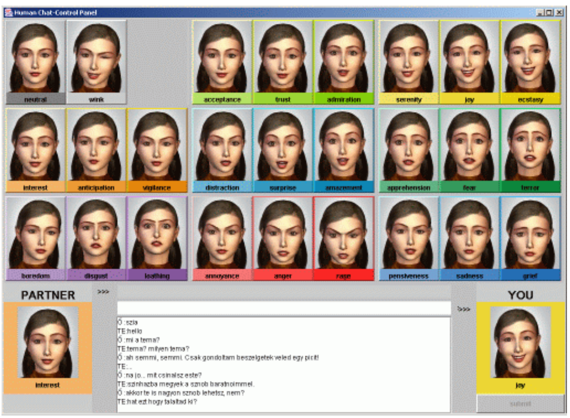


Figure 3*. The emotion transmission interface with the virtual 3D chatter robot face.*

Pictures on interfaces represent the most expressive phase of the emotion animation. At the start of the conversation the animation default is the neutral face. During the dialogue, users send various emotions to each other so then a short animation sequence is played to express the selected emotion. Faces are embedded in a frame whose color corresponds to the color of the appropriate segment of the Plutchik cone that stands for the depicted emotion. The face of Aiko, our virtual agent, was visualized by pre-rendered 3D bitmap animations for all 24 basic emotions.

The chat panel and history is located in the middle at the bottom of the screen. The user enters the message, then selects an appropriate emotion or the neutral face then sent the message. Both users could see their own virtual face (see face in the button right corner) and the virtual face of the partner (see face in the button left corner) simultaneously.

3.1 Analyses and visualization of the results

After the chatting partners labeled their sentences with the 25 (24+1) basic emotions, we receive a sequence of emotions, which we can visualize similarly to the emotional content of the continuous text.

On Figure 4, panel 1 we can follow the dialogue. The colors of the sentences are corresponding to the colors of their emotional load. On Figure 4, panels 2 and 3 we can trace the conversational partners' expressed emotions (chatter A on the top, chatter B at the bottom). Similarly to Figure 2, panel 5 in the wire-frame view we can follow the way how the communicational partner's emotion changed during the course of discussion.

On Figure 4, panel 2 we are able to follow which emotions were used, indicated by a dot in the corresponding surface segment. The size of the dot shows us the frequency of use by the chatter.

The most interesting part of the interface is Figure 4, panel 4. As we described in chapter 1.2 there is another way of calculating the distance between two emotions (out of the set of 25): calculating the difference between their intensity and the difference between their primary basic emotions (out of a set of 8). These primary basic emotions can be drawn as a circle divided up to 8 segments matching the 8 emotions with identical neighboring relations with their counterparts in the cone.

On panel 4 we visualized the user's primary basic emotion on a series of circles placed one above the other as they followed each other in the discourse. Showing the circles themselves would have made it confusing, so the center points of the segments are being showed only. We connected these points placed on above the other (chatter A's emotion to chatter B's emotion response), and we also connected each chatter's emotion to his/her next emotion being expressed in the dialogue.

This way the emotional loads of the sentences resemble a double chain of a DNA molecule, where the two sides are the two dialogue partners, and distance between two consecutive points is the distance between the two partner's consecutive sentences in the discussion. The lines beside show the intensity of the users' emotion (every second line belongs to chatter A, and every other to chatter B).

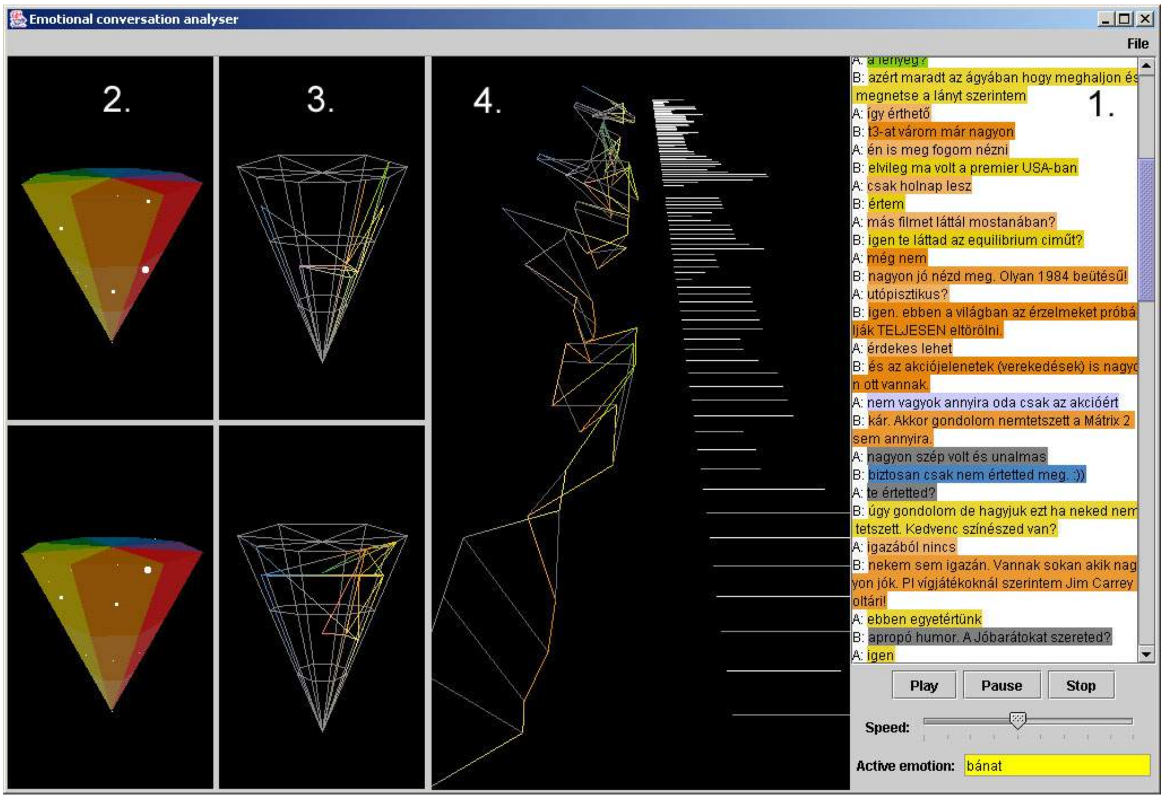


Figure 4. *Visualization tool for the emotion chat interface.*

3.2 Approaching empathy

Besides visualization, important statistical variables can be calculated from these data. If we measure the primary basic emotion differences between chatter A and B, we can conclude the willingness of chatter A to follow B's emotion. It is also possible to calculate the emotional intensity difference between A and B.

We can also calculate both the primary basic emotion and the emotion intensity difference between two consecutive sentences of A, and two consecutive sentences of B. These data show us how the chatters' emotional state was changing during the discussion, and how was it effected by the partner's emotional change. It is also informative to compare these data of the chatters to each other. A certain emotional behavior can be caused by the partner, or it can persistent during chat sessions with other partners.

It is not the objective of this paper to share our results in this field, but we found interesting correlations amongst those variables, which can provide further clues to create a computerized model of this very subjective notion called empathy.

4 Summary

In anthropomorphous computer interfaces it is important to utilize the emotional aspect of communication. In order to be able to detect and to calculate with human emotions in computer systems with textual input it is necessary to establish an emotion text database, and to examine the emotional aspect of human-human dialogue, in order to design the flow of communication. In this paper we were describing two applications created for that purpose. Besides visualization of the emotional aspect of communication we computed variables that can be calculated from the recorded emotional data. These variables can lead us to the measurement of empathy, probably the most important aspect of intelligent computer interfaces.

References

1. Brave, S; Nass, C (2003) Emotion in Human-Computer Interaction in *Handbook of Human-Computer Interaction* Under revision , Stanford University Press
2. Picard, R (1997) *Affective computing*. The MIT Press, Cambridge, Massachusetts, USA.
3. Plutchik, R (1980) A general psychoevolutionary theory of emotion. Plutchik, R. and Kellerman, H., eds.: *Emotion theory, research, and experience*. vol. 1, Theories of emotion. pp 3-33. Academic Press.
4. Plutchik, R (2001) The nature of emotions. *American Scientist*, **89(4)**, 344-350.
5. Tatai, G.; Csordás, A; Szaló, A; Laufer, L (2003) The chatbot feeling - Towards a usable emotional model for Internet ECAs. Proceedings of *EPIA'03 - 11th Portuguese Conference on AI*, Beja, Portugal, Springer Verlag.
6. Tatai, G.; Csordás, A; Kiss, Á; Laufer, L; Szaló, A (2003) Happy chatbot, happy user. Proceedings of the *4th International Working Conference on Intelligent Virtual Agents (IVA'03)*, Irsee, Germany, Springer Verlag.

The usability observation cap

G. Leitner, M. Hitz
Department of Informatics Systems, Alpen-Adria University of Klagenfurt, Klagenfurt, Austria

Abstract
In this paper we describe the development of a framework for mobile usability observation. The central device – an observation cap – is designed for hands free, non disturbing observation of tasks which are performed e.g. using a handheld device. We present the assumptions of our approach, discuss its benefits and weaknesses realized so far and sketch directions of future improvements.

Keywords
Mobile Usability, Observation Techniques, Wireless Observation.

1 Introduction
Due to the increasing availability and performance of mobile devices, the number of non-trivial applications implemented on such platforms is steadily rising. Thus, the need for usability testing these devices, mobile services and contextual conditions has gained importance. The related literature has come to the conclusion that *"usability evaluations should be done in the field"* (cf. e.g. [2]). At the beginning of mobile device testing, the approach has been to do testing in standard usability labs, placing the device to be observed at a fixed position on a desk in order to allow for systematic observations, especially video recording [4]. Test subjects had to operate the device in a way more or less alien to a realistic task setting in order to avoid impairing the observational environment.

This approach has been necessary due to technical constraints but is obviously influencing one of the main operational goals of usability testing, namely keeping the circumstances as close to realistic as possible.

At the risk of loosing controlled lab conditions (which are necessary for strong experimental assumptions) the benefit of conducting usability tests in more realistic circumstances is much higher, although one has to consider the limitations of usability testing in general [1].

Different alternatives to usability testing in lab conditions have been developed, e.g. the recording of click streams [2] or other forms of tracking data transfer. However, these approaches fail to consider important contextual information.

Another development related to mobile usability testing is to use specially designed equipment which enables mobile video observation [2, 3]. The equipment employed is still not satisfying because the cameras used either disturb the visual field and tactile operation area of the subject or – in other systems – is heavy weighted and has to be carried in an unnatural way.

Another disadvantage of this kind of systems is that no or insufficient data logging can be done. All of the approaches mentioned have some benefits, but do not support all sources of data observation a state of the art stationary usability lab is offering.

2 Initial Situation
In the context of a project on mobile e-learning (Mobilearn – www.mobilearn.at), the authors were confronted with the necessity to develop an infrastructure for conducting usability observations in realistic settings. In the course of the project focusing on the possibilities of collaborate learning in different (mobile) situations, an e-learning system has been developed and had to be tested against a set of pre-defined usability criteria.

Based on requirements analyses conducted by means of focus groups, we – not surprisingly – found that learning in mobile environments, e.g. busses, trains, park benches etc., is significantly different to sitting on a desk in an office or at home. The "atmosphere" in various environments has been identified as one of the important factors in learning situations. For instance, representatives of the target groups have stated within focus group discussions that (besides technical needs) flexibility of the system is one of the most important issues, since this represents the most significant weakness of existing systems in relation to paper based learning material. Learning is – in general – not a task where a person has a stable position for a long time. This aspect has been a major issue for the project so that the e-learning platform developed was based on PDAs. Consequently, for the conduction of the usability-evaluation of the system, we had to provide highest possible flexibility for the subjects.

Due to the lack of appropriate existing systems, the usability team decided to develop its own observation system aiming to fulfill the needs at hand:

- Highest possible flexibility regarding the selection of a learning location and position (sitting, standing etc.) of the subject.
- Lowest possible influence or disturbance of the subjects by technical equipment.
- Affordable equipment costs.

3 The observational Framework

3.1 Overview of the system
As shown in Figure 1 below, the system is based on a standard stationary usability lab with additional wireless apparatus of different kind. The user is (within the constraints of range of the different wireless systems) free in movement and position.

The following signals are transferred between user and control desk:

Data. The mobile device is connected through WLAN to a desktop computer, where a monitoring application is tracking the screen content of the mobile device.

Video. A light weight radio video camera is mounted on a standard baseball cap.

Audio. A radio microphone is worn by the user.

Conversely, signals transferred between control desk in the stationary lab and user are:

Audio. A radio microphone is connected to a receiver and earphones worn by the user. With this equipment, it is possible to instruct the user or give feedback if a question or problem arises.

Data. With the software on the control desk, it is also possible to convey visual messages to the user which are displayed on his mobile device.

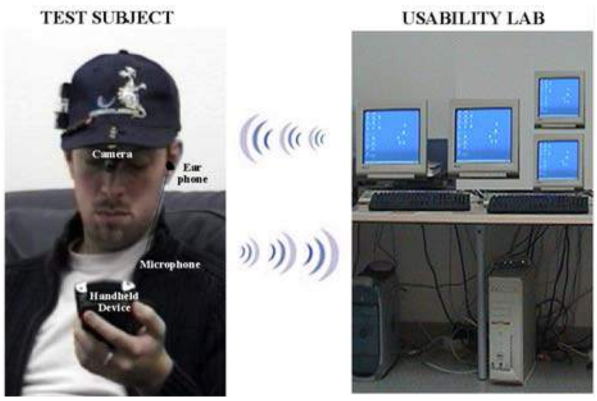


Figure 1. *Schematic overview of the system*

3.2 The prototype system in operation

The pictures in Figure 2 portray the prototype from different perspectives. The top-left Picture shows the position of the camera on the cap, the top-right picture shows the data the camera is recording and transferring to the control desk. As shown on this picture, details of the content of the mobile device are not discernible. However, this is not necessary, because the main purpose of the camera is to give information of the environment of the user, since the exact screen content of and data transferred by the mobile device are tracked by software. As it can be seen, the subject is sitting on a sofa in a very comfortable way (all pictures were taken in the real test situation). Movement is not restricted since everything necessary for the camera to work is mounted on the cap.

Figure 2. *Various perspectives of the observation system*

The bottom picture in Figure 2 includes the content of the PDA screen provided by the control software. The picture represents a combination of screen content and camera perspective of the user, mixed together by means of the video processing equipment which is part of the control desk.

Due to real-time processing, data and camera signal are always synchronous. The picture is displayed either on a computer screen, TV set or beamer. Therefore, content and controls can be recognized in detail and text is readable.

4 Preliminary Experiences

As stated, the observational equipment has been developed out of the need of evaluating a mobile e-learning platform. Therefore, we did not yet conduct any quantitative measurements, e.g. on the practicability and usefulness of the system. However, we got some qualitative results which turned out to be promising for further development of the system:

a.) None of the eight subjects refused wearing the cap. Most of them were interested in the technique.

b.) The adaptation of the cap to different persons did not cause any problems. The cap could be fitted to every subject taking part in the test.

c.) None of the persons mentioned inconveniences like uncomfortable pressure related to the cap, even though the test lasted for about one hour.

d.) The accomplishment of the tasks to be performed within the usability test has obviously not been influenced by the equipment, neither in positive nor in negative direction. This fact could be observed by comparing overall performance of a group of subjects which had to perform other tasks of the same platform in a standard lab condition (standard PC, remote video observation).

e.) Some of the subjects mentioned spontaneously possibilities of using the camera for more or less funny purposes. We interpret these statements as some kind of “liking” the system.

Besides these promising results, some shortcomings became obvious during the test. One of the major problems has been that the frequency range of the radio camera is too close to the WLAN frequency spectrum. This fact caused disturbances in the video signal of the camera when subjects worked on the mobile device intensively, e.g. by scrolling, writing on the touch screen etc. We could manage to reduce interference by changing WLAN channels of the mobile device, but the results where not fully satisfactory. The second problem was the limited range of the radio and WLAN systems. Due to the fact that we were limited to one transmitter and one receiver in each system, the flexibility of subjects to take a position or place they would like best could not be fully provided.

The range of the system has been additionally limited due to the fact that we had to conduct the tests in a ferroconcrete building. Another shortcoming is that only devices can be software tracked which are using a Microsoft Operating System (Windows CE, Windows PocketPC etc.). This fact is limiting the domain of observation, especially with regard to mobile phones.

5 Conclusions and Future Work

Based on the promising results as well as on the problems discussed, we are planning to develop a successor system with the following enhancements:

- Employing devices with different radio frequency to inhibit interferences.
- Establishment of a network cluster with a certain number of receivers and transmitters to provide a higher flexibility within a larger area.
- Usage of a broader software platform (e.g. Java) to increase the amount of trackable devices.

References:

1. Dicks, R.S. Mis-Usability: on the uses and misuses of usability testing. *In Proceedings of the 20th annual international conference on Computer documentation* (SIGDOC ’02) (Toronto, Ontario, Canada, October 20-23 ,2002) ACM Press, New York, NY, 2002, 26-30.
2. Kjeldskov J., Skov M.B., Als B.S. and Høegh R.T. Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field, *Lecture Notes in Computer Science, Volume 3160, Sep 2004*, Springer, New York et al., 2004, 61 - 73
3. Waterson, S., Landay, J.A., and Matthews, T. In the Lab and out in the Wild: Remote Web Usability Testing for Mobile Devices. *In Proceedings of the SIGCHI conference on Human factors in computing* systems (CHI ’02) (Minneapolis, Minnesota, USA, April 20-25, 2002). ACM Press, New York, NY, 2002, 796-797.
4. Weiss, S., Kevil, D., & Martin, R. (2001). Wireless Phone Usability Research. New York: Useable Products Company.

Ecological validity in usability testing

G. Lindgaard[1], K. BharrathSingh[2], B. Tsuji[1], & S. Khan[1]
[1]*Human Oriented Technology Lab (HOTLab) Department of Psychology, Carleton University Ottawa, ON Canada*
[2]*Nortel, 3500 Carling Ave. Ottawa, ON Canada*

Abstract

Identifying behavioral metrics to assess user preferences or user frustration can defy the most experienced researchers concerned with the evaluation of novel technologies. Ecological validity, known to be difficult to achieve in laboratory studies, poses an even greater challenge in such studies. A modern version of walkie-talkie technology was evaluated here. Situations were created in which we expected users to exhibit signs of frustration. Experiment 1 was equivocal with respect to the behavioral measures, but it needed improvements to ecological validity, which was addressed in Experiment 2. Even after priming participants with a dramatic video clip portraying a very urgent situation, no statistically significant results were obtained. Discrepancies between participants' comments and ratings of quality motivated Experiment 3 in which the volatility of subjective ratings was investigated. This manipulation did not result in changes in attitude towards the service. We conclude that the lack of ecological validity may account for this; participants who take part in laboratory studies who are paid for their services are unlikely to respond to the artificially created situation in a manner that resembles reality.

Keywords

Ecological validity, demand characteristics, walkie-talkie technology, Push-to-Talk over Cellular, PoC, channel switching.

1 Introduction

Innovative network-services are proliferating like never before. Small mobile devices are increasingly capable of providing access to sophisticated and complex telecommunications services. However, networks are complex, sometimes congested and not always stable. Therefore, in order to package services at different prices for varying service-level agreements, service providers need to know exact thresholds for acceptable delays in service delivery and also the point at which a deteriorating visual or auditory signal becomes unacceptable to users. Measures of perceived quality of service are thus important. Since many of these services are not yet available to consumers, these kinds of thresholds must be established in the laboratory. One challenge researchers conducting laboratory research face is that of identifying appropriate behavioral measures that indicate signs of user impatience or frustration. Even though physical stimuli are being manipulated and observable behavior is being monitored, the experimental objective is to understand the unobservable mechanism underlying that behavior [2].

Ecological validity has also long been recognized as another challenge in experimental laboratory research involving thinking human beings [11, 12]. The fact that participants agree to take part in an experiment; go to a laboratory to perform certain more or less meaningless actions; are observed throughout a session; and are paid at the end; is known to impact motivation [1]. Participants typically try to guess the experimental purpose and behave in ways they believe will support the experimenter's hypothesis as well as trying to impress the experimenter by performing as fast and accurately as they possibly can, which may say nothing about the same unobserved behavior performed in a natural setting. These add up to what Orne [11, 12] called "demand characteristics". Availability of sufficient monetary and human resources can occasionally alleviate some of the artificiality of laboratory research, especially in research involving time-, safety-, or mission-critical applications, for example studies of passenger evacuation in large aircraft [14]. Performing ecologically valid and reliable laboratory research that can effectively generalize to field settings with limited resources and short time frames remains a significant issue for human factors research in the telecommunications field.

We report three experiments here aiming to establish both behaviorally observable and opinion-based delay tolerance thresholds as well as grappling with ecological validity issues. Experiments 1 and 2 concern a novel mobile service to the mass consumer market referred to as "Walkie-Talkie-Like" (WTL) services . Basically this service allows someone with a wireless cellular device enabled with the 'walkie-talkie-like' functionality to use the device essentially like a walkie-talkie by simply pushing a button and talking. This technology is also generally known as Push-to-talk over Cellular (PoC). In Experiment 3 we investigate channel switching in digital TV systems. Space limitations force us to focus exclusively on experimental designs and methods.

2 Experiment 1

In a PoC call a user presses a phone button to call someone in a pre-defined group. After a variable delay, a one-way (or half-duplex) voice call is established. The called party responds with a similar button press, variable wait, and a verbal response. Once a user presses the 'talk' button, there is a delay of 0.5 - 6 seconds before they hear a "floor granted" tone indicating that they are now "on" [3]. This study attempted to identify overt behavior that could be attributed to annoyance with delays in PoC call setups. We studied three performance measures that, based on a literature review [6] and a set of field observations of our own behavior, we thought might capture user frustration [7]. Thus we asserted that frustrated participants might (a) press the PoC 'talk' button repeatedly, or (b) start to speak before receiving the 'floor granted' signal. Finally, (c) if sufficiently excited, participants might speak without pressing the 'talk' button at all. Response context was manipulated by varying the situational 'urgency', hypothesizing that high urgency situations would lower delay tolerance in receiving the 'floor granted' signal compared with a low urgency situation.

2.1 Method

Space limits prevent a detailed discussion of the technical and software programming details of all three studies. These may be obtained from the authors or in Lindgaard et al. [8, 9, 10]. Participants used a joystick and a microphone to simulate a PoC situation. Stimuli were presented on a computer screen. Software simulating network performance varied delays between 1 and 15 seconds approximating an Erlang distribution to mimic real-life network performance as much as possible [3]. The 20 participants (10 male, 10 female) were randomly assigned to the urgent or the non-urgent condition. Thirty sentences, roughly equal in length and complexity, were prepared for each condition and presented one at a time on the screen to avoid reliance on working memory. The experimenter controlled the presentation of sentences from a wireless keyboard in the adjacent control room. The time between the floor granted signal and onset of the participant's voice was timed; the frequency of joystick button presses was counted.

To start a trial, a sentence was shown on the monitor. Participants pressed the 'talk' button and held it down until hearing the 'floor granted' signal. They then read the displayed sentence aloud and released the talk button, ready for the next sentence to be displayed. Five practice trials were given first. In the high urgency condition participants were told to assume the role of an emergency coordinator responding to a suspicious package placed in a government facility. Sentences in this condition reflected this context (e.g. "The package is making a 'ticking' sound!") The low urgency condition was simulating a family conversation. The low urgency sentences reflected a banal context (e.g. "Do we need some milk or bread?"). The sentences in both conditions were presented in random order.

2.2 Results and Discussion

The overall response pattern approximated an Erlang distribution as intended. However, participants did not behave as predicted. Across all 20 participants (600 responses), a total of 59 repeated button presses occurred; no instances of no button speech were recorded. An average of 0.64 instances of early speech was observed. These were categorized by delay into Short; <3 sec; Medium; 4-8 sec; and Long; 9-12 sec and submitted to a split-plot ANOVA. Although the data suggested a main effect for delay, with longer delays leading to more button presses, this was not statistically reliable (see Figure 1). There was no effect of urgency; indeed, the number of button presses was slightly, but not statistically, higher in the low- than in the high urgency condition.

Overall, there was thus no evidence of the annoyance and frustration that we had experienced in our field study despite variable delays. Likewise, there were no signs of serious engagement with the stimuli. Evidently, the setup was not convincing; the necessary level of ecological validity to evoke the predicted behaviors was not achieved. This was addressed in Experiment 2.

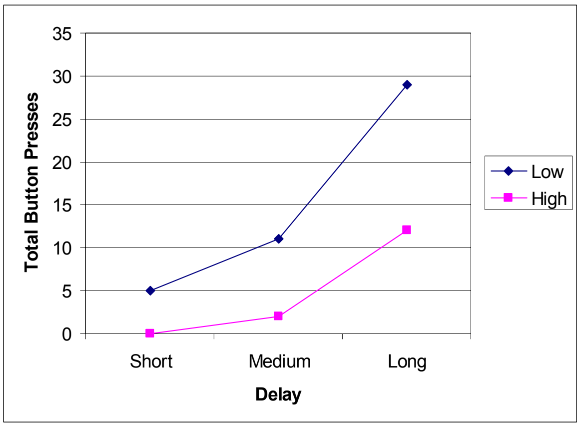


***Figure 1**: Number of button presses by Low- and High-Urgency conditions*

3 Experiment 2

Research on elicitation of emotions in laboratory studies indicates that stimuli such as pictures and films are capable of eliciting the target emotions [4, 5, 13]. Therefore, in an effort to increase ecological validity, we created a more extreme high urgency condition in which experimental trials were preceded by a short video clip related to the subsequent simulated conversation. The recorded conversation voices appeared to be quite excited and were accompanied by appropriate background sounds to enhance the effect. The computer equipment was hidden from the participant's view to avoid undesired effects. The video clip was shown on a 32" Plasma screen TV. Participants wore a cell phone-like monaural headset with a microphone for receiving and sending verbal responses. Randomized delays and behavioral measures were the same as before. Some 37 experimental sentences and five practice trials, approximately equal in length and complexity, were prepared for the two urgency conditions.

Twenty new participants were recruited from the same pool as before. After five practice trials, participants listened to a pre-recorded scenario before viewing the video clip reflecting the relevant urgency situation, either showing several firefighters in a warehouse battling an out-of-control fire or a non-urgent everyday event concerned with a business meeting. As before, the experimenter managed the session from the control room. The participant pressed the PoC Talk button, waited for the floor granted signal, read the appropriate sentence, and then released the button. Different recorded actors with background sound effects "responded". The participant then flipped to the next page and read the printed sentence after obtaining the floor granted signal. At the completion of all the sentences, the experimenter returned to the observation room to complete the debriefing. Conversation scripts were printed in 24-point font and placed in a 3-ring binder with one sentence per page. Unlike the sentence stimuli in Experiment 1, the stimulus sentences in Experiment 2 were presented in a logical order as a narrative that we anticipated would help to provide a more seamless and realistic context for the study.

3.1 Results and Discussion

As in Experiment 1, none of the data analyses approached statistical significance. The number of button presses increased slightly as a function of urgency, but again with the low urgency- resulting in more button presses as delays increased than the high urgency condition. The urgency manipulation did not affect subjective ratings either, although participant's comments clearly indicated that they were annoyed with the longer delays. This lack of a statistically robust effect is puzzling, especially since other researchers have successfully employed films to evoke particular emotions in the laboratory. It is possible that the very novelty of the PoC service may have influenced participants preventing them from displaying behaviors that prolonged usage in the field might. Either way, our efforts to increase ecological validity did not elicit the predicted behaviors. How to do this without disproportionate expenditure of resources remains unresolved. Because of the apparent discrepancy between subjective ratings and participants' comments, a Mean Opinion Score (MOS) questionnaire was administered several times in Experiment 3 to see if opinions changed as a function of varying delays.

4 Experiment 3

This experiment employed a paradigm involving channel switching in digital TV. In existing TV systems channel switching involves time delays of 30 to 50 msec. These are expected to increase to several seconds with the introduction of telco-based video services because delays will involve several switches between different multicast streams and delivery to the user's set-top box as well as digital buffering and decoding. Very few data on user tolerance of such delays are available [9]. We hypothesized that a variable delay of 1-7 seconds from the time a user selects a channel to it being displayed should lower participant measures of acceptability and of quality.

4.1 Method

Sixteen new participants (9 males, 7 females, average age 26 years), all of whom watch at least 10 hours of TV per week, were recruited. Nine used digital cable/satellite TV; the remainder analog cable TV. On average participants watched 16 hours of TV weekly; they had access to roughly 90 channels. The computer-driven digital TV software simulated a TV set to control start and stop of the program, select or change experimental trial blocks as well as managing channel switching delay times, or the time between key presses on the remote control and appearance of the video clip on the TV screen.

For each trial, a randomly-generated menu (channel listing) of seven consecutive channels chosen from a pool of 40 was displayed, one of which was highlighted at random. Participants selected the highlighted channel by pressing the relevant number keys on the wireless remote control. This activated the relevant video clip, simulating actual TV channel switching. Upon completion of a trial, the Experimenter switched the program back to the menu mode for the next trial. After each block of 15 trials, the program displayed questions one by one from a Mean Opinion Score (MOS) questionnaire. These concerned TV attributes such as picture quality, sound quality, and delays. The participant answered using the remote control.

Delays were calculated from the moment the participant pressed the second digit identifying a channel (e.g. 03) until the relevant video stream was displayed. The appearance of the video stream was delayed according to experimental condition. After the show had played for approximately 5 seconds the experimenter advanced to the next trial. Participants worked through four trial blocks following five practice trials. Three blocks employed a consistent delay of one-, three-, or five seconds for each trial, and one employed a randomly variable delay of one, three or five seconds. Blocks were presented in a Latin square to minimize serial order effects. Two participants saw each order.

4.2 Results and Discussion

The results showed that overall quality ratings and mean ratings of individual characteristics addressed in the MOS questionnaire deteriorated as delays increased from 1 to 5 seconds. Thus, even with consistent-quality video clips, quality was perceived to differ as a function of delays. The variable-delay condition did not differ from the 3.0 sec fixed-time delay. One would have thought that participants who were all seasoned TV watchers may not be as patient as our results suggested when surfing TV channels at home. While our lab is comfortable, it does not resemble a cozy lounge room in a private house. It is thus possible that Orne's [11, 12] reservations about laboratory studies applied here and perhaps overwhelmed any systematic effects that delay might exert on participants' behavior. The demand characteristics of our laboratory situation meant that participants could react with indifference to otherwise annoying delays--they knew they were taking part in a laboratory study and that they would be compensated for their time.

5 Conclusion

While the data from all three experiments suggest that our efforts to simulate network delays were successful both in the PoC and the digital-TV contexts, it remains unclear how to measure the frustration we know we felt when using ourselves as participants in our field study preceding the research reported above. Even with a very carefully designed experimental setup, a better understanding of ways to lend laboratory experiments a higher level of ecological validity in novel technologies and services remains a tremendous challenge. The issue of discrepancies between how participants tell us they felt and their numerical ratings, as well as answers to questions in a more formal questionnaire underscores the lack of reliability of, and individual differences in, scores in opinion scales. Although the lack of statistical significance in the studies we report here is somewhat unconventional, we felt the greater good was to be served by reporting these results, questioning as they do the application of traditional empirical methods in some areas of applied research.

Permission by Nortel to publish this paper is gratefully acknowledged.

References

1. Adair, J.G. (1984). The Hawthorne effect: A reconsideration of the methodological artefact. *Journal of Applied Psychology*, **69**(2), 334-345.

2. Chow, S.L. (1987). Science, ecological validity, and experimentation. *Journal for the Theory of Social Behaviour*, **17**(2), 1-8.
3. Ericsson, Siemens, Motorola, Nokia (2003). *Push-to-Talk over cellular (PoC); User requirements V1.1.0* (2003-08) PoC Release 1.0 Technical specification. 2003.
4. Gross, J.J. & Levenson, R.W. (1995). Emotion elicitation using films. *Cognition and Emotion*, **9**(1), 87-108.
5. Kim, J., Lee, J. & Choi, D. (2003). Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions. *International Journal of Human-Computer Studies*, **59**, 899-940.
6. Lindgaard, G., Khan, S, & Tsuji, B. (2004a). Push-to-Talk over Cellular (PoC): *A Review of Relevant Literature*, Unpublished HOTLab Technical Report HCIGLSKBT04POC03050527.
7. Lindgaard, G., Khan, S, & Tsuji, B. (2004b). *HOTLab observations of the Telus Mike/Direct Connect (PoC) Service,* Unpublished HOTLab Technical Report HCIGLSKBT01POC01050527.
8. Lindgaard, G., Khan, S, & Tsuji, B. (2004c). *Initial call setup delays in Push-to-talk over Cellular (PoC),* Unpublished HOTLab Technical Report HCIGLSKBT02POC02050527.
9. Lindgaard, G., Khan, S, & Tsuji, B. (2004d). *Telco-based video delivery: A Review of Relevant Literature*, Unpublished HOTLab Technical Report HCIGLSKBT05TVD01050527.
10. Lindgaard, G., Khan, S, & Tsuji, B. (2004e). *Channel switching delays in digital TV systems*, Unpublished HOTLab Technical Report HCIGLSKBT03DTV01050527
11. Orne, M.T. (2002). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Prevention & Treatment*, **5**(35), (reprint of article below).
12. Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, **17**(11), 776-783.
13. Philippot, P. (1993). Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and Emotion*, **7**(2), 171-193.
14. Wilson, R., Thomas, L. & Muir, H. (2003). *VLTA emergency requirements research evacuation study*, Work Package 3, Task 3.1 Report – Test plan for evacuation trials, European Commission, VERRES Project.

Video-based quantification of escape behaviors in groups of fish exposed to a simulated visual predator

J.D. Salierno[1,2] and A.S. Kane[1,2]
Aquatic Pathobiology Center
[1] *Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine*
[2] *Virginia-Maryland Regional College of Veterinary Medicine, College Park, USA*

Abstract

We evaluated predator escape behaviors in groups of mummichog, *Fundulus heteroclitus*, before and after a simulated visual aerial predator event using a computer-based video tracking system. The videography system recorded 15 seconds of baseline movement and 15 seconds of movement after the presentation of wood bird silhouettes that "flew" over the exposure arenas. Video data were then analyzed using Videoscript Professional© 3.1 and custom software algorithms designed by our laboratory for the analysis of fish movement. Simulated predator presentation stimulus elicited a significant response whereby 95% of groups exhibited a startle response followed by cessation of movement ($p < 0.001$). Movement cessation after predator presentation lasted for 8.2 (± 0.3) seconds. Additionally, groups of fish responded with significant decreases in average velocity, shoal velocity, interactions, percent movement, and percentage shoaling, along with significant increases in percentage of aggregation ($p \leq 0.10$). This predator avoidance hardware stimulus, when coupled to a videography and tracking system, accurately quantified predator escape behaviors in the mummichog.

Keywords

Group behavior, aerial predator avoidance, mummichog, fish

1 Introduction

The ability to evade predation is essential to the success of an organism. An animal captured and consumed by a predator has it's cumulative fitness abruptly terminated, and therefore strong selective pressure exists to evolve escape mechanisms to mitigate predation [1]. As a result, organisms need to maintain a constant and consistent suite of predator avoidance tactics in order to react to threats and to survive. However, these behaviors must also be adaptive and maintain the ability to change depending on the environmental conditions, predator and threat type, and social organization among and between species. Different fishes have evolved over evolutionary time to live in aggregations, shoals, and schools as an attempt to increase fitness and survival [2].

Teleost fishes utilize schooling and shoaling behaviors to detect, confuse and escape predators, as well as locate mates and resources, and for aid in migration [2, 3]. Fish respond to an attack with a suite of behaviors that minimizes contact with the predator and maximizes predator confusion [4]. Social groups tend to become more compact to reduce the amount of staggering in the presence of a predator [1, 5-8].

In order to respond to alterations in predation pressure, fish must actively adapt avoidance strategies. Natural selection would presumably favor individuals capable of recognizing the degree of threat posed by a predator [9]. This would avoid expending more energy than required for a possible threat, keeping energy allocation in balance. It has been demonstrated that fish can adapt predator avoidance behaviors as predation level changes [9]. This phenotypical plasticity allows fish to vary inter-individual distances, allowing different degrees of group cohesiveness to balance relative advantages and disadvantages of schooling behavior [8, 10].

We subjected groups of mummichog, *Fundulus heteroclitus*, to a simulated visual aerial predator in order to determine and quantify group avoidance behaviors using a video-based tracking system. Mummichog are essential for the success and productivity of many mid-Atlantic estuarine marsh environments [11]. These small cyprinidontid teleosts regulate the trophic structure of marsh systems through control of invertebrate populations (crustaceans and annelids), as well as serving as prey for larger predators such as blue crab, striped bass, and shore birds [12]. Therefore, mummichog provide a good model for the investigation of predator avoidance behavior because they live in shallow estuarine habitats and are vulnerable to aerial predation threats.

2 Methods

A computer-based video tracking system designed for fish behavioral recording [13] was used to record 15 seconds of baseline movement before, and 15 seconds of movement after presentation, of the visual "fly-by" stimulus. The stimulus consisted of wood bird silhouettes that "flew" over the exposure arenas via a system of motor-driven monofilament lines and pulleys. In order to test for habituation to the stimulus, 12 groups of 5 fish were presented with the predator stimulus 4 times with 5 minutes between each presentation. Analog video data were digitized on a Macintosh platform (PowerPC G5 2GHz, 4GB SDRAM) at 3 fps and cut into 15 second clips using Adobe Premiere© 6.5. Behavioral endpoints, pre- and post-predator threats, were then quantified using Videoscript Professional© version 3.1 in conjunction with custom software designed for the study of fish movement [Salierno et al., *in prep*]. Data from the first presentation were analyzed for alterations in group behaviors resulting from presentation of the stimulus. The sequence of 4 repeated presentations was analyzed to investigate for habituation to the stimulus. Response variables included immobility duration, percent stopping, group percent movement, shoaling velocity, percent startle, shoaling and aggregation, and nearest neighbor distance and angle (NND and NNA). Data were then analyzed separately and compared for differences at $\alpha \leq 0.10$. Wilcoxon 1-tail signed rank sum tests were used to evaluate differences in behaviors and habituation pre- and post-stimulus presentation.

3 Results

3.1 Habituation to Repeated Visual Threats

Repeated "fly-by" stimuli resulted in significant reductions in both percent movement and velocity in groups of fish. Groups of fish demonstrated significantly reduced percent movement in all 4 trials ($p < 0.002$, Figure 1) and significantly reduced velocity in 3 out of the 4 trials ($p < 0.01$, Figure 2). The consistent response over repeated trials demonstrated that fish did not habituate to the predator stimulus over time. These results additionally demonstrated that a simulated aerial predation threat can achieve a constant behavioral response over repeated presentations.

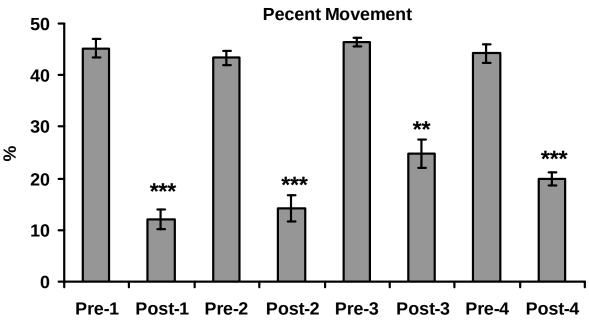


Figure 1. *Alterations in the percent movement in groups of mummichog before (pre) and after (post) a simulated predator stimulus. Trials were run 4 consecutive times separated by 5 minutes in order to test repeatability of the response. **, p = 0.01 and ***, p = 0.001.*

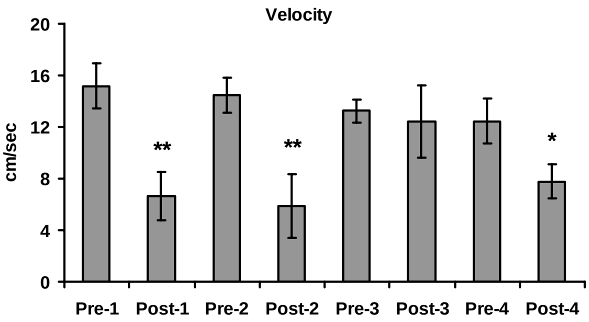


Figure 2. *Alterations in group velocities of mummichog before (pre) and after (post) a simulated predator stimulus. Trials were run 4 consecutive times separated by 5 minutes in order to test repeatability of the response. * , p =0.05, and **, p = 0.01.*

3.2 Group Responses to a Simulated Visual Threat

The visual "fly-by" stimulus elicited a significant response whereby 95% of the groups of fish stopped swimming ($p \leq 0.10$). This cessation of movement post-predator presentation, lasted for 8.2 (± 0.3) seconds, and 41% of groups evoked a startle response similar to a flash expansion behavior [14]. Groups of fish responded with significant decreases in shoal velocity, interactions, percentage shoaling, and shoaling NND, along with significant increases in percentage of aggregation ($p \leq 0.10$, Table 1). No significant differences in NNA between groups of fish were observed. Overall, groups of fish tended to slow down and become less active and organized, i.e., there was a shift from shoaling to aggregating, with fewer interactions. However, in cases where fish are shoaling, a significant increase in polarization was observed (shoaling NND, $p = 0.013$, Table 1). These results demonstrated that movement and shoaling behavior decreased significantly when exposed to a simulated predator.

Behavioral Endpoint	**Pre-stimulus mean (±S.E.)**	**Post-stimulus mean (±S.E.)**	**P value**
Shoal Velocity, (cm/sec)	7.8 (± 0.56)	4.8 (± 1.8)	**0.097**
Interactions, #	37.9(± 3.75)	23.9(±4.03)	**0.010**
Percent Shoaling	32.0(±4.45)	5.30(± 2.59)	**0.0017**
Shoaling NND, (cm)	17.9(± 1.86)	7.98(± 3.08)	**0.013**
Percent Aggregated	50.0(± 7.23)	72.3(± 10.8)	**0.006**

Table 1. *Alterations in group and shoaling behaviors pre- and post-predator stimulus.*

4 Discussion

Punctuated predation is characteristic of natural systems where contact between predator and prey is episodic, such as with fish prey and avian predators [1]. Mummichog reside in shallow, tidal, estuarine habitats and are always subject to avian predators. Predator response behaviors of mummichog observed in this study, associated with the visualization of an "avian predator," were quantitatively assessed using our system hardware and software. It is unclear that experimental fish discerned the wood figures as avian predators. However, startle response in fish is a unique, innate escape mechanism to avoid adverse visual or acoustic stimuli and, as evidenced in the present study, is clearly associated with the presentation of a shadowless, overhead visual stimulus. All fish tested in this study, upon presentation of the simulated predation threat, notably reduced activity as a response. An animal is more visible when moving than when stationary, and decreasing activity lowers predation rate [15].

Startle responses to aerial models in fish have been documented previously, but quantitative differences in response to aerial models in groups of fish using a computer tracking system have not been previously investigated. Groups of mummichog responded with a startle response, which resulted in a rapid dispersion of fish. In addition, groups of fish displayed higher swimming velocities and spent less time motionless after threat presentation.

Results of this study prove that visually-cued "predation avoidance" is a discernable and quantifiable behavioral response in groups of mummichog. These data have formed the basis of a "visual stimulus" endpoint in our laboratory that can provide significant and reproducible results using the mummichog model. This endpoint is an integral part of ongoing studies that investigate behavioral and movement alterations that are associated with contaminant and stress exposure in fish. Behavioral data from these controlled laboratory experiments are important in developing a better understanding of low-level exposure effects on fish, and serves to bridge laboratory studies to field exposure scenarios. In addition, the use of ecologically relevant, behavioral endpoints provide quantitative data to show possible detrimental individual- and population-level effects other than mortality.

The authors thank Madeline Sigrist and Patrick McGlone for their assistance with the physical construction and data collection. Support for this project was provided byThe U.S. Environmental Protection Agency (STAR grant #R82-8224) and by the State of Maryland. All experiments complied with the regulations set by the Institutional Animal Care and Use (IACUC) committee of the University of Maryland (Protocol #R-00-36B).

5 References:

1. Ryer, C.H. and Olla, B.L. (1998). Shifting the balance between foraging and predator avoidance: the importance of food distribution for a schooling pelagic forager. *Environmental Biology of Fishes*, **52**, 467-475.
2. Radakov, D.V. (1973). *Schooling in the Ecology of Fish*. John Wiley and Sons, New York.
3. Masuda, R. and Tsukamoto, K. (1998). The ontogeny of schooling behaviour in the striped jack. *Journal of Fish Biology*, **52**, 483-493.
4. Litvak, M.K. (1993). Response of shoaling fish to the threat of aerial predation. *Environmental Biology of Fishes*, **36**, 183-192.
5. Lima, S.L. and Dill, L.M. (1990). Behavioral decisions made under the risk of predation - a review and prospectus. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **68**, 619-640.
6. Magurran, A.E. and Pitcher, T.J. (1987). Provenance, shoal size and the sociobiology of predator-evasion behavior in minnow shoals. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **229**, 439-465.
7. Pitcher, T.J. (1983). Heuristic definitions of fish shoaling behavior. *Animal Behaviour*, **31**, 611-613.
8. Sogard, S.M. and Olla, B.L. (1997). The influence of hunger and predation risk on group cohesion in a pelagic fish, walleye pollock Theragra chalcogramma. *Environmental Biology of Fishes*, **50**, 405-413.
9. Helfman, G.S. (1989). Threat-sensitive predator avoidance in damselfish-trumpetfish interactions. *Behavioral Ecology and Sociobiology*, **24**, 47-58.
10. Nemtzov, S.C. (1994). Intraspecific variation in sand-diving and predator avoidance-behavior of green razorfish, *Xyrichtys splendens* (Pisces, Labridae) - effect on courtship and mating success. *Environmental Biology of Fishes*, **41**, 403-414.
11. Nixon, S.W. and Oviatt, C.A. (1973). Ecology of a New England salt-marsh. *Ecological Monographs*, **43**, 4, 463-498.
12. Abraham, B.J. (1985). Species profiles. Life histories and environmental requirements of coastal fishes and invertebrates (mid-Atlantic): Mummichog and striped mummichog. Washington D.C., U.S. *Fish and Wildlife Service*. FWS/OBS-82/11.40.
13. Kane, A.S.;Salierno, J.D.;Gipson, G.T.;Molteno, T.C.A. and Hunter, C. (2004). A video-based movement analysis system to quantify behavioral stress responses of fish. *Water Research*, **38**, 3993-4001.
14. Pitcher, T.J. and Parrish, J.K. (1993). *Functions of schooling behavior in teleosts. Behavior of Teleost Fishes*. London, Chapman and Hall.
15. Magnhagen, C. (1988). Changes in foraging as a response to predation risk in 2 gobiid fish species, *Pomatoschistus minutus* and *Gobius niger*. *Marine Ecology-Progress Series*, **49**, 21-26.

Identification of patterns in cod behavior

G.K. Jonsson[1], V. Thorsteinsson[2], & M.S. Magnusson [3]
[1] *Human Behaviour Laboratory, University of Iceland, University of Aberdeen & Pattern Vision*
[2] *The Marine Research Institute, Reykjavik*
[3]*Human Behaviour Laboratory, University of Iceland & Pattern Vision*

Abstract
The Atlantic cod (*Gadus morhua L.*) is historically one of the most important commercial species known [2]. The behavior of this species, important for fisheries, research and stock assessment, is in many ways masked by extensive horizontal and vertical dispersion in its habitat [3]. The potential of using the Theme software in behavior studies of commercial fish stocks was tested using data from tagging experiments with adult cod in Icelandic waters. The cod was tagged with Data Storage Tags (DSTs) with memory capacity of up to 260,000 records measuring temperature and depth at 10 minutes intervals. The time series were prepared for T-pattern analysis, including detection and delimitation of tidal influence in the data and event basing raw data according to predefined events. A high number of temporal patterns were detected, patterns of repeated vertical movements and speed and acceleration changes. Number of specific temporal patterns were also identified within and across vertical movements of individual cod. Special attention is given to a) behavioral patterns that are tidal-wave related and the comparison of tidal-wave related patterns to other patterns and b) behavioral patterns related to temperature and depth data.

Keywords
Cod, *Gadus morhua*, Behaviour, temperature fronts, DSTs, T-patterns

Introduction
The report discusses a new approach to analyze behavior such as vertical movements of tagged cod in Icelandic waters. The approach, known as T-pattern detection, has successfully been used within other research fields but never before in this particular field. The data presented show that specific temporal patterns can be identified within and across vertical movements of individual cod in relation to environmental parameters.

What Theme does
Theme looks for relationships between events. It takes into account the order and relative timing (critical interval relationship). If the critical interval is less than would be expected by chance, it defines a pattern called T-pattern. Theme starts with simple patterns and gradually adds them together to form more complex patterns. Less complete patterns do not survive, longer chains are "fitter"; as the patterns recombine and grow, only the fittest survive.

Method

Organization of work
- Selecting and preparing time-series of behavior information
- Event basing data
- Changing original data files into compatible files for Theme
- Running data in Theme
- Select patterns of interest
- Repetitions with adjusted search parameters
- Verification of pattern detection by superimposing the events or patterns on to original data

Event Basing
Event basing depth data involved partitioning time-series into defined periods/events (e.g. events concerning vertical movements in the range of <10m, 10 to 20 m, > 20 m).
The events defined for cod vertical movement were the following:
- Slow: Vertical movement within defined criteria
- Fast-up: Speed of vertical movement exceeding a defined criteria
- Fast-down: Speed of vertical movement down exceeding a defined criteria
- Acc-up: Vertical acceleration up exceeding a defined criteria
- Acc-down: Vertical acceleration down exceeding a defined criteria

Correcting data for tidal waves
The tidal wave comes to the shores of Iceland from south and goes one circle around the island in 12.4 hours. The timing of the max and min of the waves varies accordingly and magnitude of amplitude is also variable being greatest off the west coast of Iceland and smallest of the East coast. 12 selected cod time series spanning from 3 to 4 weeks and 7 spanning from 9 to 12 months were selected for analysis. An algorithm was developed for the correction of the sea-tide (see Figure 1).

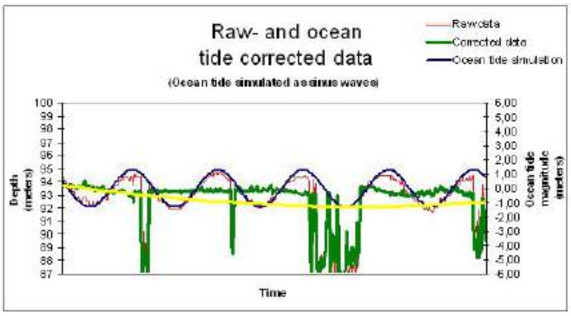


***Figure 1**. Raw data (red) minus adjustable correction data set (blue) resulting in corrected data (green). A data set resulting in the best fit for a straight line over a pre-defined time span decides the magnitude and the time phase shift of the tide (an example of data correction of file L 548).*

Temperature
Either the fish stays within a certain ambient temperature range or the ambient temperature changes at some rate. When Event basing time series of temperature data were partitioned into ranges (-1°C to 0°C, 0°C to 1°C etc.) or events that describe rate of change (x °C / time).

Changes in depth caused by tidal wave

In this study the tidal wave is partitioned into detectable events for the Theme program, such as a) slack waters at high tide, b) slack waters at low tide, c) change from low tide to high and d) change from high tide to low.

Results

A high number of temporal patterns were detected in the cod DST data. These patterns were of repeated vertical movements, speed and acceleration changes as well as resting at the same defined vertical level. An example of pattern of an individual cod is displayed in Figure 2. and an example of a pattern of vertical movements detected across a number of individual fish is displayed in Figure 3.

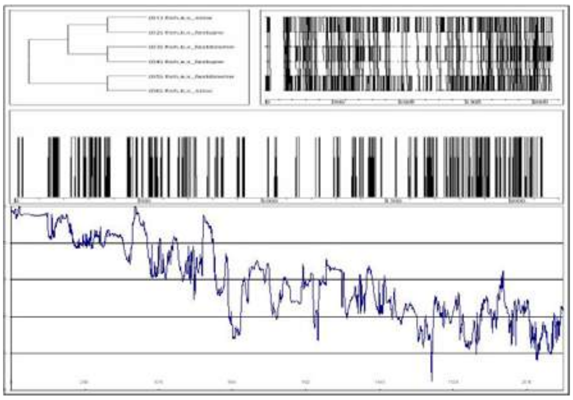


Figure 2. *An example of one of 120 patterns found in the time series (top) and raw data depth measures of Cod dstC493 in a 2 week period from 28.05.03 to 11.06.03 (bottom). The fish is migrating towards feeding grounds after spawning. At the beginning there is a rest period indicated by 16 hours of a tidal pattern after which the vertical migrations increase considerably through the period when the fish is changing depth range from 90-120m down to 250 - 400m.*

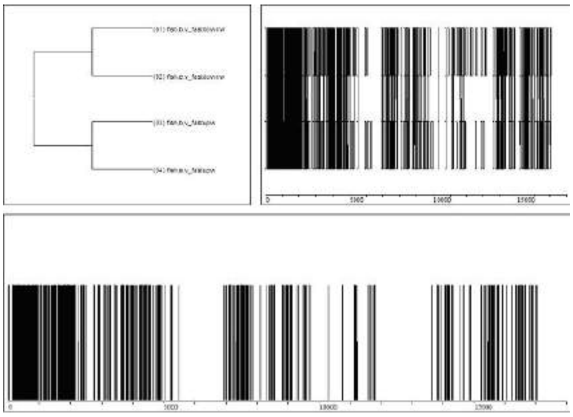


Figure 3. *Example of speed/acceleration/slow patterns found across cods.*

Discussion and Conclusion

Preliminary results indicate that a Theme analysis can make a significant contribution to the analysis of cod behavior, offering an increased advantage to view and understand hidden patterns within a large number of data points. A high number of temporal patterns were detected, patterns of repeated vertical movements and speed and acceleration changes. Number of specific temporal patterns were also identified within and across individual cod vertical movements. Future objective is to further search for patterns of vertical movement in relation to environmental parameters emphasizing a) behavioral patterns that are tidal-wave related and b) behavioral patterns related to temperature and depth data.

The detection and delimitation of tidal patterns in the DST data is very important for investigations into behavior of fish living in tidally influenced waters and also for tidal location of fish [1]. One of the tasks of the pre-project was to try to correct time-series of depth for tidal influence and then to have the tidal wave parameter as an independent actor besides the depth-time series. According to our findings in these pioneer studies however, correcting time series of depth for tidal influence would be very problematic if not detrimental because of the variability and irregularity in the amplitude of the tidal wave. The application of an imperfect tidal wave model to correct for tidal influence may therefore cause a secondary tidal pattern in data that may cause false results in pattern-analysis of Theme.

References:

1. Hunter, E., J.N. Aldridge, J.D. Metcalfe & G.P. Arnold. 2003. Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables. I. Tidal location method. *Marine Biology*, **142**, 601-609.
2. Kurlansky, M. 1997. *Cod: a biography of the fish that changed the world*. Walker and Company, New York. 294 pp.
3. Pálsson, Ó.K. & V. Thorsteinsson. 2003. Migration patterns, ambient temperature, and growth of Icelandic cod (*Gadus morhua L.*): evidence from storage tag data. *Canadian Journal of Fisheries and Aquatic Sciences*, **60**, 1409-1423.

Sex, Flies and no Videotape.

J.D. Armstrong, D.A. Baker, J.A. Heward and T.C. Lukins
School of Informatics, University of Edinburgh, Edinburgh

Abstract

The characterization and quantification of animal behavior is one of the most resource intensive activities in modern biological science. One of the best known examples is the courtship ritual of the small fly *Drosophila melanogaster* which has been discussed in the academic literature for almost 100 years. It is an exemplary behavior for the functional dissection of the nervous system as it is a natural behavior that requires the animal to integrate and learn from multiple sensory modalities. Object tracking software is extremely good at tracking single moving objects and describing their behavior particularly where previous examples have been well characterized. However, tracking multiple, interacting objects is more challenging. Here we describe the application of a bounding box segmentation algorithm that works well with Drosophila courtship behavior. We show that once segmented, just a few simple parameters are sufficient to classify the behavior in real-time, requiring no off-line storage or post processing.

Keywords
Drosophila courtship mating tracking

1 Introduction
Drosophila melanogaster males (and although less obviously, also the females) display a complex courtship ritual that uses a range of cues from multiple sensory modalities, is naively instinctive but then subsequently modified by experience. The behavior itself was first reviewed in 1915 [7] who cited what were presumably the earliest studies in the field by Barrows in 1907, Lutz in 1911 and Payne in 1911. We are now close to 100 years since these first publications in the field and interest has increased each year. There are now established laboratory methods to disentangle many of the interlinked sensory cues resulting in a very naturally embedded assay for perception of sensory stimuli and behavioral adaptation (for review see [5]). When combined with the genetic toolbox used by Drosophila researchers to mutate genes or modify cellular physiology in identified neurons (for review see [1]), we have one of the most sophisticated assays for mapping molecules to naturally occurring behavior.

However, there is a drawback - it is expensive, almost prohibitively so. In the 100 years or so of research into courtship behavior, the methods employed in laboratories to quantify the behavior itself have remained essentially the same. Courtship activity is affected by time of day therefore video recording has been employed to capture datasets in parallel which are later analyzed in serial by expert observers sometimes with the use of ethnographic annotation software packages. Otherwise the methods are more or less identical (e.g. [3]).

There are a number of behavioral events that can be recorded but most studies focus on three measures. The first, and most general measure is the Courtship Index (CI) and is the combined percentage of time the male fly under investigation spends exhibiting any recognizable courtship behavior towards a target fly or object during the observation period. These activities include orientation of the body, following, tapping with the legs, licking, wing extension or attempted copulation. The second measure sometimes reported is the Sex Appeal Parameter (SAP). This is the percentage of the observation time that wig extension is observed. Finally, latency is the time taken from introduction to the first interaction between the two animals.

Thus, for medium throughput genetic screens, assays that segregate a group of animals based on their behavior are most commonly employed. As insects move through the apparatus their behavior influences choices that alters their exit or final location. Perhaps the best known example of these is the choice point of the olfactory test [8] where flies have a choice of moving towards one odor or another. After they have chosen, flies are simply collected and counted. More complex examples are the mazes used to measure phototaxis or gravitaxis (for review see [2]) where the flies move sequentially through a series of T shaped choice points that are orientated with respect to light or gravitational stimuli and finally exit into a series of collection tubes and are simply counted. Such techniques allow for large numbers of flies to be tested but lack any analysis of individual behavior.

For mammals, almost all behavioral experiments require detailed quantitative and qualitative analysis of the behavior of individuals. The emphasis for automation has been on rodent behavior with many methods employed for tracking and describing behavior from video footage. Here, simple methods for the movement of objects in two dimensions can deal with many of the problems of occlusion and fitting active contours to the animal can be used to describe in high detail and with high precision the body movements over time. When combined these techniques can describe accurately the movement of a rat through a maze and identify specific behaviors.

The analysis and annotation of human behavior has also attracted much recent research within the Computer Vision community. Particularly in determining potential dangerous and anti-social activity as recorded in real-time on CCTV footage. This is primarily a two-stage process. Firstly reliably tracking and differentiating multiple people must be robustly handled [4]. Secondly, on the basis of good segmentation, analysis and statistical modeling of the relative interactions between people can occur, leading to a classification of the type of interaction [6,9].

Clearly an automated, real-time, behavioral tracking system for Drosophila would be extremely useful. Beyond the obvious benefit to anyone who has spent a considerable proportion of their graduate studies scoring video footage of insect courtship there are a number of other very important benefits. Observers will differ in their opinions of exactly when an insect is courting. Over long periods, observer fatigue is reported. Collecting very detailed information about courtship activity such as a fine breakdown of all the component behaviors, the velocities of the animals involved is possible but extremely time consuming. We have no empirical evidence for many of these issues thus cannot evaluate their importance but they are widely acknowledged by experts in the field.

Here we describe the development of a solution directed at all of the problems described for courtship behavior tests in Drosophila. Our aim is to produce a measure for courtship activity that is not observer dependant, will scale to medium throughput activities (i.e. 1000's of tests per day with a small team) and will also collect some of the technical information such as velocity that is difficult to do with manual annotation of video footage. We designed our test to demonstrate whether we could measure CI in a variety of samples and how well this correlated to that of human observers.

1 Methods

1.1 Video Capture

We chose to use a networked video camera (Axis 205 www.axis.com) as it supports a range of resolutions and frame rates all of which can be controlled by sending a simple URL call from a web browser or application. The camera then sends back a video file in MJPEG format. No specialist lighting or optics were used. We used incident lighting in the observation room and moved the apparatus around to limit any shadows or high contrast reflections. Unless stated otherwise, the resolution used was 320x240 pixels, captured at a rate of 10 frames per second.

1.1 Drosophila courtship

We focused our attention on normal behavior. Thus the flies described here are all Canton-S wild-type. To obtain a range of CIs we used mated and virgin target female flies in combination with naïve and experienced males (e.g. experienced males will not court mated females thus giving a low CI).

Flies were raised on standard cornmeal medium at 25'C on a 12 hour light:dark cycle. All courtship recordings were done in the morning and within 4 hours of lights on.

We used cylindrical plexiglass observation chambers, 10mm in diameter and 8mm deep. The chamber allowed for enforced separation of the insects prior to recording. For details refer to [5]. Flies were aspirated into the chambers without anesthesia and given 15 minutes to relax after transfer. The divider between the male and the target fly was then removed and activity recorded for 5 minutes.

1.1 Computation

Clearly any algorithm can be performed in real-time given sufficient computational power. We used a range of operating systems and hardware all of which was purchased between 2003 and 2004. Given that some of our development computers are multiple CPU architectures we reconfirmed all real-time tests using single CPU computers. 1) Apple 12" Powerbook 867 MHz G4 running Mac OS 10.3 with 640MB Ram. 2) Dell Optiplex GX270 2.6GHz P4 running Windows XP Pro with 512GB Ram.

1.1 Expert annotation

We designed a small JAVA annotation program that replayed the video and recorded keyboard activity to a log file. We recruited several experts to observe video footage and indicate via pressing keys when courtship activity started and stopped during the 5 minute period and thus a CI. For a small number of samples we also recorded wing extension thus obtaining an SAP score.

1 Results

1.1 Image Processing and Thresholding

The grayscale video footage we obtained did not require background subtraction since there was significant intensity difference between the fly and the background to identify it. Adaptive thresholding would have suited this task but was found to be too computationally intensive for live video work. Instead, a pseudo-adaptive thresholding procedure was implemented that split frames into smaller non-overlapping regions before mean thresholding each. We found a significant speed improvement with no loss of sensitivity for the flies (data not shown).

1.1 Segmentation

We searched the image space for blobs of approximate fly shapes and return their centers and longest angles to give a rough position and longitudinal body axis (flies can be approximated to ovals). Identifying flies within the binary image is achieved by filtering the list of blobs using one or more of the following criteria; 1) Position within frame, 2) Area of the blob.

1.1 Tracking

Flies move in 3D and are equally likely to be walking on the ceiling of the observation chamber as they are on the floor. Hence occlusion of flies directly on top of each other is common. As a result, methods for dealing with occluded blobs often applied to rodents such as erosion, water-fill etc did not perform well. We tested the bounding box algorithm of [5] and found it performed exceptionally well with the footage we obtained (see figure 1 and supplementary video file). For details of this, please see the accompanying short paper by Heward et al in these proccedings.

Using out default settings for video capture, tracking mistakes are extremely rare with approximately one error per 3000 frames.

1.1 Descriptive analysis

The most distinctive features of the fly courtship ritual (see video for an example) are following behavior, the relative angles of body orientation relative to the other fly and the extension of a single wing during the courtship song. We decided to focus our attention on relative body orientation and following behavior.

We found the relative body orientation and distance between the flies to be very predictive of the expert annotation. Just on the distribution of relative angle and relative distance on a frame by frame basis the majority of courting and non courting frames can be distinguished (figure 2).

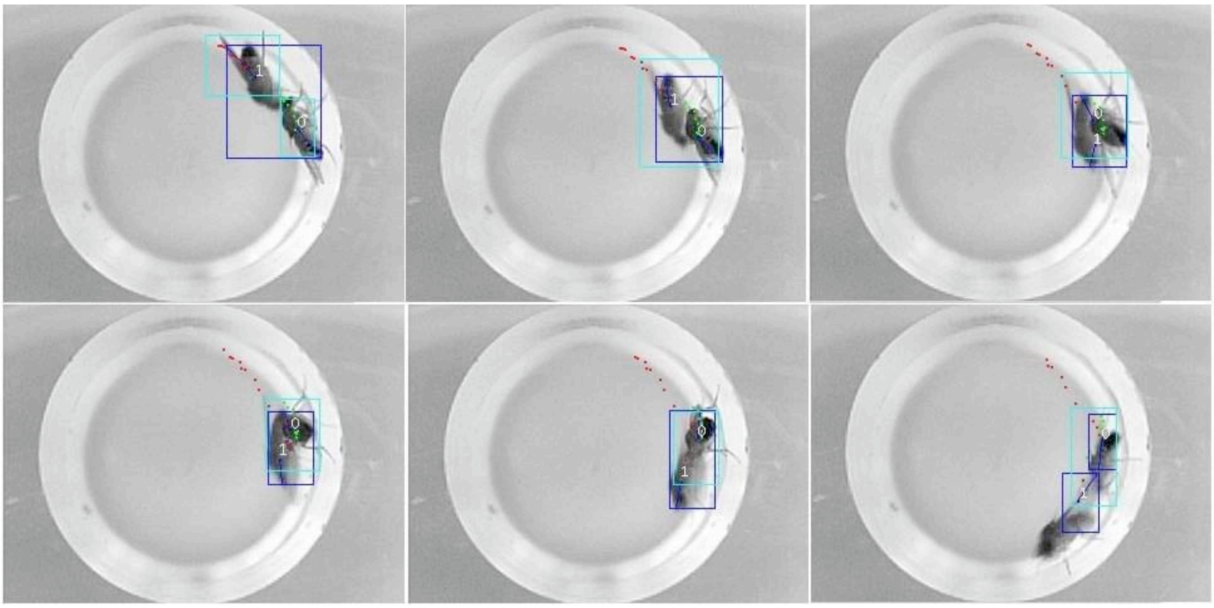


Figure 1 *Six successive frames with bounding boxes following the flies. In each frame, the previous boundary is in cyan and the current boundary in dark blue. When the boxes merge (frames 1-5) the position of each fly is estimated using K-means (K=2).*

.

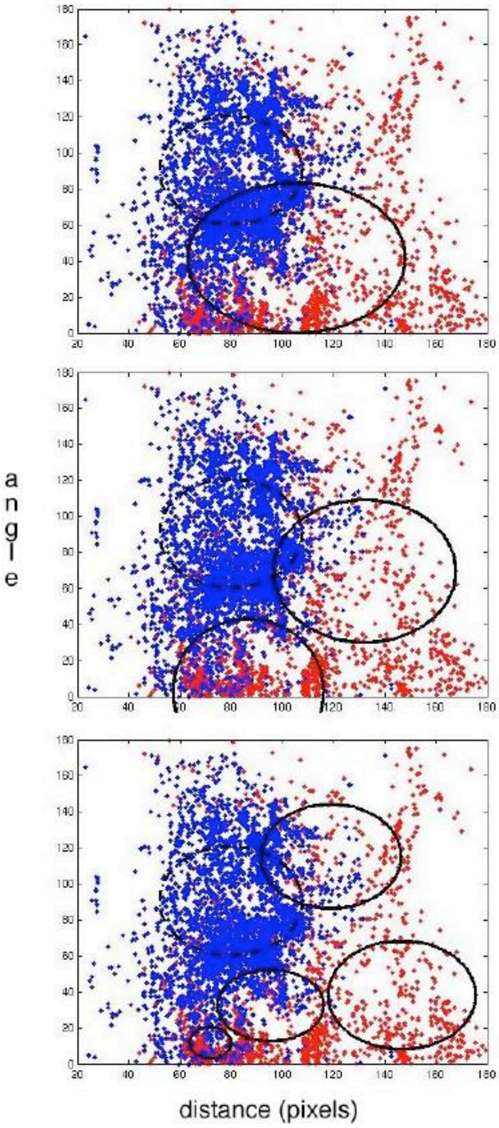


Figure 2 *95% Gaussian distributions fitted to the relative angle and relative distance parameters of frames annotated as courting (blue) or not courting (red) by one expert in one example video.*

1.1 Automatic Classification

We implemented and tested a Genetic Algorithm and K-means Clustering approach to extract predictive rules using the expertly annotated footage. The learning algorithms were given the X, Y distances, the relative distance and the relative angle between the longest axes for each blob, trained on two sets of expertly annotated datasets leaving residual datasets aside for validation. The expert annotation was offset by up to one second to compensate for slight time lags when the experts react to changes in behavior. For the GA, parameters from the previous nine frames were also available for rule design.

The CI values returned using the learned rules had correlations with the two expert datasets used of 96% and 98%. This is in line with the observed variability between our experts. Further validation is on-going.

1.6 Performance

The computational complexity increases with the number of pixels, the frame rate and with the number of objects to be segmented in the arena (all in a linear manner). Preliminary experiments suggest that our existing packages will handle 20-30 frames per second and three fly – ménage a trios - type experiments (figure 3) on standard desktop systems.

The tracker normally operates in one of three modes. The first is a live capture mode where the algorithm aims to match the speed of the live video feed. If insufficient compute or network resources are available then frames will be dropped and this is logged. In the second mode, the tracker works on previously recorded MJPEG files and will process them as fast as resources permit. In the third mode we can replay the video file at any specific rate and observe the tracker in operation.

The supplementary video file is in the third of those modes, replay frame rate was set to 10 frames per second (same as the original recording). This was recorded on a 867Mhz G4 based Apple laptop using SnapzPro2 (www.ambrosia.com) demonstrating that this system can easily handle both the tracking and video capture tasks in real time. The specific example was chosen as the flies are very actively courting resulting in a large number of partial occlusion events, highlighting the value of the bounding box algorithm.

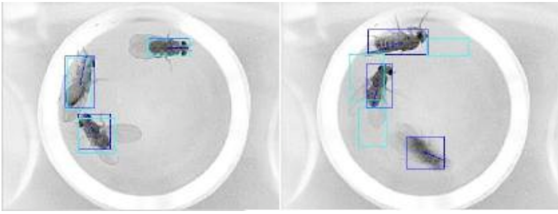


Figure 3 *The algorithms scale to multiple flies, e.g. ménage a trios type experiments.*

2 Conclusions

We have demonstrated that we can segment and process the video data at a rate suitable for live analysis of *Drosophila* courtship behavior. The application of the bounding box tracking algorithm has proven extremely useful in dealing with the occlusion problem. Moreover both the Genetic Algorithm and K-means classifier systems result in simple rule sets that perform within the accuracy bounds of experts. When all are combined the system can capture and process in real time to produce a reliable CI thus freeing research staff to focus on much finer details.

Acknowledgements

We thank Jean-Christophe Billeter, Tony Dornan, Stephen Goodwin, Kevin O'Dell and Jo Young for their help as expert annotators and system testers. This work was supported by grant no. S18944 from the BBSRC (JDA).

References

1. Armstrong J.D. and Goodwin S.F. (2004) Studying neuronal function using the Drosophila genetic system In Davies RW and Morris B (Eds) *Molecular Biology of the Neuron* Oxford University Press
2. Beckingham K.M., Texada M.J., Baker D.A., Munjal R. and Armstrong J.D. (2006) Genetics of Graviperception in Animals. *Advances in Genetics.* in press.
3. Ejima, A., Smith, B.P., Lucas, C., Levine, J.D. and Griffith, L.C. (2005) Sequential learning of pheromonal cues modulates memory consolidation in trainer-specific associative courtship conditioning. *Curr Biol* **15**, 194-206.
4. Fuentes, L.M. and Velastin, S.A. (2001) People tracking in surveillance applications in 2nd IEEE International Workshop on Performance Evaluation on Tracking and Surveillance, PETS 2001 (Kauai, Hawaii-USA, 2001).
5. O'Dell, K.M. (2003) The voyeurs' guide to Drosophila melanogaster courtship. *Behav Processes* **64**, 211-223.
6. Oliver, N.M., Rosario, B. and Pentland, A.P. (2000). A Bayesian Computer Vision System for Modelling Human interactions. IEE Transactions on Pattern Analysis and Machine Intelligence, vol.22
7. Sturtevant, A.H., (1915). Experiments on sex recognition and the problems of sexual selection in Drosophila. *Anim. Behav.* **5**, 351–366.
8. Tully T. and Quinn W.G. (1985). Classical conditioning and retention in normal and mutant Drosophila melanogaster. *J Comp Physiol [A]* **157**, 263-277.
9. Van Vu T., Brémond F. and Thonnat, M. (2003). Automatic Video Interpretation: A Recognition Algorithm for Temporal Scenarios Based on Precompiled Scenario Models. Proceedings of ICVS

Measuring nighttime spawning behavior of chum salmon using a dual-frequency identification sonar (DIDSON)

K.F. Tiffan and D.W. Rondorf[1]

[1]*Western Fisheries Research Center, U.S. Geological Survey, Cook, Washington, USA*

Abstract

In this paper we describe the use of a dual-frequency identification sonar (DIDSON) to compare daytime and nighttime spawning behaviors of chum salmon (*Oncorhynchus keta*). The DIDSON forms near-video-quality images using sound instead of light, which permits the collection of behavioral data in complete darkness. Using the DIDSON, we observed chum salmon enganged in digging to construct spawning nests, courtship behavior, chasing, and spawning both during the day and night. There were no statistical differences between the diel measures of these behaviors. The DIDSON proved to be an effective tool for collecting behavioral information with minimal intrusion to the fish.

Keywords

Chum salmon, spawning, DIDSON, nighttime, behavior

Introduction

Mate selection, courtship, and spawning behavior of chum salmon have been well described [2,7,8,10]. Female chum salmon will dig 4-6 nests that are collectively called a redd [5], which may take 30-40 h to complete [10]. A female fish selects an area for nest construction and excavates a depression in the riverbed by turning on her side and making 4-6 powerful flexures of her tail to dislodge and clean gravels. After spawning, digs made to cover the eggs occur in rapid succession, are smaller in magnitude, and are made immediately upstream of the nest. Visual cues presumably play a role in female selection of a male mate based on body morphology and coloration [8]. A dominant male fish will court a female fish by swimming back and forth over her caudal peduncle region and by quivering—a lateral display in which the male makes a series of high-frequency, low-amplitude body undulations [10]. Under high spawner densities, satellite males will hold position downstream of a spawning pair and will rush in and attempt to fertilize eggs during the spawning event [7,9]. Both male and female chum salmon will also chase intruding fish to defend their mate and territory.

To date the diel aspects of chum salmon spawning behavior have received limited attention because of the difficulty of collecting observational data at night without affecting fish behavior. Other investigators have used hydrophones to record the sound and frequency of digs made by chum salmon in an experimental channel and concluded that spawning activities occurred both during the day and at night [10]. However, this observation was never confirmed for fish in the wild. We used a new technology—a dual-frequency identification sonar (DIDSON; Sound Metrics, Incorporated; Figure 1) [1,6,11]—to describe the diel spawning behavior of wild chum salmon in the Columbia River. Our objectives were to determine if chum salmon spawn at night in Columbia River habitats, and if there were diel differences in spawning behavior.

Figure 1. *DIDSON used to collect behavioral information on spawning chum salmon in 2003.*

Methods

We collected diel behavioral data on spawning pairs of chum salmon in a small side channel of the Columbia River downstream of Bonneville Dam. This channel is about 250 m long, has a mean width of 44 m, and longitudinal gradient of <1%. The bed is composed of gravel (4-75 mm) and cobble (76-150 mm) substrate, and water velocities are generally <1 m/s. A significant amount of chum salmon spawning has occurred in this area for last 5 years [3,12].

We used a DIDSON to observe spawning pairs of fish for which the female was actively digging a nest in the company of a courting male. The DIDSON was originally developed for military use in harbor surveillance and mine detection, however, its utility in fishery investigations has been recently recognized. The DIDSON has advantages over traditional underwater video cameras for collecting behavioral information on fishes. First, because the DIDSON forms images using sound instead of light, it can image in complete darkness and zero-visibility water. Underwater video cameras cannot image in darkness without the aid of lights, which may disrupt normal behavior, or the use of infrared light, which although undetectable to fish, has poor penetration (<1 m) in water. Second, the DIDSON can image to a distance of 12 m in its high-frequency mode (1.8 MHz) and to 24 m in its low-frequency mode (0.9MHz). The typical range for an underwater video camera in relatively clear water is only a few meters. This means that the fish can be observed at greater distances while minimizing potential behavioral disturbance. Finally, the horizontal field of view of the DIDSON is generally greater than that of underwater video cameras. For example, at a distance of 10 m, the field of view is 5 m, which allows observation of fish over a greater spatial area.

The DIDSON measures 30 cm long by 20.5 cm wide by 17.5 cm high and weighs 5.5 kg in air but is slightly buoyant in water. The DIDSON forms images by simultaneously transmitting and receiving 96 acoustic beams set 0.3º apart in the horizontal plane. The horizontal field of view is 29º and the vertical field of view is 8º. Images collected by the DIDSON are sent via a cable to routing hardware where images can be output to video equipment or to a laptop computer using an Ethernet connection. We mounted the DIDSON on a weighted bracket that enabled us to adjust its aspect and tilt. We set it perpendicular to the current to continuously observe pre- and post-spawning behavior of chum salmon in about 1 m of water from 13 November to 6 December 2003. Fish were observed with the DIDSON from a distance of 5 m, and we set the field of view to begin at 3 m from the DIDSON and extend to a maximum distance of 7 m. Images of spawning behavior were collected at a frequency of 1.8 MHz and downloaded to a laptop computer at a rate of 8 frames per second. Images produced by the DIDSON appear as though the spawning pairs were viewed from above even though the DIDSON was oriented laterally toward the fish (Figure 2, top panel).

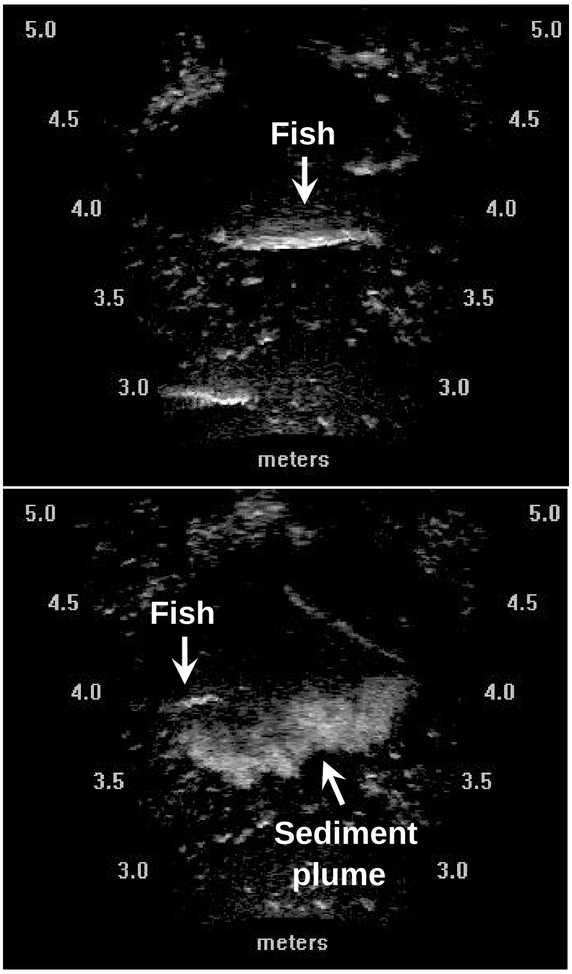


Figure 1. *An example of an acoustic image produced by the DIDSON of a digging event by a female chum salmon. The top panel shows the female fish before digging and the bottom panel shows the plume of sediment produced by the digging event.*

Therefore, the field of view appears somewhat like a triangular-shaped plane in which fish positions, movements, and behaviors can be spatially quantified. In addition, sediment plumes and substrate movement resulting from digging can be detected and measured (Figure 2, bottom panel). We used a measuring tool in the DIDSON's software to determine the size of each fish observed.

We used the DIDSON to record the number of digs made by females per 1-min interval, the number of tail crosses made by males per 15-min interval, and the number of times females and males chased intruding fish per 15-min interval. Observation of a spawning pair was discontinued 1-2 h after a spawning event. For pairs that did not complete spawning, we ended observation when the male stopped active courtship or when digging by the female declined or became erratic. We also used the DIDSON software to display of grid of known scale on the laptop screen which we used to determine the spatial area of each nest. We did this by marking the location of each dig on an acetate transparency attached to the laptop computer screen. The locations of all digs made to construct a nest were then digitized and incorporated in a geographic information system. We estimated the size of each nest by drawing the smallest possible polygon that encompassed the cluster of dig locations and calculating the area.

We marked the location of completed nests with a differentially-corrected global positioning system (GPS) and selected a new spawning pair for observation regardless of time of day. New spawning pairs were selected visually during the day, whereas the DIDSON was used to search for a new pair during the night. We used GPS to ensure the same nest was not studied twice. The local vicinity of the nest site was not revisited for 30-40 h (i.e., a female's spawning cycle) to avoid spawners that had been previously observed.

Because acoustic images did not allow us to see the release of gametes during a spawning act, we periodically used an underwater video camera during the daytime to confirm DIDSON images of each spawning activity, including gamete release. Video images from the underwater camera also allowed us to observe aspects of spawning behavior that were not apparent on DIDSON images (e.g., "quivering" by the courting male). No underwater video images were used in our analyses.

We divided the observations on digging, tail crossing, and chasing frequencies into daytime (0.5 h after sunrise to 0.5 h before sunset) and nighttime (0.5 h after sunset to 0.5 h before sunrise) categories. We did not analyze data collected during the 1-h periods encompassing sunrise and sunset due to the small sample sizes during those times and because we observed no marked changes in fish behavior during those times. Data were analyzed to test for diel differences in digging, tail crossing, and chasing frequency using two-sample *t* tests.

Results

We used the DIDSON to observe 14 different pairs of chum salmon continuously for 4.5 to 33 h. Mean female size was 80 cm (range 66 to 95 cm) and mean male size was 90 cm (range 83 to 97 cm). Females and males were active at nighttime and daytime, and were observed chasing, digging, and courting prior to spawning.

Spawners of both genders chased intruders during nighttime and daytime nest construction, often when intruding fish were over 1 m away. The courting male primarily chased other chum salmon. Females also chased conspecifics, but at night they also chased smaller fish that were probably peamouth *Mylocheilus caurinus* and northern pikeminnow *Ptychocheilus oregonensis*. We found no significant differences between mean daytime and nighttime chasing behavior for either gender. We also observed no diel differences in tail crossing by courting males. We did observe quivering on our underwater video camera, but it was not consistently distinguishable with the DIDSON.

The DIDSON enabled us to observe each time a female chum salmon dug during nest construction. A plume of sediment, which appeared as a light cloud behind the fish, was apparent each time a fish dug (Figure 2). During the pre-spawning period of nest construction, female chum salmon engaged in continuous digging of variable intensity during both daytime and nighttime, but digging frequency between the two periods were not significantly different. The post-spawning covering digs were distinct from pre-spawning digs in that they were rapid, relatively small, and the female would quickly return to the point of egg deposition after each dig. Females generally made cover digs in a cluster or arc-shaped pattern immediately upstream of where eggs were deposited. Females dug over a broad area to construct nests that averaged 2.1 m^2 (range 1.1 to 2.9 m^2).

We used the DIDSON to document spawning events for 10 chum salmon pairs, of which 3 pairs spawned twice. Of the 13 spawning events, 9 occurred at night and 4 occurred during the day. We could readily identify a chum salmon spawning event with the DIDSON because the courting male would swim quickly along side the female and both fish would remain motionless for 5-10 s, tail crossing immediately ceased, and the female would begin making covering digs. Often one or more satellite males would rush in to spawn when the female and courting male began spawning.

Discussion

The DIDSON proved to be a powerful tool for exploring the nighttime spawning behavior of wild chum salmon in a natural stream. Although the DIDSON produced acoustic images, we were able to document spawning behaviors (i.e., pre-spawning digging, tail crossing, male and female side by side and motionless, and post-spawning digging) that were consistent with those described in the literature [8,10]. Deployment of an underwater video camera was useful for confirming each spawning behavior that we observed with the DIDSON. In addition, underwater video allowed us to observe the male courtship behavior of quivering, which we could not reliably observe with the DIDSON. We recommend the use of underwater video to confirm behavioral images obtained with a DIDSON where practical.

The DIDSON has many advantages over traditional equipment for collecting behavioral information on fishes. Fish can be observed at a distance that minimizes the potential to alter behavior. In our study, we observed chum salmon at a mean distance of about 4.5 m. We never observed any aversion to the presence of the DIDSON. Observation at this distance coupled with the way in which the DIDSON builds images enabled us to spatially quantify fish movements, digging locations, and nest size that is not possible with underwater video equipment. The DIDSON was easily deployed and could be moved about in our study area. However, it was difficult to maintain the DIDSON's position in water current that exceeded 1 m/s without additional stabilization measures such as using weights and guy wires. Although the DIDSON can be deployed from shore as in this study, it can also be deployed from remotely operated vehicles or from sleds towed by a boat [11].

In spite of its advantages, the DIDSON does have some limitations that are worth considering. First, the DIDSON requires 110 V power, which may not be available in field locations. We overcame this by using a portable generator as our power supply. Second, the DIDSON generates about 1 gigabyte of data per hour when a frame rate of 7/s is used in its high-frequency mode. Data storage can be addressed by saving data to a high-capacity external storage device, regularly downloading data to a DVD, or recording data in the DIDSON's time laps mode or at a lower frame rate. We addressed this issue by having observers manually record behaviors (e.g., digging frequency) as the data were collected, which enabled us to quickly summarize data and precluded the need for subsequent reanalysis. Third, the DIDSON will not work well in environments with entrained air bubbles, which will cause acoustic interference. Finally, the DIDSON's relatively high cost ($75,000 US) may preclude its use by some research organizations.

Nighttime spawning by Pacific salmon is largely an undocumented phenomenon, which is likely due to the limitations and difficulties of collecting observational data in darkness without disturbing the fish. The DIDSON enabled us to document nighttime spawning in chum salmon for the first time in the wild. It appears that once fish begin nest construction and courtship, these activities continue regardless of diel period. Although visual cues, such as body coloration, are important in mate selection in Pacific salmon [4] and chum salmon [2, 7], we speculate that once spawning sites and mates are selected, non-visual cues (i.e., tactile, auditory, olfactory) become more important during the remainder of the spawning cycle allowing chum salmon to spawn at night as well as during the day.

Supported by the Bonneville Power Administration (Contract 00004701)

References:

1. Belcher, E., Matsuyama, B., and Trimble, G. (2001). Object identification with acoustic lenses, pages 6-11. *Proceedings of Oceans 2001 Conference,* Honolulu, Hawaii. 5-8 November.
2. Duker, G. (1982). Instream orientation and species recognition by Pacific salmon. Pages 286-295 *in* E.L. Brannon and E.O. Salo, editors. *Salmon and trout migratory behavior symposium*. University of Washington, Seattle, Washington. 3-5 June.
3. Garland, R., Tiffan, K., Rondorf, D., Skalicky, J. and Anglin, D. (2003). Assessment of fall Chinook and chum salmon spawning habitat near Ives and Pierce islands in the Columbia River. *Report to the*

Bonneville Power Administration, Contract 00004701, Portland, Oregon.

4. Groot, C., and Margolis, L., editors. (1991). *Pacific salmon life histories*. University of British Columbia Press, Vancouver.
5. Helle, J. (1981). Significance of the stock concept in artificial propagation of salmonids in Alaska. *Canadian Journal of Fisheries and Aquatic Sciences* **38**:1665-1671.
6. Moursund, R., Carlson, T., and Peters, R. (2003). A fisheries application of a dual-frequency identification sonar acoustic camera. *ICES Journal of Marine Science* **60**:678-683.
7. Schroder, S. (1973). *Effects of density on the spawning success of chum salmon (Oncorhynchus keta) in an artificial spawning channel*. Master's thesis, University of Washington, Seattle.
8. Schroder, S. (1981). *The role of sexual selection in determining overall mating patterns and mate choice in chum salmon*. Doctoral dissertation, University of Washington, Seattle.
9. Schroder, S. (1982). The influence of intasexual competition on the distribution of chum salmon in an experimental stream. Pages 275-285 *in* E.L. Brannon and E. O. Salo, editors. *Proceedings of the Salmon and Trout Migratory Behavior Symposium*, School of Fisheries, University of Washington, Seattle, Washington. 3-5 June.
10. Tautz, A., and Groot, C. (1975). Spawning behavior of chum salmon (*Oncorhynchus keta*) and rainbow trout (*Salmo gairdneri*). *Journal of the Fisheries Research Board of Canada* **32**:633-642.
11. Tiffan, K., Rondorf, D., and Skalicky, J. (2004). Imaging fall Chinook salmon redds in the Columbia River with a dual-frequency identification sonar. *North American Journal of Fisheries Management* **24**:1421-1426.
12. Van der Naald, W., Clark, R., Brooks, R., and Duff, C. (2003). Evaluation of fall Chinook and chum salmon spawning below Bonneville Dam. *Annual Report to the Bonneville Power Administration*, Contract 99BI15006/99BI15007, Portland, Oregon.

Validation of methods of measuring physical activity using The Observer software and Actiwatch in young children

C P Khokhar[3], R Juyal [1,3], U Dhingra[2],A Dutta, M Osmany [1], R E Black [2] , A Sarkar [1], P Dhingra [1], P Verma[1] ,M Black[4], R.M.Angulo-Kinzler[5] , VP Menon [1],J Kumar[1] ,GS Hiremath[2], S Sazawal[1,2]

[1]*Center for Micronutrient Research, Annamalai University, India*

[2] *Johns Hopkins Bloomberg School of Public Health, Baltimore, USA*

[3.] *Department of Psychology, Gurukul Kangri University, Haridwar, India*

[4] *School of Medicine. University of Maryland, Baltimore, Maryland, USA*

[5]. *Division of Kinesiology, Center for Human Motor Research, University of Michigan, USA*

Abstract

Objective To validate the method of measuring physical activity using Actiwatch (Mini-Mitter Co.) in young children by actual observation using observer software.

Research Methods and Procedures Among 100 children aged 13-76 months from peri- urban population in Delhi, one hour direct observation of activity at home using observer software was undertaken while the child was also wearing an omni directional sensory monitor on Arm (A) as well as Leg (L). The records from active watch and The Observer file were clipped for exact same time duration. CARS scores were estimated from observation.

Result The activity counts from watch were correlated well with CARS (C) scoring **A**: β 0.18 (p=0.07), **L** β 0.30 (p=0.003), Total score (A+L): β 0.31, (p=0.001).

Conclusion Our results indicate that activity assessment using actiwatch especially if arm and leg counts are combined are valid measures of child activity and correlate well with observational assessments.

Keywords

Physical Activity, Actiwatch activity monitor, Activity in Young children, Validation activity.

Introduction

Physical activity is a multidimensional human behavior. Because of its complex nature, it is difficult to assess precisely under free-living conditions. Children's patterns of physical activity have been historically assessed by direct observation and questionnaires. Activity monitors were developed in response to the lack of reliability of self report measures, and the intrusiveness of direct observation. With the advent of small accelerometer-based activity monitors, the ability to monitor children's activity has improved greatly[1]. Recent advances in integrated circuitry and memory capacity have produced sensitive, unobtrusive, accelerometer-based devices that measure continuously the intensity, frequency and duration of movement for extended periods. However the data evaluating qualitatively and quantitatively the scores produced by actiwatch especially in young children are lacking. There is also not a consensus or data on merits and demerits of tying the monitor on arm versus leg. To fill this gap in data, we evaluated actiwatch assessments with monitor tied to both arm and legt with a simultaneous observation of activity by an observer.

Methods

Study was conducted in Sangam Vihar, a peri urban locality of New Delhi, India. The study was performed as an add on to an ongoing clinical trial in which 24 hours activity of children was being assessed. In a sub sample of 100 children aged 13-76 months in addition a one hour observation of activity was performed. Parents were explained the study and informed consent was sought. Following which appointment for observation time was made.

The Observer Training

Five observers were trained to record the activities in The Observer software through direct observation. The Observer training included the use of videotapes and group discussions. Training videotapes involved children performing various activities. Rules were established to accommodate variations in specified activities or unexpected activities. Training was continued till agreement between observers was more than 80%.

Actiwatch

Actiwatch is an accelerometer that is capable of sensing any motion with a minimal resultant force of 0.01g/rad/s. Actiwatch stores this information as an activity counts. The weight and dimensions of this device are small (27×26×9 mm), weight (17g) and durable casing looking more like a watch. It is omni directional, resulting in sensitivity to motion in all direction. Epoch length: The period of time Actiwatch will accumulate activity counts before saving the sample and setting the counter to zero. In this study 1 min epoch length were set in all the actiwatches. There also is an event recorded that was used to indicate start and end of observation.

The Observer Software

The Observer 5.0 (Noldus Information Technology, B.V), an automated manual event recorded installed in a laptop for collecting, collating and analyzing observation data. In this software we configured activity codes in to five broad area i.e. **Gross Motor activity, Fine motor activity, Behavioral states, Contact with care giver and Location. Gross Motor activity included 16 activities** (Being carried, Lying, sitting squatting, hanging, crawling, running, climbing, rough-tumble, dancing, kneeling, stand with support, walk alone, stand alone, swinging and walk with support**), Fine motor consist 12 activities** (eating, drinking, self care, dressing, game with little movement, game with active movement, mild movement with object, vigorous movement, repetitive movement, No movement, out of view and other fine movement) **Behavioral states consisted 3 states**(

Sleeping, alert and fussy),**Contact with care giver consist 5 states**(Beyond arm's length, within arm's length, Holding hands, Clinging and touching with hand) and **Location consist 4 states** (Home restricted, home non-restricted outside restricted and outside non-restricted). Within each group the codes were mutually exclusive and start of second would mean termination of the first.

CARS (Children's Activity Rating Scale)[2]

The CARS is a five-level rating system designed to evaluated children's physical activity in observation and has been previously validated. In order to make a quantitative comparison observation of activity was transformed into CARS score. The five categories of physical activity established were Stationary 1, Stationary 2, Slow Movement, Moderate Movement and Fast Movement.

Procedure

Prior appointments were taken before visiting the child home for activity assessment. Activity assessments were done in the natural environment of the child i.e. at the home of the child. A trained scientist gave necessary instruction to the caregiver and loaded the identification information (user Identity information and numbers, age, gender), ensure data and time which automatically changes was right and set epoch length. Actiwatch was tied on arm and leg of the child for 1hour. Simultaneously the same identification information was loaded in The Observer software. As soon as the marker was pressed on the actiwatch the observer scientist also started the coding. Time of actiwatch and laptop was synchronized. Activities of the child were coded using The Observer 5.0 (Noldus Information Technology) with the help of laptop for one hour.

Figure 1. Observer scientist tying Actiwatch to the child.

Analysis

Actiwatch data was downloaded with the help of actimeter. Using a in-house developed software in visual basic the activity score, start time and end time of each epoch was appended to a oracle data base which had child ID information already. From direct observation, we selected an interval of 1 minute to match the epoch of acti watch and within that estimated time spent in each activity codes. For each interval then scoring using Children's Activity Rating Scale (CARS) was estimated and these data were also appended to Oracle data base. Data were analyzed as means ± SD of the scores. Pearson correlation and multiple regression analyses were performed using STATA, version 8 (Stata Corp, Union Station, TX). To further evaluate consistency children were categorized by quartiles of CARS score and within these groups actiwatch scores were compared.

Results

A total of 100 children had 1 hour of direct observation while wearing a sensor on arm and leg. The mean (± SD) age and SES score for 100 children were 49.0±15.8 and 7.22± 2.35 respectively. The activity counts from actiwatch were correlated well with the CARS scoring (Table 4) and were not affected by age [**A**: β -0.04, (p=0.68), **L:** β -0.04, (p=0.72), **T:** β -0.05, (p=0.63)]. When all the children were categorized into 4 quartiles based on CARS scores [(Group1:0-25), (Group 2: 25-50), (Group 3: 50-75), (Group 4: 75-100)] actiwatch mean scores across four categories showed a significant rising trend (Table 1, 2, 3).

Table 1. *Mean, SD of CARS & Total score in four percentile group.*

Categorization	CARS Scores	Total Scores
	Mean (SD)	Mean (SD)
Group1 (0–25)	14326.95 (2735.36)	31838.32 (27251.94)
Group2 (25-50)	17355.17 (348.35)	37818.24 (24268.97)
Group3 (50-75)	18197.30 (209.63)	39196.72 (30593.84)
Group4 (75-100)	19298.73 (688.73)	78731.88 (45460.01)

Table 2. *Mean, SD of CARS & Arm score in four percentile group.*

Categorization	CARS Score	Arm Scores
	Mean (SD)	Mean (SD)
Group1 (0–25)	14326.95 (2735.36)	14636.88 (11060.81)
Group2 (25-50)	17355.17 (348.35)	17849.12 (14340.34)
Group3 (50-75)	18197.30 (209.63)	15291.8 (15355.90)
Group4(75-100)	19298.73 (688.73)	23692.56 (18897.87)

Table: 3 Mean, SD of CARS & Leg score in four percentile group.

Categorization	CARS Scores	Leg Scores
	Mean (SD)	Mean (SD)
Group1 (0–25)	14326.95 (2735.36)	17201.44 (24742.81)
Group2 (25-50)	17355.17 (348.35)	19969.12 (18428.33)
Group3 (50-75)	18197.30 (209.63)	23904.92 (24885.32)
Group4 (75-100)	19298.73 (688.73)	55039.32 (34970.84)

Table 4. *Correlation of Activity counts with CARS for leg, arm and total activity.*

Activity Scores	Coefficient (Beta)	R^2	P Value
CARS vs (Arm)	0.18	0.03	0.07
CARS vs (Leg)	0.30	0.09	0.003
CARS vs (Total)	0.31	0.10	0.001

Discussion

Our result indicated that the motion sensor counts determined by the Actiwatch are correlated with direct observation of activity as assessed by CARS score in preschool children. These results agree with Finn and Specker [3] who validated Actiwtch (Mini Mitter co. In) in 40 children 3 to 4 years of age against 6 hours of direct observation using the CARS. Further when we categorized the children into 4 quartiles based on CARS score a significant rising trend of activity mean scores with actiwatch tied on either on arm or leg was observed. Our data indicate that actiwatch tied on arm or leg does recorded activity validly however combined score of the two is better correlated with Observed activity.

The study was approved by the human research review committees at the Center for Micronutrient Research, Annamalai University, and Society for Essential Health action and Training (a non government organization in Delhi, India), and the Johns Hopkins Bloomberg School of Public health, and the WHO.

References:

1. Puyau MR, Adolph AL, Vohra FA, Butte NC (2001). Validation and calibration of physical activity monitors in children. *Obesity Research Journal* **10**: 150-157.
2. Phul JS, Greaves K, Baranwski T, Gruben D, Sealle D. (1990). Descriptions and calibration of a children's activity ratng scale (CARS). *Res Q Exerc Sport.* **61**: 459-477.
3. Finn, K.J and B Specker. (2000). Comparison of Actiwatch activity onitor and Children's Activity rating Scale in children, *Med Sci. Sports Exerc.* **32(10)**: 1794-97.

Application testing of a new three-dimensional acceleration measuring system with wireless data transfer (WAS) for behavior analysis

K. M. Scheibe, C. Gromann
Institute for Zoo- and Wildlife Research Berlin (IZW), Germany,

Abstract
The WAS (greenway systeme) is a three-dimensional acceleration measurement system with wireless data transfer and additional on board data storage capacity. It was tested on free moving cows and horses as a collar and as a flat box for measuring leg or trunk movements. By means of the collar sensor, basic behavior patterns like standing; grazing, walking, ruminating, drinking, and hay uptake could be identified in cows. Lameness could be detected in cows and horses by means of the leg sensor. The system can be used for the detection and analysis of even small movements of free moving humans or animals over several hours. It is convenient to analyze also other movement characteristics like emotional reactions, events releasing flight or frightening, or to compare different housing elements like floors or fences.

Keywords
acceleration measurement, radio telemetry, behavior, lameness

1 Introduction
The movements of points at an animals or humans trunk or at single extremities are often recorded by optical methods [e.g. 1, 7]. This requires a fixed measuring parlor and the tree-dimensional record is possible only under certain conditions. A tree-dimensional acceleration measuring system for biological as well as for technical purposes is the MT9 system [9], but this system requires a wire connection to a fixed recording station. It has been used for motion analysis for example in humans [4, 2]. Systems measuring accelerations in only two axes are not convenient for a complete description of a movement, since the resulting force cannot be computed. Furthermore, the application on free moving subjects such as unrestrained humans, domestic, or even wild animals requires a wireless data recording system. Such a system could be used for automatic behavior identification or for the identification of disturbed movement patterns resulting from diseases or environmental disturbances. For this purpose we tested a wireless three - dimensional acceleration measuring system (WAS) [8] and different analytical procedures.

2 Material and Methods
The WAS is capable to record accelerations of up to +-2g linearly, accelerations between 2g and 4g with a damped characteristic. A version for +- 10g is also available. Recording intervals of 10ms, 100ms or 1000ms can be selected. Results can be transmitted simultaneously on a laptop over a distance of about 50 to 200m or stored in the 8MB internal memory. Data are transmitted by a wireless data link, the receiving station connected to a PC. In the case of radio transmission, results are displayed on a graph simultaneously. Data are stored in EXCEL - compatible text files containing the time and the tree acceleration values for each time point. Two sensors were available, one as a collar, and another as a flat box for fixation at legs or at the trunk. The sample rate in all reported experiments was 10ms. For each sample point, the resulting force was determined from the tree-dimensional accelerations.
The collar sensor was applied on cows on free range for the identification of basic behaviors. The leg sensor was applied in cows and horses guided on concrete for lameness detection, using two lame cows and a horse. These animals were compared to healthy individuals under identical conditions (two cows, two horses).
From the recorded data files, series of 1000 values were selected for behavior identification, resp. series of 500 values for lameness detection. As analytical procedures, frequency distribution, fractal dimension, and spectral analysis were applied. From the spectral analysis, harmonic and non-harmonic components to the step frequency were separated and the power compared.

3 Results

3.1 Function of the system
The application of the system was safe and simple. The basic functions like sampling interval could be selected in an initial menu. Data transmission was possible over up to 200m in open terrain, but at least over 50m even with obstacles or in stables. The data storage function worked reliable and enabled simultaneous recordings from several sensors at the same animal. The collar sensor could easily fixed; the leg sensors were fixed easily and safe by means of usual horse boots. For the collar sensor, the 2-g range was sufficient in most conditions; in case of the leg sensor, results lay sometimes beyond 2 g and in extreme situations even beyond 4g. Measurements were possible over periods of several hours without battery exchange.

3.2 Behavior detection in cows
Even by simple visual inspection of the original curves of accelerations of the collar in the tree axis, basic behaviors like standing; grazing, walking, and rumination could be identified. For this purpose, the gravitation must be regarded to determine the position of the head. The head position together with the frequency pattern enables a differentiation of grazing against rumination and walking (Figure 1). The frequency distributions of the three single accelerations and the resulting force differed for these behavior patterns. The standard deviation was also useful to distinguish between these behaviors. The standard deviation was highest in grazing, lower in walking and jet lower in ruminating. Power spectral analysis of the resulting force could also be applied for behavior

identification. Behaviors like walking, grazing and rumination were characterized by regular head movements with different periods. But not only the basic frequencies differed between the behaviors walking, grazing, and ruminating, but also the other components of the spectrum. The basic period was very clearly expressed in walking, related to the other components. In grazing two periods were expressed, and in ruminating a broader spectrum was found.

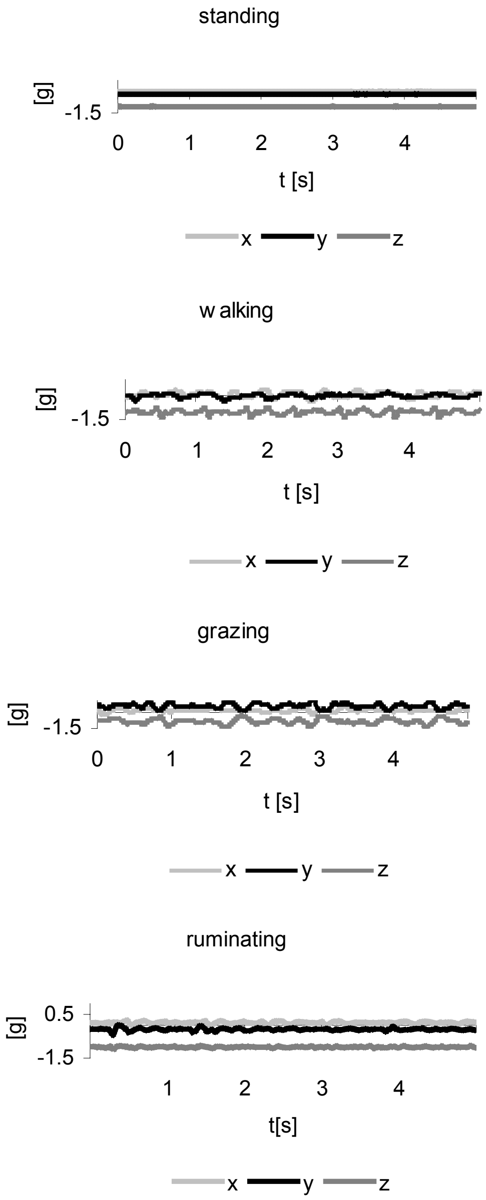


Figure 1. *Original acceleration patterns from a collar on a cow during different behaviors.*

The fractal dimension of two-dimensional acceleration patterns was also compared. It differed for such similar behaviors like hay uptake, grazing, and drinking (Figure 2).

3.3 Lameness detection

The original time series of accelerations looked fairly similar between lame and healthy cows and horses. Differences became obvious in the power spectra of the resulting forces, especially if the basic frequencies (representing the step frequency) were expressed in relation to the harmonic and non-harmonic components. In lame animals, the power of the basic frequency was less expressed, but the harmonic and especially non-harmonic components were higher than in healthy individuals. This was true for both species. In the case of horses, an animal with very slightly expressed lameness could be tested. Even in this individual, these components differed clearly form the control animals and the difference increased after a provocation test (Figure 3).

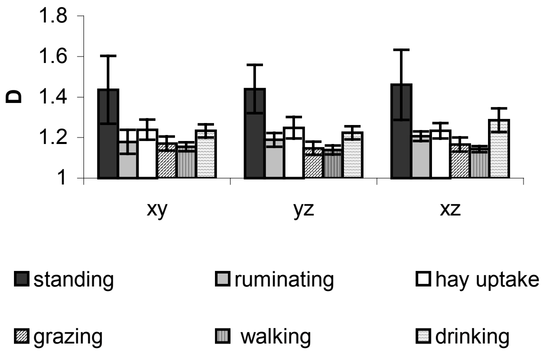


Figure 2. *Fractal dimension of two-dimensional acceleration patterns in the three spatial planes during different behaviors.*

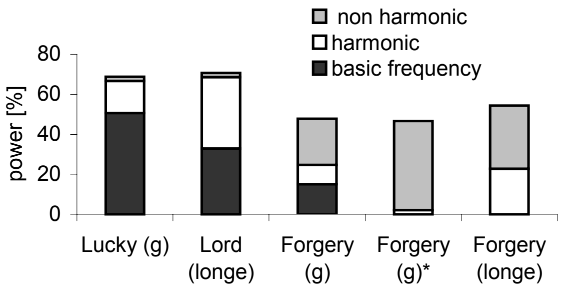


Figure 3. *Classification of power from spectral analysis of the resulting force from a leg sensor in horses walking guided (g) and running on longe line (longe). Lucky and Lord: healthy individuals, Forgery with a low degree of lameness, * after a provocation test.*

The fractal dimension of the accelerations was found to be lower during walking in lame individuals compared to healthy animals, but not in the horse during gallop.

4 Discussion

The acceleration measurement system was simple to apply and the measurement range of up to +-4g sufficient in most cases. For intensively moving animals or expected high accelerations, the 10-g version may be recommended. A temporal resolution of 10ms seemed sufficient for all the recorded behaviors and movements. The data storage capacity of the instruments offers the possibility to take simultaneous records from several points at one and the same individual, but a radio transmission gives the advantage of a direct record on the screen of a laptop or PC.

The importance of the system consists of wireless recording and transmitting accelerations in all tree axis from which the resulting forces can be computed for each time point. Some other systems for lameness diagnosis only transmit accelerations in two axes [3]. The tree-axial recordings from a collar can be used for automatic identification of behavior patterns, but several procedures should be combined. By means of procedures like position detection, spectral analysis and fractal dimension, most basic behaviors in grazing animals could be identified. The investigation of lameness is either conducted on fixed measuring places like treadmills [6] or only visually by veterinarians in practice. The WAS offers a simple opportunity to record and analyze movements under field conditions. Nevertheless, a certain standardization of the measurement conditions will be needed, such like the floor condition for lameness diagnosis. The applied procedures, especially the power spectral analysis of the resulting forces and subsequent frequency classification give examples for simple interpretation methods for records taken under such conditions. By means of methods like power spectral analysis and fractal analysis, values for a complex description of movement characteristics can be obtained. Fractal analysis was already successfully applied to gait analysis in humans [5]. But the possibilities of interpretation are not limited to these procedures, since a complete acceleration record is available with a high temporal resolution.

As the system can be easily applied and is not restricted to special environments, other applications than the described may be possible. We assume, that the system can be used for the investigation of emotional reactions as well as for evaluation of elements of keeping systems for domestic animals, for example different floors in stables. It can be also recommended for comparing the walking capacity of individuals from species like pigs or turkeys for selection purposes.

Acknowledgement

The investigation was supported by the Federal Ministry for Industry and Technology of Germany (program PRO INNO, grant KF 0331701KRF1). We acknowledge Dr. Lebelt, Veterinary Clinic Havelland, the cattle clinic of the Free University Berlin, Mrs. A. Wessel, Trappenfelde and Mr. A. Broja, Liebenthal for the experimental animals and kindly supporting the investigations.

References

1. Falaturi, P.H. (2001). Computerkinematographie (CKG) als geeignetes Verfahren zur objektiven Bewegungsanalyse. Beschreibung und Ergebnisse. Pferdeheilkunde 17, 30-41
2. Keshner, E.A. (2004). Head-trunk coordination in elderly subjects during linear anterior-posterior translations. Experimental brain research 158, 213-222
3. Lindner, A. (Ed.) (1999). Beiträge aus Orthopädie, Ganganalytik und Doping. Lensing, Dortmund
4. Mathie, M.J.; Coster, A.C.F.; Lovell, N.H.; Celler, B.G. (2004): Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiological measurement 25, R1-R20
5. Ripoli, A.; Palagi, G.; Bedini, R.; Rizzo, V.; Kraft, G.; Franchi, D. (1995). A fractal analysis approach to obtain effective algorithms usable in biotelemetry on-line signal analysis. Biotelemetry XIII. March 26-31. Williamsburg, Virginia, 407-413.
6. Schobesberger, H.; Peham, Ch. (2002). Computerized detection of supporting forelimb lameness in the horse using an artificial neural network. Veterinary Journal 163, 77-84
7. Szalay, F.; Back, W.; Barneveld, A.; Schamhardt, H.; Hajos, F. (2002). Comparison of VHS video recording system with apple macintosh-based image analysis and modified CODA-3 systems in equine motion analysis. Acta Veterinaria Hungaria 50, 167-176.
8. Greenway systeme (2004). WAS - Wireless accelometry system. Technical specification.
9. Xsens Technologies B.V. (2003). MT9 Real-time 3D motion tracker. Technical data.

What is invisible? Electrical monitoring of penetration activities of insect mouthparts

W.F. Tjallingii
Laboratory of Entomology, Wageningen University, Wageningen, The Netherlands

Abstract
In this paper it is explained how the electrical penetration graph (EPG) technique visualizes the activities and tissue positions of aphid stylets (piercing mouthparts) in plants. Insect and plant are made part of an electrical circuit and signals can be monitored and recorded as soon as the stylets start penetrating. Eight waveforms have been distinguished and correlated with activities and tissue positions so far, combining a variety of techniques with EPG recording. Mayor correlations are presented. Here we explore concurrent digital recording of video images and EPG signals as a new tool for waveform correlations.

Keywords
Feeding behavior, Aphid, stylets, plant penetration, phloem.

1 Introduction
Mouthparts of aphids and other piercing-sucking insects penetrate their host and are often involved in a number of complex activities. These activities are difficult to observe since the mouthparts are embedded in a mostly non-transparent tissue of the host. In order to observe and study such activities the electrical penetration graph (EPG) technique, which is briefly introduced here. Though developed for aphids, it has been applied now for several other insects as well.

1.1 The EPG System
In the EPG set up the plant and aphid are incorporated in an electrical circuit (Fig. 1) by inserting one electrode into the soil of a potted plant and attaching another electrode – a thin (20 µm diam.) gold wire – with conductive silver glue to the aphid dorsum [1]. Once the aphid stylets (mouthparts) penetrate the plant – referred to as 'probing' – an electrical contact is established, since the stylet's capillary canals are fluid (electrolyte) filled. Thus completing the circuit the electrical properties and changes in the stylet canals are transduced in an electrical signal. However, this only works well when there exist no other short-circuiting electrical contacts. In aphids and in a number of other insects with piecing mouthparts no other electrical contacts are made by the legs as their chitin is a good insulator and the plant surface is covered by non-conductive cuticular wax. The insect/plant represents also two important metal/electrolyte interfaces generating a certain voltage (E in Fig.1). The circuit is provided with an adjustable voltage source (p) to control the total (system) voltage (V).

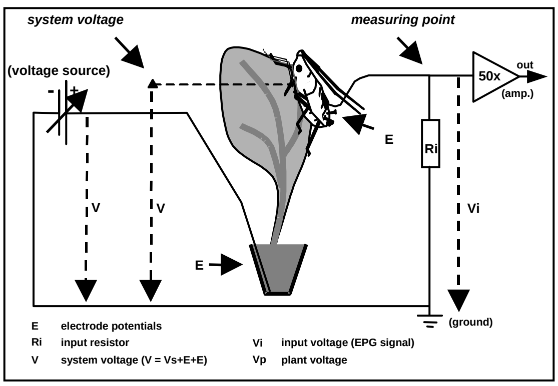


Figure 1. *Set up of the EPG recording circuit. The output (amp) is connected to a recording device, mostly a computer with screen display and hard disk data acquisition.*

1.2 Origins of EPG signals
In principle, the electrical resistance (conductivity) changes of the insect will cause a fluctuating voltage across the input resistor at the measuring point resulting from voltage division by the aphid-plant resistance and the input resistor (Fig. 1, Ri). The biological background of such resistance changes can be due to, for example, opening and closing valves in two capillary stylet canals, the food and salivary canal. These valves are associated with food and salivary pump in the aphid, respectively. This is the R-component of the EPG signal, which is measured by the original 'feeding monitor' – and more recent equivalent devices – that was introduced by McLean & Kinsey [2], the AC system. In contrast, the DC-EPG system we discuss here, records more than the conductivity fluctuations only. It also records bio-potentials, voltages at the measuring point that are generated inside the insect-plant combination, i.e. the electromotive forces (emf). The voltages of the emf-component in the EPG signal are caused; for example, by the membrane potentials of punctured plant cells and also, by streaming potentials caused by fluid movements in the stylet's capillary canals. Thus as the DC-system records both R- and emf-components it is sensitive for a more complete range of biological signals than the AC-system. Nevertheless, it can be useful sometimes to record both DC- and AC-system signals simultaneously [3]. A more complete comparison of the two systems is given elsewhere [4]. Although the membrane potentials of plant cells punctured are clearly reflected in the EPG (waveforms pd, E1 and E2; Fig.2), no biopotentials from the insect are recorded. The reason is that the plant cell is incorporated in the electrical circuit when punctured whereas the electrically active elements (muscles, nerves, gut potentials) inside the insects always remain outside the circuit. Their biopotentials will be short-circuited by the heamolymph in which they are bathing. The emf-components related to insect activity are indirectly related to muscle activity, i.e. they are caused by the streaming potentials in the stylet canals resulting from muscle contraction, not by muscle potentials themselves. The rather low waveform frequencies in the EPG (up to 20 Hz) as compared to muscle potentials (up to 75 Hz) appear to support this.

1.3 Behavioral impact of EPG recording
The feeding behavior of aphids is hardly affected by the system. Especially, the electrical current does not stimulate any muscles or nerves and the effect on particle movements has been found negligible. Aphids survived 10 days continuous EPG recording while constantly phloem sap feeding and giving birth to numerous offspring. The most serious effect is actually caused by tethering, which limits the area of moving and avoids an escape from a

plant, which is relevant in case of poor plant conditions or resistant plants [5].

1.4 EPG Waveforms

The EPG typically shows a baseline (zero volt) when the aphid does not penetrate its stylets (np). By convention the plant potential is adjust to get a maximally positive signal at the start of a probe (a period of stylet penetration). The system voltage (Fig. 1) is positive then and due to the fact that the early signal is dominated by increased conductance the first waveforms. A one-hour example of the EPG signal is given in Fig. 2.

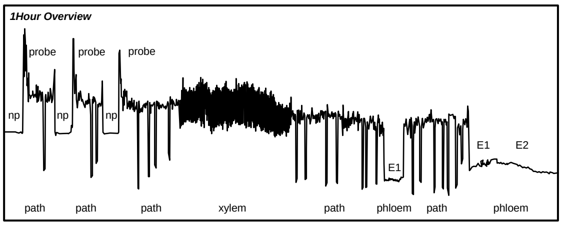


Figure2. *Electrical penetration graph of aphid feeding behavior. One-hour example of probing and non-probing alternation. A probe is a stylet penetration period, np is non-probing. First two probes contain pathway phase (path) only, the third probe also includes a xylem phase and two phloem phases. Many potential drops occur during path periods. The first of the two phloem phases is short with E1 only, the second E1 is followed by E2.*

Probing, i.e. periods of stylet penetration, alternates with periods of non-probing (np), which normally implies walking and selecting a new probing site in free non-wired aphids. Aphids feed on phloem sap, which is located in the sieve elements, deep inside the plant. First the stylets have to reach that tissue and the early waveforms in probes have been related mainly to stylet pathway (path) activities, indeed. Within probing, two more behavioral phases can be distinguished, each related to a ´target tissue´ – xylem and phloem – without further stylet advancement. Pathway phase, includes waveforms A, B, C, pd, and F (in Table 1 but not Fig. 2), xylem phase has only one waveform, G, and phloem phase includes two waveforms, E1 and E2.

1.5 Correlation studies

After characterizing EPG signals of aphids in terms of amplitude, frequency, voltage level and electrical origin, eight distinct waveforms have been distinguished so far. Many techniques have been used in combination with EPG recording in order to correlate the waveforms with activities of the insect and stylet tip positions in the plant tissues. The main waveforms with their experimentally correlated positions and activities are listed in Table 1.

Table 1. *Main correlations of aphid EPG waveforms with stylet tip positions in plants and aphid activity.*

waveform	*position*	*activity*
A, B, C, (F)	**path**	**intercellular penetration**
pd	**path**	**intracellular puncture**
G	**xylem**	**sap ingestion, active**
E1	**phloem**	**saliva secretion**
E2	**phloem**	**sap ingestion, passive**

Techniques used in these correlation studies were, feeding experiments with Parafilm® covered fluids labeled with radio-isotope [6] and myogram recording [6]. Stylectomy (fast stylet amputation) was executed during different waveforms and exuding sap appearance from the stylet stump was observed (and collected) or plant tissue was dissected and processed for sectioning and subsequent transmission electron microscopy [7,8]. Another powerful method appeared the use of aphids that had acquired plant viruses, the inoculation of which forms a self-replicating marker, indicating saliva secretion in cells of different tissues [9,10]. Honeydew excretion was another simple and good indicator for feeding, using a rotating drum to collect the droplets (honeydew clock).

1.6 Present ideas on plant penetration by aphids as derived from EPG recording

On basis of EPGs a number of new aspects of plant penetration appeared and several other aspects needed adjustments. Our present idea is that the stylet are inserted into the cell walls between two adjacent epidermal cells. They appear to proceed intercellularly in the secondary cell material between cellulose and hemicellulose fibers. Sometimes, they pass trough an intercellular air space or inside a cell between the cell wall and the plasmalemma (intramural-extracellular) [7,8]. During pathway phase the stylets are protruded, forcing a space between fibers and retracting briefly to fill this space with gelling saliva (waveform B). This gel forms a lining, salivary sheath surrounding the stylets in the tissues. Thus pathway phase represents a continuous cyclic activity of alternating mechanical protruding and sheath secretion. This activity is frequently interrupted by brief intracellular punctures (lasting typically 5-8 s) into most cells along the stylet track.. The punctures are reflected in the potential drop (pd) waveform. During a pd puncture some watery saliva is injected into the cell and subsequently a small amount of sap is sampled. It remains unclear what function the saliva has for the aphid or what its impact is on plants but apparently non-persistent virus particles – if present – are release from the stylets and flushed with the saliva into intact plant cell. Sap sampling presumably has a function in host plant selection. Xylem ingestion maximally lasts for one hour and mainly seems to serve compensation of water losses, which is important since aphids are confronted with the high osmotic value (mainly sugar) of the sap during phloem feeding. Phloem sieve elements are living cells that require delicate handling in order to prevent sieve tube sealing. Presumably this is the main reason why watery salivation into the sieve element (waveform E1) always occurs first. E1 salivation allows persistent viruses (different from non-persistent ones) to enter the intact phloem cell. Saliva injection is only possible if the throat valve is closed, preventing an immediate inflow of phloem sap due to its very high hydrostatic pressure. Later, when waveform E2 starts, the valve opens and then the plant actually feeds the insect passively. During E2 valve movements, presumably controlling the sap inflow occur concurrently with continued watery salivation but this saliva is immediately mixed with the incoming sap does not reach the plant since the outlet of salivary canal is located in the food canal just before (downstream) the stylet tip. Possibly, the watery saliva in E1 and E2 prevents clogging (coagulation) of some phloem proteins that protect plants against ´bleeding´ [11].

1.7 EPG application

Most information derived from the EPG technique has directly contributed to our fundamental knowledge of the process of stylet penetration itself and understanding the functional features of stylet morphology [12]. Directly linked to the use of plant viruses as markers in EPG correlation studies is the spin off to a better understanding the mechanisms of virus transmission and the damage avoiding mode of stylet penetration [8,10,12]. As natural counterparts of entomologists, phloem physiologists became interested in EPG work and aphid-plant interactions especially, in the mechanism of avoidance or suppression of phloem wound responses by aphids. Salivary components can be distributed easily once excreted into the sieve tubes and appear to induce both resistance and susceptibility to aphids. Most work is very recent and has not been published so far.

From the beginning, a main objective was applying EPGs in studying mechanisms of host plant resistance and this is a fast growing field. From EPG information the location of resistance factors in specific plant tissues can be derived. Even susceptible plants can have (partial) resistance factors and though giving only a temporary protection combining genes from these plants with genes of other resistance factors (pyramiding) can be rather valuable. In the development of new insecticides, both synthetic and natural products, EPGs are used to study behavioral effects.

2 Video revisited.

2.1 Old attempts

Visual observation of stylet penetration in plants is obscured by the non-transparency of the tissues. Several old live visual observations by light microscopy used Parafilm® covered fluids to observe the penetrating stylets. Another visible aspect during stylet penetration is the contraction of muscles in the head of some aphids that have a semi-transparent cuticle. Concurrent EPG recording made a concurrent playback desirable in order to correlate visual and physiological information. Split-screen video with one camera on the microscope and the other on the EPG displaying chart recorder or oscilloscope was a feasible combination but zooming in into EPG details was difficult. Using time markers on independently taped video and EPG traces gave synchronization problems due to tape stretch differences. Another combination used was putting an amplitude modulated EPG signal (AM-EPG) on the audio trace of the video tape in addition to separate EPG recording on PC disk. Direct digitalization of video recording has now opened new prospectives to record physiological and video signals and to analyze them in a synchronized way off line. Our aim is to investigate these new possibilities using The Observer-XT software (Noldus Information Technology) and to compare them with earlier results.

2.2 Materials and methods

During concurrent video and EPG recording aphids were allowed to penetrate plants or artificial diets. Video recording was mediated by The Observer-XT whereas EPGs were acquired by Probe 3.0 software (Wageningen University: WU). EPGs were obtained by a multi channel DC-EPG system (Giga-4, WU) using a separate computerwhile adding to a free channel a random time code signal from COM port of the Observer-XT running computer. Later the EPG file with the time code was imported into the Observer-XT in analysis mode.

2.3 Results

Video-EPG combinations obtained by the old video & AM-EPG technique are compared with the new system.

3 Conclusions

EPG recording has demonstrated to provide reliable information on the invisible activities during plant penetration by insects with piercing mouthparts. This information has contributed and adjusted ideas of stylet penetration in plant tissue considerably. Nevertheless, many questions are left about the nature and function of aphid probing activities and their impact on plants. Further research using combined video and EPG recording appears a promising approach to tackle such questions by contributing to EPG waveform correlations.

Thanks to Noldus Information Technology for allowing the use of Observer-XT and especially, for technical help and ideas during the experiments.

References:

1. Tjallingii, W.F. (1988). Electrical recording of stylet penetration activities. In *"Aphids, Their Biology, Natural Enemies and Control", A.K. Minks and P. Harrewijn (Eds.), Vol.2A.* pp.95-108. Elsevier, Amsterdam.
2. McLean, D.L.; Kinsey, M.G. (1965). Identification of electrically recorded curve patterns associated with aphid salivation and ingestion. *Nature* (London), **205**, 1130-1131.
3. Kindt, F.; Joosten, N.N.; Tjallingii, W.F. (2005) Electrical penetration graphs of feeding behaviour in thrips: concurrent recording by DC- and AC-systems. *Journal of Insect Physiology* (accepted)
4. Tjallingii, W.F. 2000. Comparison of AC- and DC-systems for electronic monitoring of stylet penetration activities by homopterans. In *"Principles and Applications of Electronic Monitoring and Other Techniques in the Study of Homopteran Feeding Behavior", Walker, G P.;. Backus, E.A.. (Eds.).* Thomas Say Publications in Entomology, Entomological Society of America, Lanham, pp. 41-69.
5. Tjallingii, W.F. (1986). Wire effects on aphids during electrical recording of stylet penetration. *Entomologia Experimentalis et Applicata,* **40**, 89-98.
6. Tjallingii, W.F. (1978). Electronic recording of penetration behavior by aphids. *Entomologia Experimentalis et Applicata,* **24**, 521-530.
7. Kimmins, F.M.; Tjallingii, W.F. (1986). Ultrastructure of sieve element penetration by aphid stylets during electrical recording. *Entomologia Experimentalis et Applicata,* **39**, 135-141
8. Tjallingii, W.F.; Hogen Esch, Th. (1993). Fine structure of aphid stylet routes in plant tissue in correlation with EPG signals. *Physiological Entomology,* **18**, 317-328.
9. Powell, G. (1991). Cell membrane punctures during epidermal penetrations by aphids: consequences for the transmission of two potyviruses. *Annals of Applied Biology*, **119**, 313-321.
10. Martín, B.; Collar, J.L.; Tjallingii, W.F.; Fereres, A. (1997). Intracellular ingestion and salivation by aphids may cause acquisition and inoculation of non-persistently transmitted plant viruses. *Journal of General Virology*, **78**, 2701-2705.
11. Knoblauch, M.; Van Bel A.J.E. (1998). Sieve tubes in action. *The Plant Cell,* **10**, 35-50.
12. Prado, E.; Tjallingii, W.F. (1994). Aphid activities during sieve element punctures. *Entomologia Experimentalis et Applicata,* **42**, 157-165.

Investigations of daily biorhythm in different horse keeping systems for well-being measured with ALT pedometer

U. Brehme[1]; U. Stollberg[1]; Brigitte Strickler[2]; R. v. Niederhäuser[2]; H. Zurkinden[2]
[1] *Institute of Agricultural Engineering Bornim e. V. (ATB), Division Engineering for Livestock Management, Potsdam, Germany*
[2] *Haras National, stud-farm, Avenches, Switzerland*

Abstract
Electronic identification and measuring systems represent key technologies for progressive automation in animal husbandry in modern, future-oriented livestock farming. Modern sensors (sensors, bio sensors) and increasingly non-invasive measuring and transfer methods make crucial improvements in the potential for measuring data. A new type of pedometer, named ALT pedometer, was developed for using in cattle and horses. **ALT** is the synonymous for **a**ctivity, **l**ying time, and **t**emperature. The pedometer system register the three parameter activity, lying time, and temperature for an accurate determination of the daily biorhythm of horses. It was tested in stud-farms in Germany and Switzerland with good success. The main reason for these investigations was the question - is the present dimensions of single boxes, single boxes with run and pasture area sufficient and suitable for horse keeping? The investigations took place in Germany and in Switzerland in three experiments.

Keywords
Precision Livestock Farming, sensors, well-being

Introduction
The growing distribution of the sports and saddle-horse keeping in the complete EU area require a customization of the legal bases and framework conditions for the individual countries for the horse keeping under the aspects of well-being in horse keeping systems in stable plants and the keeping claims of horses on pasture keeping.
Well-founded statements concerning the daily biorhythm - activity behaviour, rest periods - must basis new, revised guidelines and regulations of the keeping order of horses scientifically for these animals.
For the on-stables and accommodation of the horses modern, suitable for manners keeping systems should find, the both the needs of the animals, the needs of man and have the protection of the environment in the look. Horse keeping suitable for manners is not only a demand of the protection of animals law but takes a high general request into account particularly at the aspects of the sustainability and the environmental compatibility. This has to be taken into account at all planning and measures for the making for new buildings or reconstruction plans. Behaviour problems of the horses which aggravate the contact or the use and can be explained by insufficiencies to a considerable part in the keeping as one knows today don't have to be also neglected, however. The most frequent form of the individual keeping is the single box but also outer boxes and outer boxes with run, a directly adjacent little pad dock, are used. The group housing for horses in great boxes (closed stable building without permanent entry to a run) and the loose barn belongs to the group of keeping systems. One understands an increase area system by a loose barn with a lying area in a stable building, an eating area into or outside the stable building and for a run area it is outside the stable building, permanently accessible to this one.
Newer investigations give tips to the fact that especially rank-low animals show a decreased rest behaviour in the lying time in the group, because they are disturbed over and over again by more high-ranking during her quiescence and sleeping phases [6,7]. Sleep is an essential need, and deep sleep is of vital importance for all mammals. This is possible with horses only in the lying in the belly position and side position as on the basis of EEG and muscle tone measurements could be ascertained [9,14]. Adult horses rest about seven hours during the day from what they spend approx. 20 % in the lying. Although horses rest, indeed, predominantly in the standing position (doze), must be still guaranteed also in the group keeping that an enough largely limited, as well as dry and malleable lying surface is available, so that all horses are at a distance unhindered at the same time and get up as well as can sleep in side position. Pilot studies show that structuralizations of the lying areas can affect in loose barns positively the frequency of lying periods and duration, especially rank-low horses.
New development of a pedometer system, called ALT pedometer, to register activity, lying time, and temperature for an accurate determination of the daily biorhythm for horses was tested in stud-farms in Germany and Switzerland.
The main reason for these investigations was the question - is the present dimensions of single boxes, single boxes with run and on the pasture sufficient and particularly for horse keeping.

Material and Methods
It is main concern of the research project on hand to provide the legislation in the area of the protection of animals and the appropriate to the species keeping of horses, by a scientific analysis of the keeping systems used in Germany to provide a sound data material to the movement activity and to the lying phases or quiescence of horses in the daily biorhythm. In addition, broader knowledge shall be collected by horses to the optimization of the group keeping so that the wording is made possible more clearly for guidelines and recommendations also in this area. In addition, nearly optimum keeping terms favour well-balanced and affable horses and contribute with it to an accident prevention not to be neglected (holder, owner, ready, orderly, blacksmith, veterinarian among other things). For an animal which is daily on an average just 19 hours in movement under natural conditions an everyday movement duration of less than one hour shows a considerable restriction of appropriate to the species behaviour patterns. Different investigations have shown that lack of exercise in the horse can cause a whole row of illness and with it connected suffering and pains and is often early departure cause for horses [7,10,11,13].
This is also reflected in insurance statistics again in which illnesses of the movement apparatus with an interest of more than 50 % are listed as a main departure cause for riding horses [6,9]. By lacking movement offer the tendons, tapes and joints lose her elasticity and strong charges as they appear today in the horse-riding, affect quickly damaging. Only with the neutral horses can do her well-being motion sequences as the preferential slow walking in the step, but also humps and plays freely unfold. Investigations show that only the loose barn can cover with separate functional areas and the all-day pasture way the movement need [1,2,3,8,11]. Hence, the veterinary union

also recommends for protection of animals that in all single keeping systems an everyday several hours of movement possibility is to be offered to the horses to the balance for the activity loss. And the necessary movement is to be guaranteed by free running, if necessary complements with work or training.

The mode of operation of pedometer for the measuring parameter "animal activity" works according to the principle of the impulse counting. To this gets a horse at an arbitrary foot, as advantageous the forefeet has proved itself, for safety (hits out) an ALT pedometer laid out with a ribbon or in a gaiter. ALT stands for activity, lying time and temperature at it as synonymous. The decisive advantages of this type of pedometer lie in the following features:

- Measuring of three animal individual parameters (activity, laying time, surroundings temperature at the pedometer), instead of one feature only (activity) in the conventional pedoemter;
- eligible time interval for the recording of all parameters in the measurement range between 1 and 60 min;
- the defined assignment of all data sets to the day of the biorhythm permits the real time clock in the ALT pedometer;
- continuous measuring data acquisition, data storage and manual or cyclical automatic data transmission over arbitrary time periods after predefined measuring interval by means of radio modem to the PC;
- the high correspondence between activity and lying time permit secure statements for daily biorhythm of horses.

The pedometer contain four sensors to the recording of the step activity, the lying time in two different lying positions and the surrounding temperature. The results from temperature relations permit conclusions in the lying area of the animal and with it on the well-being. A μ processor, the data memory, the real-time clock, a lithium battery and the radio module to the wireless data transfer complete ALT pedometer. The μ processor grasps temperature, step activity and lying positions of the animal continuously and sums these up about the measuring interval configured at attempt beginning (1 to 60 min). The sum of the step activity, the lying times and the surrounding temperatures unite a data set. The storage capacity of an ALT pedometer amounts to 740 data sets [4].

First investigation took place in Neustadt/Dosse (Germany) in single inside boxes only with stallions and mares over two weeks [4]. In Switzerland three investigations were curried out. First in Bellelay with stallions and geldings in single inside boxes and boxes with run over three months [5]. Second with stallions at the pasture take place in Avenches over four months in autumn and wintertime. Last investigation over two months is a test to the daily biorhythm with the same stallions after wintertime on pasture in inside single boxes in the stud-farm from Avenches.

Through continuous automatic recording with an eligible time interval (five minutes) automatic data transmission from pedometer to PC a successful and rapid control of the exact daily biorhythm with detailed results for activity and lying time per day is possible.

Results

The statistical evaluation of the data was carried out with EXCEL 97 and S-plus 6.1.

Differences in movement activity of the animals under the experimental conditions with and without became run was tested with the help of the Wilcoxon omen rank. This not parametric test was elect because:

- no normal distribution of the results had to be presupposed;
- every animal under two experimental conditions was watched and therefore paired samples were presented.

Per animal and experimental condition a correlation (Spearman correlation) became of the measured activity and this one by direct observation of the horses.

The graphic representation from the data used by EXCEL 97.

The movement activity of our using is an incorruptible indicator for animal welfare of the elective keeping system regarding the assessment of animal health, fertility, performance, well-being and balance.

One knows individual or single and group keeping systems for horses today. At the individual animal keeping the horses are substantially restricted in their social contact to other horses by the structural design of the boxes. In addition, the movement need and the reconnaissance behaviour of the horses are taken into account little in this keeping system. The most frequent form of the individual keeping is the single inside box but also outer boxes and outer boxes with run, a group keeping systems under each other in the same keeping unity. From investigations is known that approximately 84 % in single keeping and only about 16 % in group keeping are held by the whole horses continuance [2,3,7,12,15]. The horses in single keeping are moved on average five hours per week beyond the single box. This means that horses stay in single keeping systems approximately 23 hours daily in her single box [1,4,6]. For an animal which is daily on an average just 19 hours in movement under natural conditions an everyday movement duration of less than one hour shows a considerable restriction of appropriate to the species behaviour patterns. Different investigations have shown that lack of exercise in the horse can cause a whole row of illnesses and with it connected suffering and pains and is often the early departure cause for horses. In the examinations the trend could be confirmed obviously that, generally, horses showed a higher movement activity than in the inside box ($P = 0.0547$) in the individual box with run. All available minute values per keeping system and animal groups (stallions/geldings) was summed up and by the number of available minute values split. From this one of the central results of this work follows: Animals with run showed according to tendency a higher average movement activity than without run. With the stallions 17.95 impulses became with run in the average measured per minute, without run only 12.29 impulses. This one was at the geldings average movement activity 6.37 impulses per minute and without run 3.64 impulses. One compares the mean average values of the movement impulses of the animals in the variants or without run the stallions showed with run 33.82 % more movement impulses so that without run and the geldings 52.54 %. It is a strong trend act, however, that horses moved under condition with run more than in the variant without run [5].

Movement activities the calculations to the whole activity proved that the animals showed altogether one higher movement activity than without run in the variation with run. Because of big differences between stallions and geldings the animal groups were looked individually. Per animal group and keeping system were calculated from all available data on an interval of 20 minutes in each case of averages. The analysis pointed, that the movement activity of the stallions and geldings about distributed the whole day with run was higher than without run. With the stallions fell at them that only in the morning at start of work of the staff an increase this one activity was measured while the geldings weren't reacting to the presence of the staff not in the movement activity. No entry to the run had the horses from 5:00 p.m. hours till 9:00 a.m. hours. The results showed that all stallions and geldings behaved quietly at night. The low of the movement activity was reached at 2:00 a.m. hours. The results show a clear trend for the fact that the movement activity can be raised by horses by the offer of a run accessible during the day. It could be shown that horses often stayed at choice in the run. Clear differences in the behaviour between stallions

and geldings were established. Stallions carried out 3.36 more changes between box and run as geldings so, for example. The geldings stayed in the run during on an average 37.63 % of time, the stallions during 65 % of time. In even clearer values this trend at this is compare to recognize the keeping variants (Fig. 2) - single inside box without run and pasture keeping [5]. The comparison you see in Figure 2 between the test periods on the pasture and in single inside box from a stallion. The graphic makes clear that the daily biorhythm on the pasture is the most well-being system for horses. Horses move in the social federation to the feed admission up to 16 hours daily. Under keeping conditions is therefore an adequate compensation to make for the activity loss. With regard th the movement possibility but also the suggestion for the movement is keeping with run a judge keeping form into groups when the suitable for horses one. The individual keeping should always be a less-than-ideal solution and be used only there where a group keeping is not practicable. At examinations in increase room group run keeping put horses back between 3 km to 4.7 km per day. You ran through more than 80 changes between the box and the run at it.
Figure 1 shows the results from daily biorhythm in a single inside box over two week from a stallion, bed conditions for animal well-being.

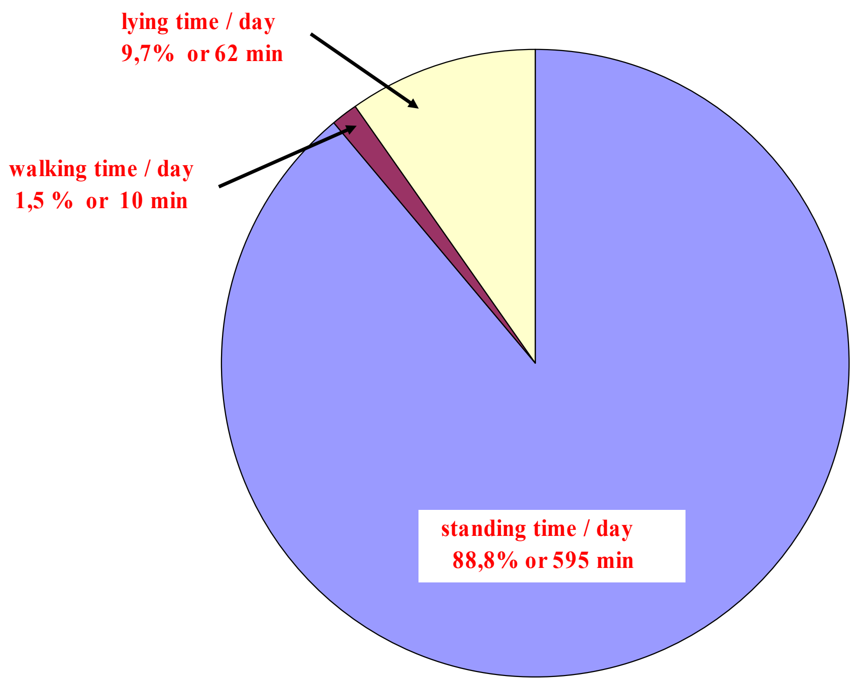


Figure 1. *Daily biorhythm from "Kitaro" - single inside box (measuring time: 00:00 to 24:00 o. c.)*

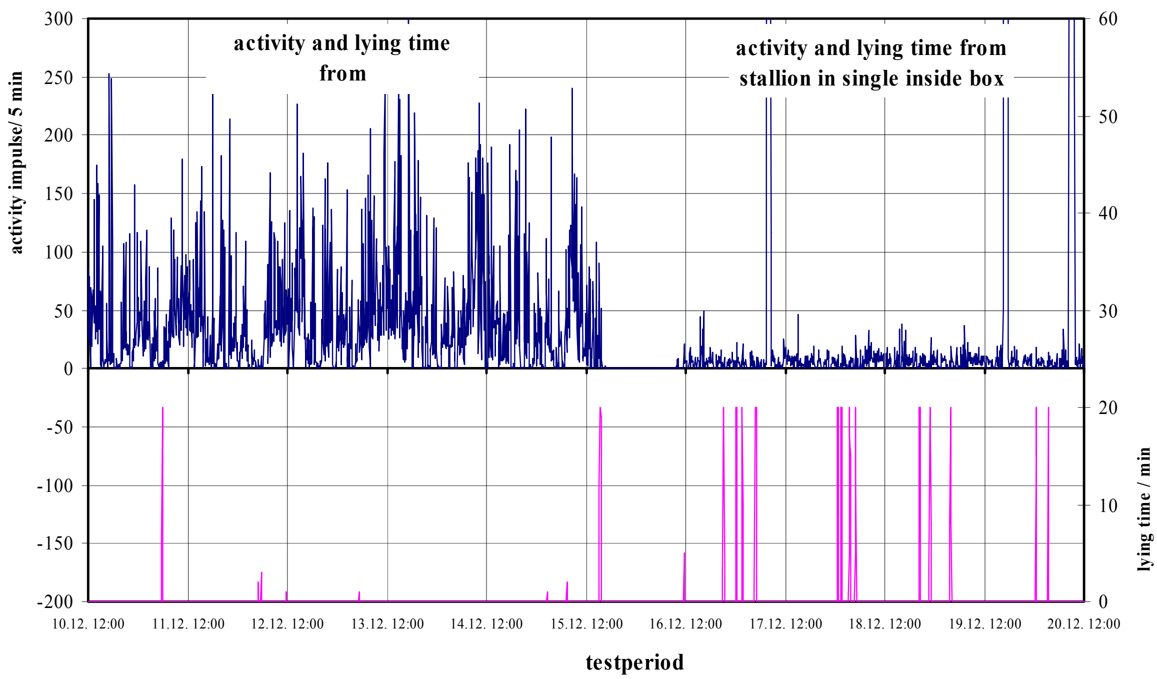


Figure 2. *Comparison between activity and lying time from stallion "Lorambo" (Parts of the test periods on the pasture and in single inside box)*

References

1. Bachmann, I. (2002b). Haltung und Nutzung von Pferden in der Schweiz: Eine repräsentative Erfassung des Status quo. Schweiz, *Arch. Tierheilkunde*, **144**, 331-347.
2. Bey, O. (1993). Nutzung des Pferdes aus der Sicht des Tierschutzes. *Diss. vet. med.*, Gießen.
3. Beyer, S. (1997). Konstruktion und Überprüfung eines Bewertungskonzeptes für pferdehaltende Betriebe unter dem Aspekt der Tiergerechtheit. *Diss. vet. med.*, Gießen.
4. Brehme, U.; Stollberg, U. (2004). Eignungstest von Fußpedometern für Untersuchungen bei Sportpferden. ATB Bornim.
5. Buschor, Y.; et. al (2004). Bewegungsaktivitäten von Pferden in Einzelboxen mit bzw. ohne Auslauf. Diplomarbeit ETH Zürich.
6. Dallaire, A.; Ruckebusch, Y. (1974). Sleep patterns in the Pony with observations in partial perceptual deprivation. *Physiol. Behav.*, **12**, 789-796.
7. Fader, C. (2002). Ausscheide- und Ruheverhalten von Pferden in Offenstall- und Boxenhaltung. Diss. agr., Weihenstephan.
8. Hommerich, G. (1995). Ausfallursachen und -frequenzen laut Schadenstatistik der Vereinigten Tierversicherung Gesellschaft AG. Gesellschaft Pferdetage, Institut für Tierzucht und Haustiergenetik, Göttingen.
9. Korries, O. C. (2003). Untersuchung pferdehaltender Betriebe in Niedersachsen - Bewertung unter dem Aspekt der Tiergerechtheit bei Trennung in verschiedene Nutzungsgruppen und Beachtung haltungsbedingter Schäden. *Diss. vet., med.*, Hannover.
10. Manton, A.; Pollmann, U.; Richter, Th. (2004). Eine Untersuchung des individuellen Ruheverhaltens von Pferden im Liegebereich eines Offenlaufstalls. 9. *Internationale Fachtagung zum ThemaTierschutz*, 19./20. Februar, Nürtingen.
11. Rodewald, A. (1989). Fehler bei der Haltung und Nutzung als Schadensursache bei Pferden in Reitbetrieben. *Diss. vet. med.*, München.
12. Ruckebusch, Y. (1972). The relevance of drowsiness in the circadian cycle of farm animals. *Anim. Behav.*, **20**, 637-643.
13. Seidensticker, C. (1999). Abgangsursachen entschädigter Pferde einer Tierversicherung aus den Jahren 1990 - 1995. *Diss. vet. med.*, Hannover.
14. Wackenhut, K. S. (1994). Untersuchungen zur Haltung von Hochleistungssportpferden unter Berücksichtigung der Richtlinien zur Beurteilung von Pferdehaltungen unter Tierschutzgesichtspunkten. *Diss. vet. med.*, München.
15. Wallin, L.; Standberg, E.; Philipsson, J.; Dalin, G. (1998). Estimates of longevity and causes of culling and death in Swedish warm-blood and cold-blood horses. Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, 750-757.
16. Zeeb, K. (1997). Anforderungen an die Pferdehaltung. In: *Das Buch vom Tierschutz*. Sambraus, H.H.; Steiger, A. (Eds.), Ferdinand Enke Verlag Stuttgart.

Enter the Matrix: How to analyze the structure of behavior

L. Hemerik[1], T. Bukovinszky[2], R. Gols[2], J.C. van Lenteren[2] and L.E.M. Vet[2,3]

[1]*Department of Mathematical and Statistical Methods, Biometris, Wageningen University, Wageningen, The Netherlands*

[2] *Department of Plant Sciences, Laboratory of Entomology, Wageningen University, Wageningen, The Netherlands*

[3]*Department of Multitrophic interactions, Netherlands Institute of Ecology, Heteren, The Netherlands*

Abstract

Several methods are available to analyze different aspects of behavioral transition matrices, but a comprehensive framework for their use is lacking. We analyzed parasitoid foraging behavior in environments with different plant species compositions.

The resulting complex data sets were analyzed using the following stepwise procedure. We detected abrupt changes in the behavioral records of parasitoids using a maximum likelihood method. This served as a criterion for splitting the behavioral records into two parts. For both parts Mantel's test was used to detect differences between first-order transition matrices, whereas an iterative proportional fitting method was used to find behavioral flows that deviated from random transitions. In addition, hidden repetitive sequences in the transition matrices were detected by using the Theme software.

We discuss the results from the example in a biological context and the comprehensive use of the different methods. We stress the importance of such a combined stepwise analysis to detect differences in some parts of behavioral sequences.

Keywords

Behavioral transition matrix, Iterative proportional fitting, MatMan, Theme

1 Introduction

Behavioral data show great variation at different levels. When having collected data on experimental animals in different environments, an important question is in what part of the displayed behavior are the differences most pronounced? In the literature, several methods are available to analyze and compare the structure of behavioral flows in complex data sets [1, 2, 3, 4]. Some methods use the behavioral records [1, 4] as the base of the analysis, whereas others use a summary of the behavioral records, i.e. the first-order behavioral transition matrices [2, 3]. In these matrices the behavioral records are summarized as the total frequencies at which a preceding behavioral element is followed by one of the other distinguished behavioral elements. All these analytical methods detect and compare different properties of the behavioral data sets.

Here, we approach the problem of analyzing behavioral records and the associated transition matrices by using and comparing the above-mentioned analytical tools. These tools are complementary and thus give answers to questions at different levels. The breakpoint analysis is used to detect one breakpoint, an abrupt change over time, in a behavioral record [1]. The correlation ***between*** behavioral transition matrices is investigated with MatMan version 1.1 [2, 5]. The proportional fitting method of Goodman [3] is used to detect differences in the suite of behaviors displayed by the animal in varying environments. This method is a ***within*** matrix method. The higher order transitions are studied with Theme version 5.0 [6].

In this study, we specifically assess the effect of different plant compositions of the environment on the behavior of a parasitoid that is used for biological control in crops. In the experiments the host containing or target plant was surrounded by plants of the same species, or by a combination of this plant species and another plant species, the so-called companion plant species. This companion plant species was either a close or distant relative of the target plant species. We hypothesized (1) that a breakpoint in behavior should coincide with the first landing on the target plant and (2) that the companion plant species has a different effect on the parasitoid's behavior until the first host is found.

2 Material and Methods

2.1 Experimental set-up

The details of the rearing and the experiment can be found in [7]. The basic set-up consisted of nine Brussels sprout plants (three rows of three plants, see circles in Fig. 1ab). The plant in the center is the target plant, with 15 host larvae of the species Plutella xylostella. This is the low-density monotypic set-up (A). In between these nine plants twelve Brussels sprout, barley or mustard plants were placed in two different configurations (mixed or row, see Fig. 1ab). These different situations will be called (B) the high-density monotypic (21 Brussels sprout plants), (C) the mixed mustard, (D) the row mustard, (E) the mixed barley and (F) the row barley set-up, respectively. At the start of an observation a female parasitoid of the species Diadegma semiclausum was released at a distance of 90 cm from the target plant. For each of the six defined experimental set-ups 25 trials were performed.

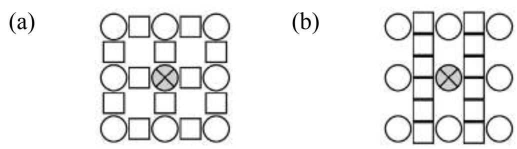


Figure 1. *The experimental set-up mixed (**a**) and row (**b**) configuration. The circles ○ represent Brussels sprout plants and the squares □ companion plants. The B. sprout plant ⊗ in the center is the target plant, i.e. the only plant with hosts.*

2.2 Recorded behavior

The behaviors distinguished during continuous observations were (1) walk – the parasitoid walks over the leaf surface; (2) preen – the parasitoid cleaned itself by grooming or scraping its body parts while stationary; (3) stand still – the parasitoid was motionless; (4) fly – the parasitoid was airborne; and (5) oviposit– the parasitoid inserted its ovipositor into the host. After the fifth oviposition the observation stopped. We used a handheld computer (Workabout Psion PLC 1995©) with The Observer 3.0 [8] to record the foraging behavior of wasps in the set-up, further data handling was carried out by The Observer 5.0 [9].

2.3 Detection of a breakpoint

Each behavioral record for the three different plant combinations was first analyzed with a test to detect one breakpoint [1]. This test is called the maximum likelihood estimate of a breakpoint in a transition matrix. Based on the behavioral record for each parasitoid, the following log-likelihood is defined

$$L = \sum_{i=1}^{5} \Big(\sum_{\substack{j=1; \\ j \neq i}}^{5} n_{ij} \log n_{ij} \Big) - \sum_{i=1}^{5} n_i \log n_i \qquad (1)$$

with n_i the number of occurrences of act i (i = 1, 2, 3, 4, 5) as defined above. The number of transitions from act i to act j is n_{ij}. The behavioral record with a total of N bouts is split in the first T bouts and the last N-T bouts. For both these parts of the record and the total behavioral record $L(T)$, $L(N\text{-}T)$ and $L(N)$ have been computed with formula (1). The breakpoint occurs at the value for which the likelihood ratio test statistic $\Lambda(T)$ is maximized.

$$\Lambda(T) = 2\big(L(T) + L(N - T) - L(N)\big) \qquad (2)$$

2.4 Analysis with MatMan

After having determined the breakpoint in the behavior we constructed the required number of different first-order behavioral transition matrices. For analysis, we assumed that behavioral sequences manifested by the individual parasitoids did not differ significantly within a given treatment and the data was pooled for all individuals in that treatment. The pooled first-order behavioral transition matrices were analyzed with MatMan version 1.1 after normalization of the data [10, 11]. The correlation between two matrices was calculated using Mantel's Z and Pearson's r tests. The significance of the correlation was based on 2000 permutations to asses how extreme the actual Mantel's Z value is relative to the Z-values generated under the null hypothesis of no association between the compared matrices [12].

2.5 Proportional fitting

At this stage we know which matrices are highly correlated, because we have performed an analysis for correlation between matrices. However, within a matrix the balance between the occurrences of behavioral elements can also differ. This kind of analysis asks for a within matrix approach.

In each transition matrix, the values on the diagonal are logical zeros, since a particular behavior cannot logically follow itself in such a record. The iterative proportional fitting method of Goodman [3] is used to find the expected values for the non-zero matrix elements. A log-likelihood ratio test (G-test) is used to evaluate their statistical significance for the overall matrix. Throughout the procedure, Yates's correction for continuity is applied. If the deviations in the overall matrix table are significant, then significant transitions can be found by collapsing the full matrix table into a 2 x 2 matrix around each transition. The significance of these individual tests is adjusted to a table wide level of 5% with the sequential Bonferroni method [13]. The goal of this analysis is to provide a clear picture of how the observed suite of behaviors differs when parasitoids forage in different environments.

This method identifies which transitions occur more or less frequent than by chance alone conditional on the observed column and row totals. To gain insight in how the behavior of searching parasitoids is organized, behavioral kinetograms [14, 15] are constructed (see Fig. 3). This allowed the visualization of differences in the behavioral sequences between parasitoids foraging in different environments.

2.6 Analysis with Theme

Thus far, only a breakpoint is detected in each behavioral record and first-order transitions from the original behavioral records are explored. The time-structure of the data has been ignored during the analysis of the first-order transition matrices. Within the behavioral records however, repeating structures and higher order transitions might occur frequently. Therefore, we used Theme version 5.0 [6] to detect so-called t-patterns [4]. Theme performs an intensive structural analysis of the behavioral data, and it not only considers the order of events, their relative and real timing, but also the hierarchical organization of the events. In Theme, a sequence of behavioral elements that is repeatedly occurring in a behavioral record is called a pattern. Among other variables Theme detects in a behavioral record (1) the number of different patterns, (2) the total number of pattern occurrences and (3) the mean number of behavioral elements in these patterns. We analyzed these results from Theme with non-parametric tests: the Kruskal-Wallis analysis of variance followed by pair-wise comparisons with the Mann-Whitney U test in SPSS 11.0. Because we performed multiple tests on the same data, the Bonferroni correction was used [16].

3 Results

3.1 Detection of a breakpoint

The maximum likelihood estimate of a breakpoint in all behavioral records is calculated and the number of the behavioral bouts that preceded the detected breakpoint is denoted. All these bout numbers are plotted against the bout numbers of the first landing on the target plant for the set-ups **(B)**, **(C)**, **(D)**, **(E)** and **(F)**.

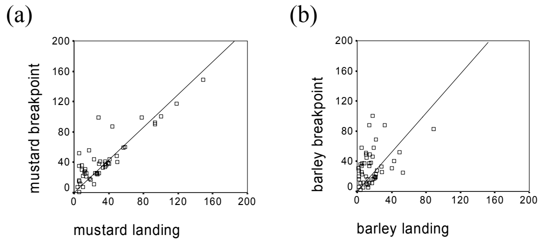


Figure 2. *The bout number of the first landing on the target plant is plotted against the breakpoint detected with the maximum likelihood method. The regression lines and points are given for the combined* **(C)** *and* **(D)** *set-ups of the mustard (a; r^2 = 0.89) and* **(E)** *and* **(F)** *set-ups of the barley (b; r^2 = 0.42).*

In all five cases (not shown) the regression line with intercept set to zero explained a large part of the variation (P <<0.001). The r^2 for these regression lines in the set-ups **(B)**, **(C)**, **(D)**, **(E)** and **(F)** were 0.42, 0.80, 0.96, 0.53 and 0.62 respectively. Figure 2 shows the result for the mustard **(CD)** and barley **(EF)** set-ups, pooled over the mixed and row configurations because of similarities. The plot for the high-density monotypic set-up **(B)** is not given but looks similar to Fig. 2b.

It should be noted that the method to detect a breakpoint in a behavioral record by maximization of $\Lambda(T)$ (formula 2) is not very good at detecting these points near the beginning or the end of a behavioral record [1]. In the barley **(EF)** and high-density monotypic **(B)** set-ups the first landing on the target plant was near the beginning of the behavioral record.

Based on the high correlation between the bout numbers of the first landing and the detected breakpoint, we continued our analysis with separate first-order behavioral transition matrices before and after the first landing.

3.2 Analysis with MatMan

For the analysis with MatMan we thus started with a total of 12 first-order transition matrices. When comparing all six normalized matrices for the behavioral records after the landing on the target plant, the null hypothesis of no association between the matrices was never rejected when row and mixed configurations, low and high-density monotypic set-up or the species of the companion plants were compared in pairs. The Pearson correlation was 0.81 or higher at a significance level for the permutation test of $P = 0.014$ or lower.

For the parts of the behavioral records before the first landing on the target plant we determined the correlation between the high and low-density monotypic set-up and between the mixed and row configurations for the two companion plant species. Table 1 shows that the three squares with a white background located around the diagonal indicate that the alternative hypothesis of association is accepted. In the off-diagonal square matrices with gray shades the null hypothesis of no association could not be rejected in all cases, suggesting an effect of the plant species in the environment.

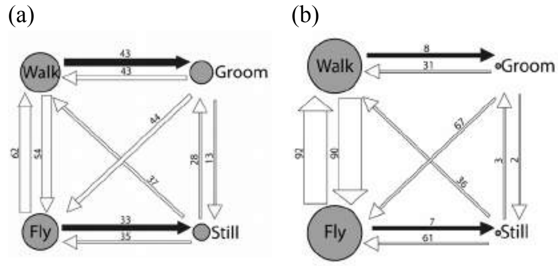
After pooling these first parts of individual behavioral records into three classes (**(AB)**, **(CD)**, **(EF)**) only representing the species composition the resulting transition matrices for the monotypic **(AB)** and mustard **(CD)** set-up were correlated significantly (Pearson's r 0.92, $P < 0.001$). Between the barley **(EF)** set-up and the other two set-ups no association could be shown (for **(AB)** and **(CD)**, P values of 0.133 and 0.086 respectively).

Table 1. *The upper triangular matrix contains the estimated values of Pearson's* r *and the lower triangular matrix the significance level. An asterisk * is given for a significance level P < 0.001.*

	SL	**SH**	**MM**	**MR**	**BM**	**BR**
SL	--	0.90	**0.62**	0.97	**0.81**	**0.76**
SH	*	--	0.74	0.86	**0.61**	**0.56**
MM	**0.08**	0.04	--	0.68	**0.35**	**0.63**
MR	0.04	*	0.04	--	0.89	**0.85**
BM	*	**0.12**	**0.27**	0.04	--	0.81
BR	**0.11**	**0.17**	**0.09**	**0.05**	*	--

3.3 Proportional fitting

We used the iterative proportional fitting method of Goodman [3] to find the expected values for the non-zero matrix elements. The observed behavioral transitions within the three set-ups were significantly different from expected transitions (G-test, $P<0.05$). Thus, the females did not display different behavioral elements by chance only. When the transitions were tested around single behaviors, certain behavioral transitions significantly deviated in a positive or negative way with respect to the expected number of transitions. The structure of the transitions between the behavioral elements is presented in the kinetograms (Figure 3abc), where black (missing) arrows represent behavioral elements that occur more (less) often than expected by chance. The white arrows show transitions that are not significantly different from the expected values. The size of the circles is proportional to the overall frequency of the behavioral elements. The thickness of an arrow is proportional to the likelihood of occurrence of each behavioural transition. Numbers on arrows are percentage transitions to succeeding behaviors.
In the monotypic set-up **(AB)**, flight behavior was often followed by a short stop that was often followed by walking or in a few occasions by preening (fig. 3a). Walking was more than proportionally followed by preening. The behavior of females searching in the mustard set-up **(CD)** was similar to that in the monotypic set-up **(AB)**: the same transitions were significantly absent or present (fig. 3b). In the barley set-up **(EF)**, the behavior of females was unlike that in the monotypic **(AB)** and in the mustard **(CD)** set-ups (fig. 3c). After landing on a plant, females started walking. Walking and standing still often occur before preening, which is often followed by flight behavior.



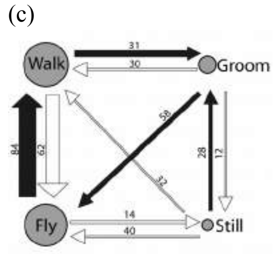


Figure 3. *Behavioral kinetograms of* D. semiclausum *females until the target plant is found. They were searching in monotypic (a), mustard (b) or barley (c) set-ups.*

3.4 Analysis with Theme

Theme is developed to find structures that repeat themselves throughout a behavioral record. With Theme we found differences between (1) the number of different patterns, (2) the total number of occurrences of patterns in behavioral records and (3) the mean number of events in the detected patterns. The Kruskal-Wallis analysis of variance showed a significant effect of the experimental set-up for all three characteristics ($P << 0.001$).

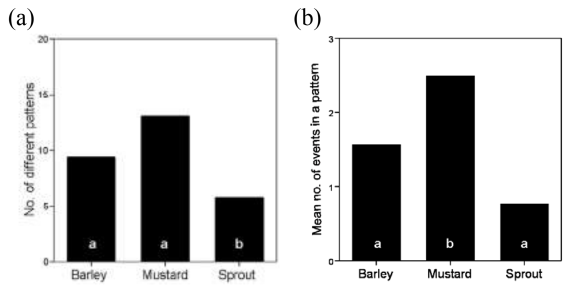


Figure 4. *Bar plots for the number of different patterns (a) and the mean number of events in a pattern (b) as detected by Theme. Bars with different letters differ significantly (3P < 0.05) according to a pairwise comparison with the Mann-Whitney* U *test and Bonferroni correction.*

Figure 4a shows that the numbers of different patterns in the mustard **(CD)** and barley **(EF)** set-ups are not significantly different from each other whereas they are different from the monotypic **(AB)** set-up. The mean length of the patterns is not different between the barley **(EF)** and monotypic **(AB)** set-up, but this variable is significantly higher in the mustard **(CD)** set-up (Fig. 4b). The total number of occurrences in the mustard **(CD)** set-up (not shown) is higher than in the barley **(EF)** set-up ($P = 0.001$) or the monotypic **(AB)** set-up ($P << 0.001$). Behavioral records from the barley **(EF)** set-up contained significant more patterns than the monotypic **(AB)** set-up ($P = 0.003$). Summarizing, the analysis with Theme reveals that the behavior of the parasitoid *D. semiclausum* is less complex in a monotypic environment than in environments containing mustard or barley plants.

4 Conclusion and discussion

The analysis of behavioral records as performed in this study uncovers a lot of information that was hidden in the behavioral data.

It is known that the method to detect a breakpoint in a behavioral record by maximization of $\Lambda(T)$ is less suitable for detecting such points near the beginning or the end of a behavioral record [1]. Our data supported this finding: the first landing on the target plant in the mustard set-up occurred after more behaviors and thus correlates well with the breakpoint. In the other two set-ups the majority of the first landings on the target plant occurred right near the beginning of the behavioral record (e.g. Fig. 2b), and the test could not always locate this as a breakpoint.

We used the results of the analysis with MatMan only for pooling our behavioral records before the first landing in low and high-density monotypic set-ups and for pooling the mixed and row configurations for both the barley and mustard set-up. The behavioral transition matrix before the first landing on the target plant in the barley set-up was different from both other set-ups. This may be explained by the fact that both mustard and Brussels sprout plants belong to the same plant family. They can both be a host plant for the pest insect *P. xylostella.* However, barley is from an unrelated plant family and never contains host larvae of *P. xylostella*.

The behavioral kinetograms constructed after performing Goodman's [3] proportional fitting analysis give support to the fact that the parasitoid is searching differently when potential host plants (Brussels sprouts) are interspersed with non-host plants (barley). Detected treatment differences in within and between matrices comparisons were similar, showing complementary results with MatMan and the proportional fitting method.

The analysis with Theme reveals mostly details about the more complex behavior of the parasitoid in the mustard set-up, where two related potential host plants are combined.

With these different analyses we are able to compose an overall picture of the behavioral elements displayed in the different environments. When a complete quantitative description with the timing of the behavioral elements is wanted, a (semi)Markov approach [1] can also be used for estimating transition rates between the behavioral elements.

We like to thank Dr. M.A. Keller for the use of his software to construct the kinetograms from the behavioral transition matrices. A. Willemsen from Noldus Information Technology is kindly acknowledged for giving us the opportunity to analyze our data with the most recent Theme version and for discussions.

References:

1. Haccou, P.; Meelis, E. (1992). Statistical Analysis of Behavioural Data: An Approach Based on Time-Structured Models. *Oxford University Press, Oxford*.
2. De Vries, H.; Netto, W.J.; Hanegraaf, P.L.H. (1993). MatMan: a program for the analysis of sociometric matrices and behavioural transition matrices *Behaviour*, **125**, 157-175
3. Goodman, L.A. (1968). The analysis of cross-classified data: independence, quasi-independence and interactions in contingency tables with or withour missing entries *Journal of the American Statistical Association*, **324**, 1091-1131
4. Magnusson, M. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection *Behavior Research Methods, Instruments and Computer*s, **32**, 93-110
5. Noldus Information Technology (2003) MatMan, Reference Manual 1.1. *Wageningen, The Netherlands*
6. PatternVision Ltd and Noldus Information Technology (2005). Theme^{TM}: Powerful tool for detection and analysis of hidden patterns in behavior Reference Manual Version 5.0 *Wageningen, The Netherlands*
7. Gols, R.; Bukovinszky, T.; Hemerik, L.; Harvey, J.A.; van Lenteren, J.C.; Vet, L.E.M. (2005). Reduced foraging efficiency of a parasitoid under habitat complexity: Implications for population stability and species co-existence. *Journal of Animal Ecology*, in press.
8. Noldus Information Technology (1993). The Observer: system for collection and analysis of observational data Version 3.0 *Wageningen, The Netherlands*
9. Noldus Information Technology (2003). The Observer: system for collection and analysis of observational data Version 5.0 *Wageningen, The Netherlands*
10. Freeman, L.C.; Freeman, S.C.; Romney, K. (1992). The implications of social structure for dominance hierarchies in red deer, *Cervus elaphus* L. *Animal Behaviour*, **44**: 139-145.
11. Van Dierendonck, M.C.; De Vries, H.; Schilder, M.B.H. (1995). An analysis of dominance, its behavioural parameters and possible determinants in a herd of icelandic horses in captivity *Netherlands Journal of Zoology*, **45**, 362-385
12. Dietz, E.J. (1983) Permutation tests for association between two distance matrices *Systematic Zoology*, **32**, 21-26
13. Rice, W.R. (1989). Analyzing tables of statistical tests. *Evolution*, **43**, 223-225
14. Field, S.A.; Keller, M.A. (1993). Courtship and intersexual signalling in the parasitc wasp *Cotesia rubecula* (Hymenoptera: Braconidae). J*ournal of Insect Behavior*, **6**, 737-750
15. Wang, X.-G.; M.A. Keller (2002). A comparison of the host-searching efficiency of two larval parasitoids of *Plutella xylostella Ecological Entomology*, **27**, 105-114
16. Sokal R.R.; Rohlf, F.J. (1995) Biometry *W. Freeman Publishing*

Using state space grids to display, describe, quantify, and analyze synchronized time series or event sequences

T. Hollenstein
Department of Psychology, Queen's University, Kingston, Ontario, Canada

Abstract

The recent increase in observational research has created a demand for methods that can visually display, describe, quantify, and analyze two or more synchronized time series or event sequences. The current paper describes a new methodology based on dynamic systems principles, state space grids, which is well-suited for the kind of data obtained via observational programs such as the Noldus Observer.

State space grids are two-dimensional grids constructed from synchronous sequences of two categorical data streams. Each axis represents mutually-exclusive categories along a single dimension, and the cells on the grid represent joint states of categories along the two dimensions. Behavior is traced as a trajectory across this space: a new point is plotted for each change in state and a line is drawn that connects the old and new state. Examples are presented from the freely available software program GridWare (www.statespacegrids.org).

Keywords

Dynamic Systems; State Space Grids; Developmental Science, Dyadic Interactions; Time Series

1 Introduction

Advances in computing have accelerated the growth in observational methodologies and have allowed researchers to capture complex sequences of events as they unfold in time. This increase in observational data has created a demand for methods that can handle the complexity of these investigations. However, there are few applications that can visually display, describe, quantify, and analyze two or more event sequences. The current paper describes state space grids, a new methodology based on dynamic systems (DS) principles that is well suited for the kinds of data obtained via observational programs such as the Noldus Observer.

1.1 Dynamic Systems

State space grids were inspired by DS principles that describe the processes of change and stability in complex, adaptive, open systems. Most simply, a dynamic system is a "system of elements that change over time" [9]. All dynamic systems have several key properties in common. Although a system may have a wide range of states available, there is a tendency to stabilize in a limited number of these possibilities. In DS terms, these recurrent, stable states are called attractors. These are "absorbing" states that "pull" the behavior of the system from other potential states [1]. The full range of all possible states is called a state space, which is configured by the presence of attractors. This is often represented as a topographical landscape on which valleys are the attractors.

1.2 State Space Grids

Based on these abstractions, Lewis and colleagues [7] recently developed a graphical approach that utilizes observational data and quantifies these data according to two ordinal variables that define the state space for the system. This technique was first used for analyzing two dimensions for one individual but was adapted to study dyadic behavior, specifically parent-child interactions [3]. With this method, the dyad's behavioral trajectory (i.e., the sequence of behavioral states) is plotted as it proceeds in real time on a grid representing all possible behavioral combinations. Much like a temporal scatter plot, the parent's coded behavior is plotted on the x-axis and the child's behavior is plotted on the y-axis. Any time there is a change in either person's behavior a new point is plotted in the cell representing that behavior and a line is drawn connecting the new point and the previous point. Thus, the grid represents a sequence of dyadic events. For example, a hypothetical trajectory representing 15 seconds of parent-child behavior is presented in Figure 1. The state space is formed by the intersection of an ordinal set of affect categories for both parent and child: High Negative, Low Negative, Neutral, Low Positive, and High Positive. The size of the plot point corresponds to the duration of the dyadic behavior. The sequence depicted begins in the mutually Low Positive cell followed by 3 seconds in the Mother Low Positive/Child Low Negative cell, 4 seconds in the mutually Low Positive cell again, 2 seconds in the Mother Neutral/Child Low Positive cell, 2 seconds in the Mother Neutral/Child High Negative cell, and finally 2 seconds in the mutual High Negative cell.

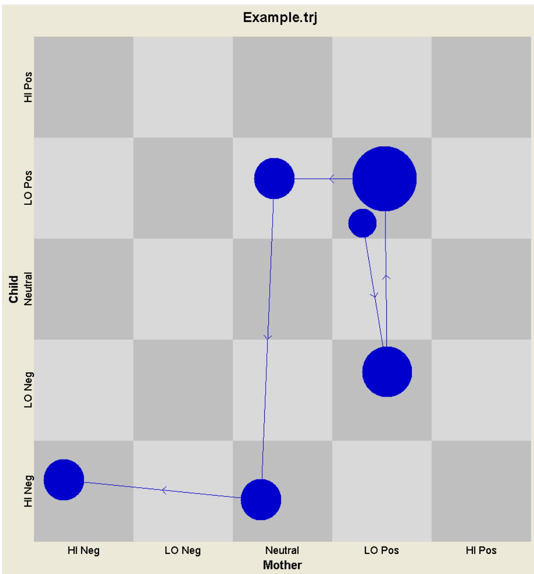


Figure 1. *Example state space grid.*

The work of myself and my coleagues has focused on using this technique to display and quantify patterns of infant emotional development [7,8], and patterns of parent-child interactions [2,3,4,5]. However, this technique is not limited and can be used for a wide range of applications. Some possibilities include eye tracking (horizontal vs. vertical plane), group dynamics, reaction time and performance combinations, bivariate economic trends, and biological and behavioral combinations. Before describing examples of this technique, I will first describe the software that can be used to create state space grids.

2 GridWare

Recently, my colleagues and I developed GridWare [6], a software program for the creation of state space grids that can be downloaded for free from the web (www.statespacegrids.org). This program allows the user to view and manipulate trajectories based on 2 or more synchronized data streams. Thus, it is quite amenable to any multiple pass coding within the Noldus Observer. In fact, we have created a Perl script for this conversion. The input for this program is a control file, called the GridWare File, and one file for each trajectory (synchronized time series data). The GridWare File provides the parameters of the grid dimensions (names and # of categories, names of dimensions, etc.) and identifies each trajectory file. It is also possible to add variables that uniquely identify characteristics of each directory (i.e., treatment group, gender, session #). The trajectory files themselves have a minimum of three columns each: onset, state variable 1, and state variable 2 (although more than 2 state variables are acceptable). When a project is opened in GridWare, three windows appear. These are described below.

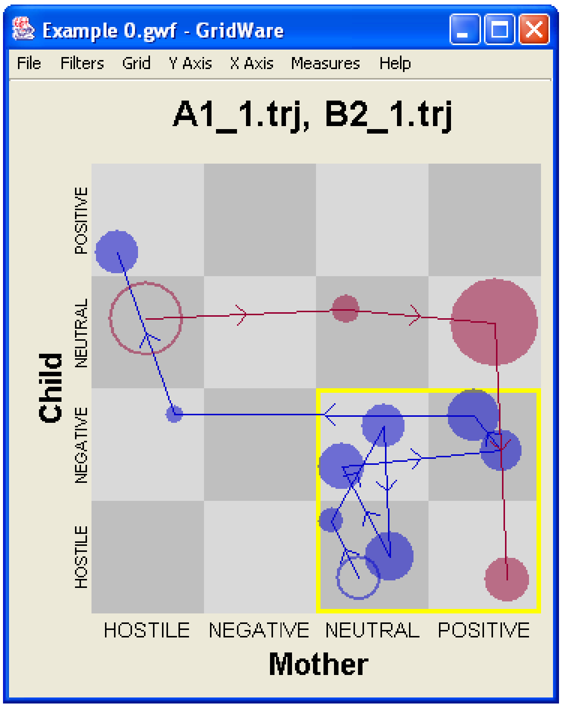


Figure 2. The Grid Window displaying 2 trajectories.

2.1 The Grid Window

Figure 2 shows an example of 2 trajectories displayed in GridWare. GridWare can display none, one, some, or all of the trajectories in a project. If there are more than 2 streams of simultaneous data (state variables), the axis menus allow the user to choose which to display on each axis. Some of the display options include: Arrows to indicate the direction of the trajectory in time, hiding the plot points or transition lines, selecting one or more cells to create a region (the highlighted yellow quadrant), and color can be applied to any selection, as seen here by one trajectory in red and the other in blue.

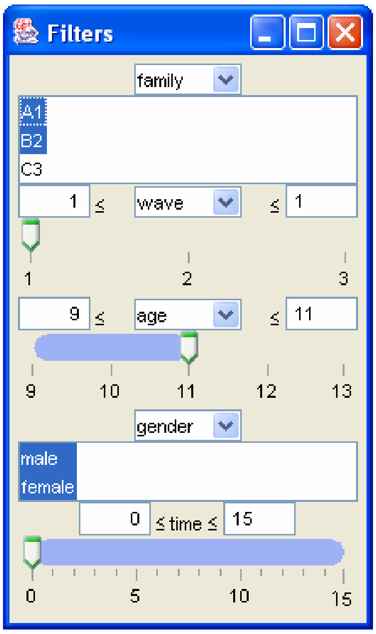


Figure 3. The Filters Window that can be used to select one, some or all of the trajectories.

2.2 The Filters Window

Figure 3 shows the filters window of the example trajectories displayed in Figure 2. In this example, the trajectories selected are from families A1 and B1 at wave 1 of the study and these children were between the ages of 9 and 11 at that time. Any number of filter variables can be included in a project, although only up to 6 can be displayed in the filters window at any one time. Categorical filters (Family and Gender in this example) allow the user to select multiple, non-adjacent values (i.e., A1 and C3) but ordinal filters only allow a continuous selection. The last filter is always time. This slider can be used to select a window of time, or it can be used to scroll through the sequence displayed by selecting a window and then moving the slider. The selection of filters also constrains the measures that are displayed or exported for analysis.

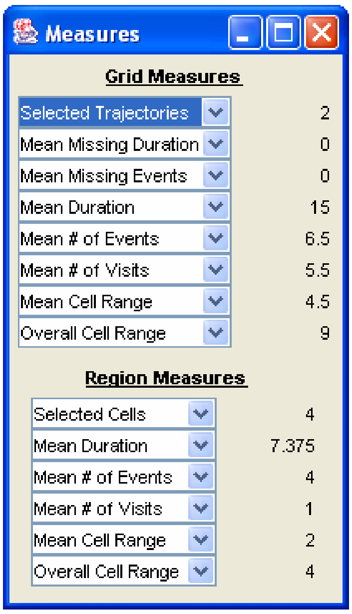


Figure 4. The Measures Window with a few of the many possible measures displayed for the whole grid and the selected region (see Figure 2).

2.3 The measures window

Figure 4 shows the measures window of the example trajectories displayed in Figure 2. Measures derived from the grids can either describe patterns across the whole grid or within a cell or region that has been selected. For the whole grid, these include: total and mean duration, number of events, transitions, number of cells occupied, and dispersion (an index based on proportional durations). These are also available for selected regions. There are specific region or cell measures as well. All of these measures, including measures for each individual cell, can be exported as a tab delimited text file for use in other statistical packages.

3 State space grid examples

State space grids open up new possibilities for research designs and analyses. The measures derived from state space grids have been successfully applied to a handful of studies on child and adolescent development and several more are in the works. The examples I have chosen represent how this technique has been applied at different time scales. In each of these examples, parents and children were coded for their emotional displays during several tasks, usually discussions. These emotion codes formed the categories of each grid dimension.

3.1 Rigidity in parent-child interactions and the development of problem behaviors in early childhood [5]

The continuum from rigidity to flexibility is easily captured by several grid measures. As shown in Figure 5, the grid on the left is more rigid (less flexible) than the grid on the right. In particular, the dyad on the left occupies fewer cells, makes fewer transitions (fewer lines), and has higher mean durations (bigger plot points) than the flexible dyad on the right. Using these measures, we were able to show that higher rigidity in parent-child interactions at the beginning of kindergarten predicted growth in aggressive and antisocial behaviors over the subsequent 18 months.

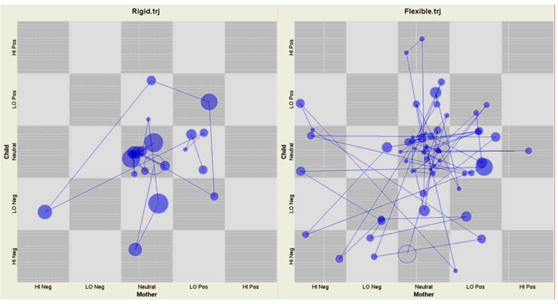


Figure 5. *Two example state space grids that differ on the rigidity-flexibility dimension. The grids on the left depicts more rigid behavior, the grid on the right depicts more flexible behavior.*

3.2 A developmental phase transition across early adolescence [2]

As an adaptive dynamic system changes in structure it goes through a *phase transition*: a period of instability and high variability observed when one stable pattern or structure breaks down and a new structure emerges in its place. The transition through early adolescence has been described in the same way. In this study we examined whether families with boys traversing the transition into adolescence would show a peak in variability characteristic of a phase transition. Boys and their parents were observed every other year for ten years, beginning when the boys were 9-10 and ending when they were 17-18 years old. The middle of the 5 waves corresponded with early adolescence (age 13-14). Thus, we examined whether the number of cells or transitions derived from the state space grids of their interactions would peak at the transition wave. As shown in Figure 6, both the number of cells and transitions peaked right at the early adolescent transition.

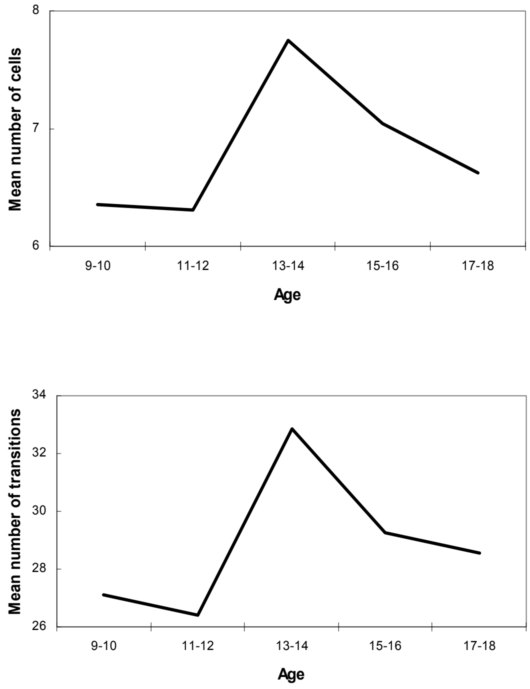


Figure 6. *The mean number of cells (top) and transitions (bottom) during parent-boy interactions from late childhood to late adolescence.*

3.3 Measuring behavioral change through comparisons of state space grids over time [8]

This study took a novel approach to examine changes in state space grid patterns for evidence of a phase transition. These grids were constructed from ordinal data on one toddler along 2 dimensions during a frustrating toy task: attention to the toy and attention to mother. Thus, repeated activity in one cell or region was considered a child's behavioral attractor. The toddlers were observed once a month for 12 months and grids were constructed for each time point. In order to test for the degree of stability or change from month to month, 2 novel methods were used. First, the durations in the cells of all of the grids across all of the months were entered into a cluster analysis. Each grid was identified as a member of one cluster. Then, each child received a month to month cluster change score (1 if the cluster changed, 0 if it didn't). The average cluster change score peaked at the hypothesized phase transition age of 18-20 months. The second method to compare grids was an inter-grid distance score. With this method, each cell in grid A was subtracted from the same cell in grid B using a Euclidian distance algorithm. The sum of the squared differences between monthly grids also peaked at the hypothesized transition age.

3.4 The effect of negative emotions on flexibility [4].

Flexibility is known to increase under conditions of positive emotions and decrease in the presence of negative emotions. Until the advent of state space grids, socioemotional flexibility has been difficult to measure. One of the goals of this study was to use state space grids to look at changes in mother-daughter flexibility as they went from discussing a positive topic to a self-identified unresolved conflict and then back to a positive topic again. State space grids can be used to look at specific content or attractors as well as overall patterns of behavior. In this study, negative emotion was measured as the number of visits to and the duration in the negative emotion region of the state space. Flexibility included cells and transitions, as in the previous study, but also mean duration and dispersion (a new devised measure that combines duration information with the number of cells). Negative emotion, as expected, was highest during the conflict and dropped off again during the last, more positive interaction. Flexibility changed in the opposite way, as expected – the dyads were most flexible during the positive discussions. Two other interesting results emerged. The dyads were less flexible (but not more negative) during the last positive discussion. Thus, the conflict had a lingering effect. The girls were grouped according to the amount of current stress (high vs. low). The low stress group responded to the changing contexts in the way just described for the whole sample. The high stress group, however, were less flexible in the positive tasks and showed little change across the different contexts.

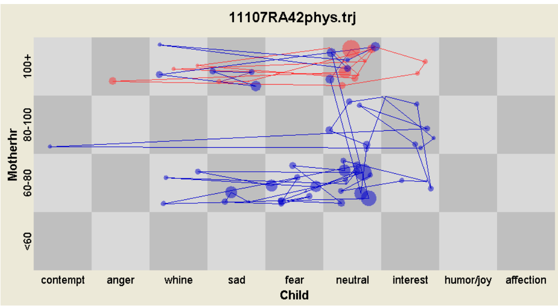


Figure 7. *Child's affect with mother's heart rate. Red indicates when child's heart rate was high.*

3.5 Combining psychophysiological data with coded observations

With the release of the next version of the Noldus Observer, researchers will be able to easily record video and psychophysiological data simultaneously. This opens up a whole world of possibilities for the application of state space grids. My colleagues and I have just started analyzing some pilot data that combines these two measurement types and an example grid is shown in Figure 7. This is a grid of mother's heart rate (y-axis) and child's affect (x-axis) and the child's highest heart rate is colored red. In this example, you can see that both the mother's and the child's heart rate was in their highest category at the same time. Of course, many issues of collapsing the continuous heart rate data emerge (the next version of GridWare will allow for continuous and categorical data), but as this is the very first image of behavior and biology on a state space grid, the possibilities are exciting. With better measurement techniques this can only improve.

4 Conclusions and future directions

Although state space grids have so far been used primarily for developmental research on social and emotional processes, there are clearly many more research areas that could benefit from this technique. I am currently working on various projects that use hourly and daily sampling methods, eye-tracking data, and group interactions. One of the tricks with this research is how to construct the state space, which informs the coding process as well. As with coding, the construction of the state space is an integral part of the research design [1]. My hope is that more researchers from varied disciplines will find this technique useful and informative for their inquiries.

Acknowledgements: The author was supported by National Research Service Award fellowship from the National Institute of Mental Health (F31-MH068046-01). Many thanks to my colleagues Marc Lewis, Isabela Granic, and Alex Lamey and for the enormous efforts of Bart van Roekel of Noldus, Inc.

References:

1. Granic, I., & Hollenstein, T. (2003). Dynamic systems methods for models of developmental psychopathology. *Development and Psychopathology,* **15,** 641-669.
2. Granic, I., Hollenstein, T., Dishion, T. J., & Patterson, G. R. (2003). Longitudinal analysis of flexibility and reorganization in early adolescence: A dynamic systems study of family interactions. *Developmental Psychology,* **39**, 606-617.
3. Granic, I. & Lamey, A.V. (2002). Combining dynamic-systems and multivariate analyses to compare the mother-child interactions of externalizing subtypes. *Journal of Abnormal Child Psychology.***30**, 265-283.
4. Hollenstein, T. (2005). Socioemotional variability at the early adolescent transition. *Unpublished doctoral thesis, University of Toronto.*
5. Hollenstein, T., Granic, I., Stoolmiller, M., & Snyder, J. (2004). Rigidity in parent-child interactions and the development of externalizing and internalizing behavior in early childhood. *Journal of Abnormal Child Psychology,* **32**, 595-607.
6. Lamey, A., Hollenstein, T., Lewis, M.D., & Granic, I (2004). GridWare (Version 1.1). [Computer software]. http://statespacegrids.org.
7. Lewis, M. D., Lamey, A. V., & Douglas, L. (1999). A new dynamic systems method for the analysis of early socioemotional development. *Developmental Science, 2,* 457-475.
8. Lewis, M.D., Zimmerman, S., Hollenstein, T., & Lamey, A.V. (2004). Reorganization in coping behavior at 1 1/2 Years: Dynamic systems and normative change. *Developmental Science,* **7**, 56-73.
9. Thelen, E., & Smith, L.B. (1998). Dynamic systems theories. In W. Damon (Ed.) *Handbook of Child Psychology, 5th Edition, Vol. 1. Theoretical models of human development.* New York, Wiley.

Measuring the Step and Kick Behavior of Dairy Cows During Robotic Milking

M. Pastell[1], M. Hautala[1], V. Poikalainen[2], J. Praks[2], I. Veermäe[2] and J. Ahokas[1]
[1] *Department of Agrotechnology, University of Helsinki, Helsinki, Finland*
[2] *Estonian Agricultural University, Tartu, Estonia*

Abstract

We have worked on automatically measuring the physical health of a dairy cow. Milking robot offers a unique possibility for a dynamic measurement of physical data. Four strain gauge scales were installed into a milking robot in order to measure the weight of each leg separately. The sensors were connected to an amplifier and the data was collected into a PC using a dedicated computer program. From the data the dynamic weight or load of each leg can be measured. The average weight, the weight variation of each leg, the total weight, the number of kicks and steps, the frequency of kicks and the total time in the milking robot can be calculated. The acquired information can be used to judge cow's restlessness and welfare, e.g. leg health. The leg weight monitoring was developed into a real time system with software that alerts the user of possible hoof diseases and other leg problems.

Keywords

Step and kick behavior, Robotic Milking, Leg weight distribution.

1 Introduction

The step and kick behaviour of a cow can be used when studying animal welfare. Rousing et al found that higher kick frequency during milking may be the result of pain or discomfort caused by e.g. teat lesions [1]. However they did not find any relation between lameness and increased stepping or kicking during milking. They suggest that step and kick behaviour measurement might be a good tool for monitoring welfare problems such as udder health and milking techniques. According to Wenzel et al. increased step frequency during milking correlates with increased heart rate and milk cortisol concentration [2].

Milking robot offers a unique possibility for a dynamic measurement of physical data. We have worked on automatically measuring the leg weights of dairy cows during automatic milking. From the data the average weight, the weight variation of each leg, the total weight and the number and frequency of kicks and steps can be calculated. The aim of this paper is to describe the measurement system and discuss about the possibilities of monitoring step and kick behaviour.

2 Materials and methods

2.1 Measurements

A system for automatically measuring the leg weights of dairy cows was developed. The system consists of four balances, wiring, a carrier frequency amplifier and a computer for collecting the data. The balances are located in the floor of the milking robot under a rubber carpet (Figures 1 and 2). Each balance is constructed from a plywood plate on top of a steel plate that is attached to a strain gauge (Tedean 1510) with the capacity of 500 kg. The size of the balances ranges from 310 mm x 310 mm to 445 mm to 390 mm depending on the location and robot's chassis structures. The plates that are located under the hind legs are a bit larger than the plates under the front legs, because it was found out that the position of hind legs of different cows during milking varies more than the position of front legs. Figure 1 shows the installation of strain gauges in the floor of the robot. Figure 2 shows the plates on top of the four balances as they are positioned in the robot.

Figure 1. *The strain gauges on the floor of the milking robot.*

Figure 2. *The plywood plates on top of the balances as they are positioned on the floor of the milking robot*

The computer and the carrier frequency amplifier (Spider, HBM) were located in the office of the cow barn so 40 m long wiring was needed. The balances were also

automatically tared during the washing of the robot, which took place at least three times a day. In spite of the long wires the system works with low noise level, but some drifting still occurs as a function of temperature.
The measurements were controlled by a dedicated computer program made with TestPoint 5.0. The software controlled the measurements based on the cow ID and the robots actions. This information was obtained from a log file created by a software provided by DeLaval. The software started to record the leg weights of a cow when the milking started and saved the recorded data on the hard disk when the milking ended. The measurement rate was set to 10 Hz in order to see the kicks properly, the duration of a kick can be as short as 0.5 s. After the measurement had ended the software automatically calculated the mean weight and the standard deviation of each leg and the number of kicks during milking. These values were also saved for each cow.
A web camera was also employed to record the cow behavior during milking. The video image was also recorded with a computer for later examination. Also a remote connection via internet was used for downloading the data as well as tracking of the measurements together with the real time video. Figure 3 shows the components of the measurement system.

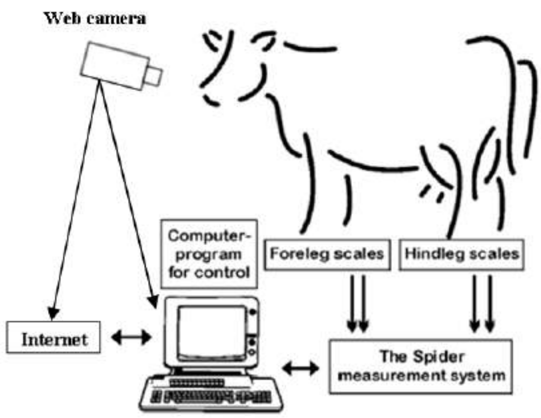


Figure 3. The components of the measurement system.

2.2 Data Analysis

Data from over 13000 milkings and 72 cows were recorded during 5 months. The recorded data was downloaded from the measurement computer and MATLAB 7.0.4 was used for analyzing the data. The number of kicks and steps during each milking was calculated.
A kick was defined as cow lifting the leg up and one kick was calculated when the leg weight of a cow decreased to below 5 kg (the drifting sometimes moved the zero level to 5 kg). A step was defined as weight shuffling and a step was calculated from the data when the weight of a leg decreased to between 5 and 20 kg.
The mean number of steps and kicks during milking were calculated for primiparous and multiparous cows. A t-test for two independent samples was conducted to find out if there is a statically significant difference in the step and kick behavior of multiparous and primiparous cows during milking.
Also the leg health of the cows was regularly inspected during this study and the effect of lameness to the cow behavior was looked into.

3 Results

Figure 4 shows the mean number of kicks and steps during one month for multiparous and primiparous cows. The average number of kicks for multiparous cows was 9.9 and for primiparous cow 4.5 times/milking. The average number of steps for multiparous cows was 25.9 and primiparous cows 14.6 times/milking. The figure also shows that the deviation in the number of steps and kicks is a lot higher for primiparous cows. Also a t-test gave a statistically significant difference ($p<0.01$) between the mean number of steps and kicks of multiparous and primiparous cows.

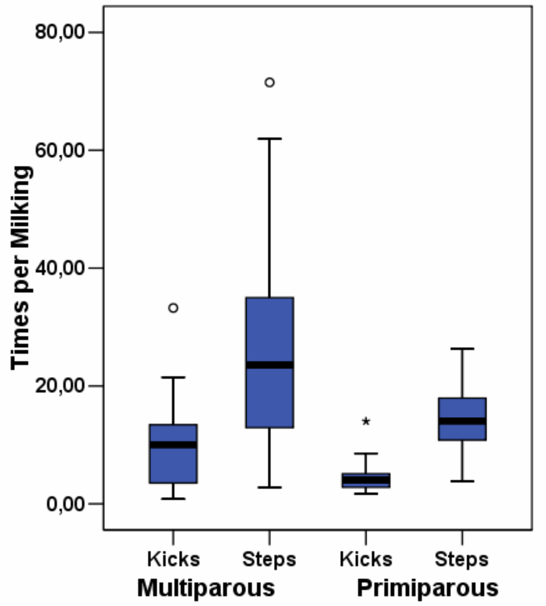


Figure 4. The mean number of kicks and steps during milking in August 2004 for multiparous (n=29) and primiparous (n=21) cows.

Figure 5 shows the number of kicks and steps of a cow that has injured a leg during the measurements. The cow injured her leg on 25th of August and that can be seen in the data as a significant increase in the stepping and kicking frequency. The high number of steps is caused by the cow constantly lifting the injured leg. The injury healed by itself and the step and kick behavior of the cow returned back to normal.

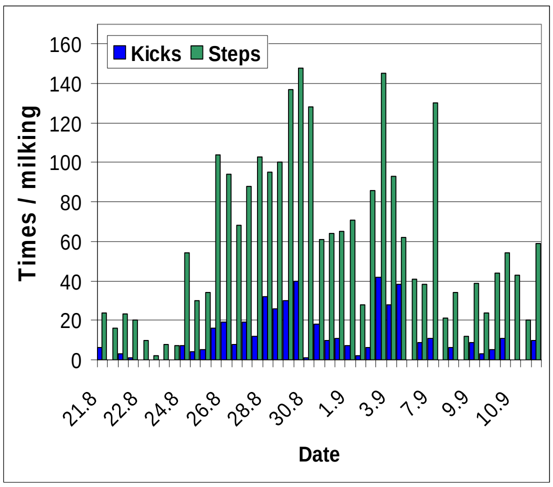


Figure 5. The number of kicks and steps of one cow during milking between 21st of August and 10th of September.

It is also possible that mastitis can be seen in the step and kick behavior of a cow. Figures 6 and 7 show the weight of hind legs of a cow with mastitis and a healthy cow during milking. The cow with mastitis has been restless during the hold milking and kick frequency has been very high, whereas the healthy cow has stood very evenly and the kick frequency is low. The kicks can be seen from the data as sharp drop and rise in the weight of the kicking leg.

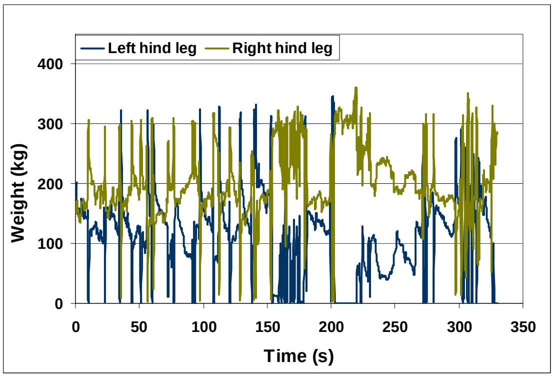


Figure 6. *The weight of the hind legs of a cow with mastitis during milking.*

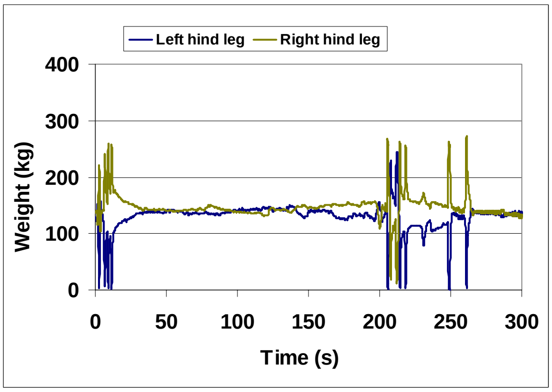


Figure 7. *The weight of the hind legs of a healthy cow during milking.*

4 Discussion

The system can be used for real time monitoring of step and kick behavior of dairy cows. Measurements have been fully automated and algorithms for counting the number of kicks and steps have been developed. However, the balances are constantly subject to severe loads and the fixing of the transducers and platforms has been problematic. This is due to the loading of leg, since the transducers were not designed for concentrated point loads but for larger load area.

A calm cow can be separated from a restless cow and changes in the behavior of individual cows can be monitored. The multiparous cows had higher step and kick frequency during robotic milking than primiparous cows and there was also more variation in the behavior of multiparous cows.

The system has also proved to be useful for lameness detection, since clear lameness cases result in very high stepping frequency, when the cows try relieve the pain on the affected leg. It might also be possible to use the system for mastitis detection.

5 Conclusions

The distribution of leg weights and their dynamics in time can provide a basis for the detection of leg disorders. To a certain extent the cow's behavior influences the leg load distribution. An automatic system based on leg weights can be used to monitor the health status of animals.

The behaviour of a cow can be seen from the data, i.e. a calm cow can be separated from a restless cow. The number of kicks can be calculated from curves and a quantitative measure of restlessness can be figured out.

In future the milking robots or feeding automates can be equipped with an automatic weight measuring system and alert system to announce changes in the behavior of a cow.

References:

1. Rousing T; Bonde M; Badsberg J H; Sørensen J T (2004). Stepping and kicking behaviour during milking in relation to response in human-animal interaction test and clinical health in loose housed dairy cows. *Livestock production science*, **88**, 1-8.
2. Wenzel C; Schönreiter-Fischer S; Unshelm J (2003). Studies on step-kick behavior and stress of cows during milking in an automatic milking system. *Livestock Production Science*, **83**, 237-246.

The Virtual Raceway Method: a novel method for assessing perception in the social environment

L. M. Collins
Department of Zoology, University of Oxford, Oxford, UK

Abstract
An animal's social surroundings are complex and dynamic. Captive animals that are kept at high stocking densities, such as the commercial broiler chicken, are forced to live in close proximity to thousands of other individuals. How do the birds themselves perceive this kind of environment? Whether high stocking densities are as distressing to animals as we humans think they are, is one of the most urgent and controversial issues in farm animal welfare, and is extremely difficult to answer with our present methodology.

Keywords
Broiler chicken, social behavior, raceway, tracking, preference.

The experimental raceway method is used by researchers to test for aversion to certain stimuli under controlled experimental conditions [2, 3]. Despite the wide-spread availability of video and CCTV, relatively little attention has been given to how to utilize and analyze it in terms of what it can demonstrate about the preferences and aversions of animals in our charge. Here I have developed and used a new *in situ* virtual raceway test that involves following the paths of individual birds from video footage taken inside commercial broiler houses. This method allows researchers to track behaviour *in the place* where there is concern for animal welfare.

Using video footage filmed inside 20 UK commercial broiler houses (for further details on video set-up methodology see [1]), I developed and used a virtual raceway method of tracking to observe behavioral differences between houses with different manipulated target stocking densities (30, 34, 38, 42 and 46 kg/m^2). Tracking involved choosing a feeding bird at random from video footage, then rewinding the tape until the point in time when the bird first stood up to approach the feeder (an obvious start point, equivalent to the start box used in many raceway experiments). Thus the chickens could be seen to be walking down an 'imaginary', *in situ* raceway, passing many other birds on their way to the feeder (an obvious end point, equivalent to the goal boxes used in raceway experiments).

Behavioral performance (type of behavior, frequencies) and body position (distance and angle) of the focal chicken in relation to surrounding birds was recorded. The behavioral performance and body positions of all birds within 1 metre radius ('neighbours') of the focal chicken were recorded in order to compare the effects of local social environment (what neighbouring birds are doing, where they are and how many of them there are), and global target stocking density on raceway performance of focal chickens.

Using the virtual raceway method, it was demonstrated that broilers are influenced by their local social environment, but also by target stocking density. A focal bird's body orientation with respect to neighbouring birds was affected by the behaviors being performed by these neighbours ($F_{10, 881}$=7.41, p<0.0001). Neighbour behavior and stocking density also influenced focal bird behavior ($F_{10, 886}$=3.03, p=0.01, and $F_{4, 886}$=7.88, p<0.0001, respectively).

It is hoped that this method will be easily transferable between different species, permitting us a valuable non-intrusive insight into how animals view the super-social surroundings that we keep them in.

L. M. Collins was supported by the Newton-Abraham Studentship, Lincoln College, Oxford University, U.K.

References:

1. Dawkins, M. S.; Donnelly, C. A.; Jones, T. A. (2004). Chicken welfare is influenced more by housing conditions than by stocking density. *Nature*, **427**, 342–344.
2. Grigor, P. N.; Hughes, B. O.; Appleby, M. C. (1995). Social inhibition of movement in domestic hens. *Animal Behaviour*, **49**, 1381-1388.
3. Rushen, J. (1986). Aversion of sheep to electro-immobilization and physical restraint. *Applied Animal Behaviour Science*, **15**, 315-324.

Early behavioral pathways in infants later diagnosed with Autism Spectrum Disorder. Home Movies analysis through The Observer 4.0®

Apicella, F.; Maestro, S.; Cesari, A.; Grassi, C.; Manfredi, A.; Muratori, P.; Petrozzi, A.; Polidori, L.; Muratori, F.
I.R.C.C.S. "Fondazione Stella Maris" – Università di Pisa - Italy

Abstract

This study investigates the influence of Autism Spectrum Disorder (ASD) on early development of social behaviors in infants later diagnosed with ASD.

Methodological issues in using retrospective video-analysis and computer-based coding system as a useful tool for early identification of autism social deficit will be discussed.

The Infant Behavioral Scale (IBS), a new grid for the detection of social behavior in the first 18 months of life, will be introduced.

Keywords

Autism Spectrum Disorder; Home-Movies; Social behaviors; Infant Behavioral Scale; Social deficit.

1 Introduction

Autism is an expression of a specific disease process that, acting through genetic factors, results in abnormal brain development and the autism syndrome characterized by qualitative impairments in social interactions and communication, and by restricted interests and activities.

These disturbance range from subtle anomalies in social reciprocity and intersubjectivity to much more obvious difficulties in the use of eye contact and facial expression. Autism affects how young children move and respond in the environment, how they develop interpersonal sympathy and collaborative skills, interfering with cultural learning and with a shared common sense of the world (1).

To figure out how this core disturbance become evident during the first year of life in children with autism is a priority in the forefront of the scientific community because of the substantial impact on long-term clinical outcome of early intervention programs. Moreover, early detection would enable us to characterize the infancy of children with autism, broadening our understanding of autistic (2). However, the earliest stages and the developmental pathways of the social deficit in early autism is still unclear.

Home-Movies (HM), made by parents before child's ASD is diagnosed, represent an ecological tool to study the earliest stages of children with autism and of their interaction with the environment. The retrospective analysis through HM represents, therefore, an excellent option for providing direct observation of social behavior in children with autism (3, 4, 5, 6).

In a precedent study, the evaluation of social behaviors through the technique of HM analysis in a group of subjects aged 0-18 months has highlighted the possibility to reveal several indicator of risk for an early ASD. These indicators concern eye contact, affective states, attention and communication. The hypothesis is that these markers represent a precursor of the deficit in joint attention, the core of the ASD social deficit, hence the need for developing tools for an early screening for identifying children at-risk (4).

2 Objectives and hypothesis

The aim of this study is twofold: (a) to provide insight regarding the influence of ASD on the development of social behaviors within parent-infant cycle of interaction; (b) to address methodological considerations of retrospective video-analysis technique and a computer-based coding program (The Observer 4.0®) (7).

To these aims, we've conducted a microanalysis of HM on the base of a new methodology already tested in other researches (4, 5, 6).

The research hypothesis is that in ASD infants the presence of dysfunctional social behaviors may represents a marker of innate deficit of autism, that may be predictive of further impairments of other inter-subjectivity abilities (imitation, joint attention, gaze following).

3 Method

3.1 Participants

We studied HM from the first 18 months of life of three groups of children. 1) 12 children with a diagnosis of ASD; 2) 9 children labeled as mental delayed; 3) 12 children typically developing.

1. The children of the first group (experimental sample) were diagnosed with ASD through the ADI-R and according to the DSM-IV expressed separately by 2 expert clinicians. In addiction, the investigators used the Children Autism Rating Scale (CARS) to access severity of autistic symptoms (those with scores below 30 were excluded).

2. The children with developmental delays (1st control sample) had documented global cryptogenic mental delays, diagnosed according to the DSM-IV by 2 expert clinicians. Those with delay secondary to neuropsychiatric syndromes (X-Fragile, Rett or Down Syndrome, etc.) or to evident neurological deficits were excluded.

The children of the first two groups were recruited among those referred from multiple community sources to the Division of Child and Adolescent Neuropsychiatry.

3. The parents of typically developing children (2nd control sample), recruited among those attending a local kindergarten, reported no history of developmental problems.

3.2 HM organization and editing procedures

HM were collected from parents of all children. After a complete description of the study to the families a written informed consent were obtained.

The video material were organized in three range of age (0-6 months (T1), 6-12 months (T2), and 12-18 months (T3). An editor, blind to the diagnoses of the children, selected for each HM all segments (standard situations) where the infant was visible and involved in human interaction. Chosen segments were comparable across groups with respect to the type, length and number of standard situation, for T1, T2 and T3.

For all the three groups will be included only videos without any editing by the parents and running for a minimum of 10 minutes for each period.

3.3 Infant Behavioral Scale

On the basis of the instruments utilized in previous researches (4, 5, 6), a scale for the detection of social behavior in the first 18 months of life were introduced to study the early interaction between infants and their caregivers and to analyze the distribution of the child's spontaneous attention to social and non social stimuli.

The IBS consists of 22 items describing the behavioral components of interactions between the infant and his/her caregivers. The items are divided in 4 sub-scale:

1) *The Grid of Basic Behaviors* is composed by 6 items (Looking, Orienting, Gaze following, Smiling, Seeking conctat and Explorative activity). This sub-scale consists of a repertoire of typical motor activity that the infant uses to know, to explore and to interact with the environment.

2) *The Grid of Complex Behaviors* is composed by 11 items (Enjoying, Sintony, Anticipation of other's intention, Communicative gestures, Referential Gaze, Soliciting, Accepting Invitation, Offering himself to another person, Imitation, Pointing, Maintaining social engagement). This sub-scale consists of a number of complex behaviors that the child display within the interactions with the caregivers. The items represent inter-subjective abilities that arise in the first 18 months of life.

3) *Grid of Vocalizations* is composed by 3 items (Simple vocalization, Meaningful vocalization, Crying). The items of this sub-scale regard the vocal activity of the infants.

4) *Grid of Caregiver's Behaviors* is composed by a number of items categorized in two general area (a. Regulation up and down; b. Stimulation by requesting behavior, vocalizing, naming him/her, gesturing, touching him/her, showing him object). This sub-scale consists of behaviors that the caregiver do within a sequence of interaction for stimulate the child or for regulate his/her affective states.

The items are also divided in two type: state (to code the start and the end of behavior) and event (to code the moment in which appears).

The items have been selected on the basis of previous studies and of the expected social competencies for the age.

3.4 Computer-based coding system

The Observer 4.0® was configured for the application of the IBS. The four sub-scale was set as behavioral class and each target behavior is assigned a unique key.

3.5 Coders training and inter-rater reliability

Four raters were trained to use the computer-based coding system by viewing and coding several sample standard situations of both pathological and typical developing children, to familiarize with type and quality of the videos and with the meaning of the items, to acquire competence in coding procedures and to learn to correctly identify the items.

Raters were blind to the diagnosis of infants.

Inter-rater reliability was calculated by The Observer 4.0®. Raters were required to achieve a satisfactory agreement (Cohen's Kappa >/= 0,7) with an expert clinicians in two matched sequences for each range of age.

The tolerance window regarding time discrepancy between codes was set at 1 second for event-behaviors and at 3 seconds for state-behavior. The same code recorded outside of this window was reported as a coding error and was considered a disagreement.

3.6 Coding procedures

The HM of the 3 groups were mixed, distributed on four CD and rated by the trained observers. The coders were blind to the diagnosis of infants. Each segments were rated by one coders; 25% of sequences were rated by two coders for the verification of inter-rater agreement.

4 Results

To investigate how the children of the three group may differ in their early interaction with the caregivers the data were examined in several ways.

Using The Observer 4.0® tools for analysis, descriptive statistic and lag sequential analyses were applied to identify specific behavioral profiles, related to the hypothesis of the study.

a) Infant's behavior were studied through the following statistical measures: 1) total number of occurrence of state or event behavior in each sequence; 2) latency (time lag) from a caregiver's solicitation to an infant's related behavior occurrence; 3) mean number of occurrence of state or event behavior per minute; 4) total and percentage duration (time between the start and the end of a behavior) for each state behavior; mean duration and standard deviation for each state behavior.

b) Interaction between infant and caregiver will be studied through lag sequential analysis (LSA) technique. The LSA will be used for calculating frequency and type of infant's behavioral response to the caregiver's stimulation in a certain time lag (LSA time-based). The LSA were also used for determining the characteristics of behavioral turns (how the infant's behavior is associated with different caregiver's stimulations) and for revealing the temporal differences in the infant-caregiver turns among the three groups.

Preliminary data show several significant differences in social competencies development at T1, T2, and T3. These differences concern both the frequencies and the mean duration of basic and complex behaviors (i.e.: looking at people, orienting to name, social engagement), and their relation with care-giver stimulation.

More detailed data on differences among the three group and on use of The Observer 4.0® will be given and discussed.

References:

1. Frith, U. (1991). *Autism and Asperger Syndrome*. Cambridge: Cambridge University Press.
2. Trevarthen, C.; Stuart, D. (2005), Rhythm and synchrony in early acting and communicating and signs of Autism and Rett Syndrome in Infancy. To be published in 2005 as a Supplement of *Brain and Develpment*.
3. Baranek, G.T.; Barnett, CR; Adams, E.M.; Wolcott, N.A.; Watson, L.R.; Crais, E.R. (2005). Object play in infants with autism: methodological issues in retrospective video-analysis. *The American Journal of Occupational Therapy*. 59(1):20-30.
4. Maestro, S.; Muratori, F.; Cesari, A.; Cavallaro, M.C.; Paziente, A.; Pecini, C.; Grassi, C.; Manfredi, A.; Sommario, C. (2005) Course of autism signs in the first year of life. *Psychopathology*. 38:26-31
5. Maestro, S.; Muratori, F.; Cavallaro, M.C.; et al. (2002). Attentional skills during the first 6 months of age in Autism Spectrum Disorder. *J Am Acad Child Adolesc Psychiatry*. 41:1239-1245
6. Maestro, S.; Muratori, F.; Barbieri, F.; et al. (2001). Early Behavioral Development in Autistic Children: The First 2 Years of Life through Home Movies. *Psychopathology*. 34(3):147-52.
7. Noldus Information Technology. (2002). *The Observer®. Reference Manual 4.1*. Wageningen.

A descriptive analysis of social referencing using The Observer® 5.0 software among preschool children

P. Dhingra[1], P. Verma[1], U. Dhingra[2] , A. Dutta[2] , R. E. Black [2], A. Sarkar[1], M. Osmany[1], C. P. Khokhar[3], R. Juyal[1,3], V.P. Menon[1], G. S. Hiremath[2], S. Sazawal[1,2], M. Black[4], B. Lozoff[5]

[1] Center for Micronutrient Research, Annamalai University, Delhi, India, [2] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore USA, [3] Department of Psychology, Gurukul Kangri University, Haridwar, India [4] Department of Pediatrics, University of Maryland, USA, [5] Center for Human Growth and Development, University of Michigan, Michigan, USA.

Abstract

Studies so far, have employed manual or video recording of observations of social referencing skills of children that have limitations due to observational errors and inconsistencies. We designed a social referencing project (CMROBS) using The Observer® 5.0 software (Noldus Information Technology) installed in a lap-top to observe and code the child's social referencing skills and parent child behavior during the session. Social referencing assessments were conducted as continuous behavioral observations for 11 minutes, among 1240 preschool children, in a playroom clinic setting. Familiarization toys for first 5 minutes and 3 different ambiguous stimulus toys for 2 minutes each were given to the child for exploration. Components of social referencing behaviors were recorded as child's social referencing skills to Mother; Contact with Caregiver; Proximity to Mother; Mother's Affect; Child's Affect; Latency to first touch to Familiarization/Stimulus toy; Contact with Toy and Distance to Stimulus toy. Results suggest that The Observer® 5.0 software was successfully used in systematic/error free recording and analysis of social referencing skills among children.

Keywords

Social referencing, The Observer software, preschool children, structured observation.

1 Introduction

Social referencing is a measure of social and emotional development wherein children use expressions from caregiver to appraise events and regulate their behavior [1, 2, 3]. This is a secondary appraisal process and operates primarily under conditions of uncertainty and novel conditions, where infants are expected to reference their mothers as a means of appraising the event. Referencing skills are generally measured in a laboratory playroom setting where the mother is present and the child is provided with stimuli specifically selected to be compelling enough to generate uncertainty [4]. Most of the studies conducted on assessing the social referencing skills of children have used manual or video recording method of observation. These methods of recording have limitations due to inconsistency, difficulty in recording duration of behavioral observations, complexity in analyzing and interpreting findings. Currently, observational softwares have gained popularity due to ease of recording, storing, collating, and analysis of complex behavioral observations and reducing data entry errors. The Observer® 5 is a widely used software [5] which provides flexible, computer-based coding system to observe the complex behavioral processes, in children and has data analytic features. We evaluated the social referencing skills in a large sample of developing country preschool children using CMROBS in The Observer® 5.0 (Noldus Information Technology).

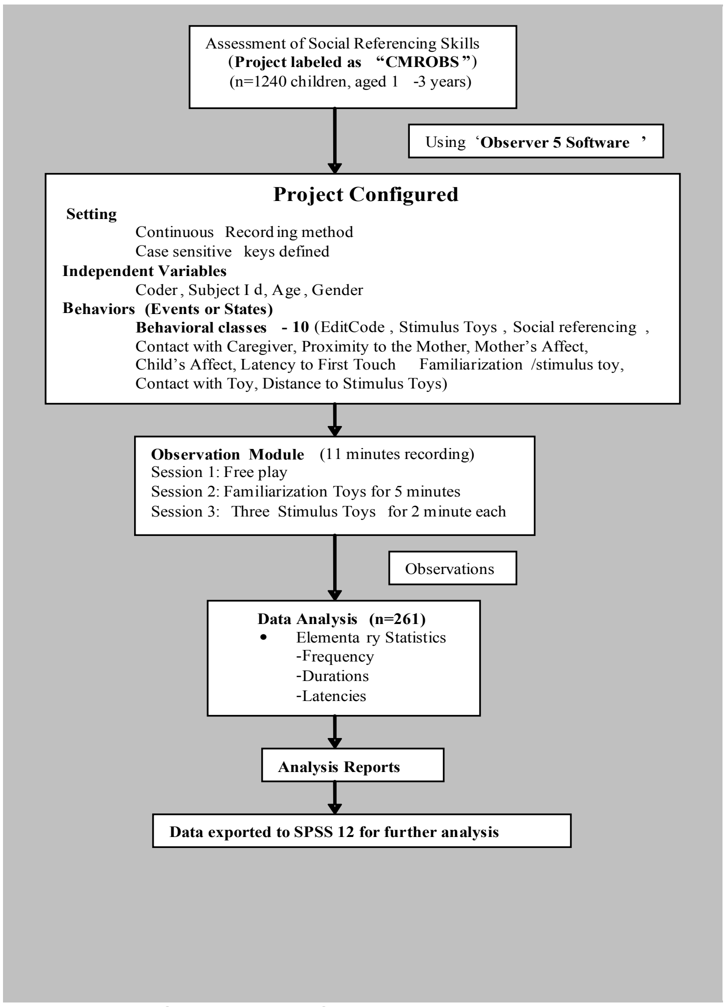
2 Methods

2.1 Project Design

We designed a social referencing project (CMROBS) in The Observer® 5.0 software (Noldus Information Technology) to observe and code the child's social referencing skills and parent child behavior during the session (Figure 1). Configuration of the project included setting, Independent variables, subjects, behavior, and behavior classes which were specified in a spread view. We defined the observation method as continuous and also specified the case sensitive keys. Independent variables were defined as child's identification information, age and gender. Based on literature, 64 behavioral elements of interest were identified and classified within 10 social referencing constructs. These were **Edit code** (mistake as an event), **Stimulus Toys** (freeplay and 3 stimulus toys recorded as events); **Child's Social Referencing to Mother/Tester/ Student** (Child Look, Show, Offer, Verbal, Bid as events); **Contact with Caregiver** (degree of closeness to mother recorded as four states recording duration and one momentary event as touching hands); **Proximity of child to Mother** (position of the child in comparison to mother recorded as state); **Mother's Affect** (recorded as Mother talk/ no talk as durational states and positive/negative initiation/ response as momentary events); **Child's Affect** (mood/ general state of the child recoded as duration); **Latency to first touch to toys** which recorded the duration from the presentation of the toy to the first time the child touched the toy; **Contact with Toys** (Stimulus toy, Familiarization toy or No Contact with toy as durational states); **Distance to Stimulus toy** (closeness to the toy recoded as durational states). Channels were used during the event recording and defined as proximity of the child to the mother, child's affect, mother's affect, distance to stimulus toy, contact with the caregiver, contact with the toy. During the social referencing observation, each channel box showed the currently active behavior from a particular behavioral class except social referencing. In the project we have not used any modifiers. We defined behavior as momentary/event if it last just a moment and noted only its occurrence/time of occurrence. By contrast, state behavior was defined as durational, and recorded as how long they last, typically recording times of onset as well as times of offset.

When the configuration was completed, it was reviewed, checked for consistency and redundancy and the configuration got locked after the first observation was stored.

Figure 1. Observation of social referencing using The Observer® 5.0 software.



2.2 Subjects and Settings

The study was conducted among 1240 preschool children (aged 1-3 years) in periurban community. Assessments were performed in a playroom field based clinic setting of the Center for Micronutrient Research, New Delhi. This study was a component of a randomized controlled trial with 4 arms (2 interventions)

2.3 Social Referencing Assessment Procedure

Social referencing assessments were done as continuous behavioral observations for 11 minutes in three divided sessions during the day time when the child was awake. Child and mother were placed on a floor mat (2m X 2m) comprised of 16 squares of equal size (0.5m X 0.5m), with toys (Figure 2).

Error!

Figure 2. The examiner is coding the social referencing done by child in 11 minute observation session.

During the observation session, mothers were instructed to remain seated and to refrain from initiating interaction with their child but to respond the child only if he asked questions or requested for help. Before the start of actual observation, the child was initially placed two squares away from the mother and was free to play with the free play toy for 1 minute to get acquainted with the setting. During the session, four familiarization toys (Whistling Toad, Four Red Cubes, Stacking Rings, and Toy Car) were placed on the mat and the child was allowed to explore for 5 minutes. Next, the first ambiguous stimulus toy (Piggy Tail Elephant) was presented, placed one square apart from the child. Followed by that the other two stimulus toys (Bunny on a Slide, Helicopter) were presented in the same manner. Each stimulus toy was given for 2 minutes duration.

2.4 Training and Reliability

Careful training of observers, and establishing their reliability, is an important part of data collection. Examiners were trained to record observations using CMROBS installed in lap-top from videotapes of social referencing observations done on sample children. Inter observer agreement was established by comparing records of social referencing observations collected by two observers simultaneously during the 11 minutes session. Training and reliability continued until observers achieved more than 80% agreement with the trainer.

2.5 Data Recording

The observer coded social referencing and mother - child's affect using the lap-top on which CMROBS was installed. Observation module of the software was opened to record the social referencing of a child and all the observations were scored in the Event Log window. Before the start of observation, the independent variables were entered. Initial behavior for each behavioral class (child' affect, mom affect, contact with the Caregiver/toy, proximity to mother and distance to stimulus toy of the dyad during the observation) was recorded. In the screen, there were two columns, one pertaining to subject and other to action which ensured subject wise continuous recording of the behavior against time. Only one of the behaviors could be active at any given point of time for a particular behavioral class. We used continuous live behavioral observations for 11 minutes with transition-activated recording because observers remain continuously alert but record only when a transition occurs, that is, when one state ends and another begins, or when a momentary behavior occurs. During the observation session, observer coded social referencing skills of the child and parent child behavior as and when the behaviors occurred. The Observer® 5 software, automatically recorded the time (to 1-second accuracy) against each record/code such that the computer program organized the records and their associated times into computer files. At the end of observation period, stop button was pressed. Observers can view the previously recorded information, and can edit those files in the event log window if, on reflection and reviewing, there was some incorrect coding that required to be changed. The data files were stored as Observational Data Files (ODF) automatically and were retrieved for analysis.

2.6 Data analysis

For the present analysis, children in one of the four arms were used. Data of 216 observations was selected and data profile was created. The frequency and duration of social referencing was carried out using the elementary statistics of The Observer® software. For the analysis, all the observations were truncated to 11 minutes. Data analysis results for each child were displayed in both report and matrix form as tables. For momentary codes, *frequencies* (how often the behaviors were displayed by children during the session). For duration codes, *average durations* (how long, on average, did the behaviors last) were calculated. Latencies to first touch to familiarization or stimulus toys (duration from the presentation to the first time that the child touched the toy) were also calculated.

Some of the elements of behavioral class (BC) were combined [**Child's affect** (BC): big smile and slight smile; **Contact with Caregiver** (BC): within mother's arm length and clinging or close, sitting on lap for the purpose of the analysis.

For further analysis, the analysis results files from The Observer ® were exported to excel and then to SPSS (version 12). The analysis provided mean frequency (how often specific behaviors occurred), duration (what percentage or proportion of time behaviors occurred) and latencies information. This was done as an example. The data set could now be merged with other data for more detailed analysis.

3 Results

Descriptive analyses of social referencing on 216 children are presented. The average duration of observations was found to be 678 seconds. During the 11 minutes observation session, children spent 78 % of time in playing with the familiarization and stimulus toys. Children spent more time with familiarization toys as compared to the stimulus toys (Table 3.1). Of all the social referencing skills used by the children during the session, child looking to mother was exhibited more often than child bid/show/verbal/offer mother (Table 3.2).

Table 3.1. *Mean frequency and duration of active exploration among preschool children.*

Variables	**Children (N=216)**
Total Duration of assessment	678.1 ± 11.3#
Total Play	527.8 ± 132.5#
Unengaged	33.16± 89.6#
Active Exploration	
Contact with Familiarization Toy	6.0 ± 3.6* 288.7 ± 114.8#
Contact with Stimulus Toy	4.1 ± 2.1* 200.9 ± 101.2#

*# Duration in sec (Mean ± SD), * Frequency (Mean ± SD)*

The constructs of child's affect was measured by hesitant behavior of the child, wariness of the child and positive affect displayed by the child (Table 3.3).

Table 3.2. *Social referencing skills observed among preschool children.*

Variables	**Children (N=216)**
Child Looking to Mother	5.75 ± 5.78
Child Offering to Mother	1.59 ± 2.75
Child Biding to Mother	0.37 ± 0.90
Child Showing toy to Mother	0.98 ± 1.55
Child Verbalizing with Mother	1.62 ± 2.59

Frequency (Mean ± SD)

Table 3.3. *Measuring child's affect during social referencing observations.*

Variables	**Children (N=216)**
Hesitant Behavior measured as Latency to first touch (duration from the presentation to the first time that the child touched the toy)	
Familiarization toy	24.5 ± 57.1*
Stimulus Toy	323.6 ± 46.1*
Wariness of the child	
Duration of Cry in the session	11.99± 48.64#
Duration of Wary behavior in the session	34.49± 106.57#
Duration of the time the child Clinged Close or remained within arm length of mother	488.2 ± 219.7#
Positive affect of the Child	
Duration of Smile	80.83± 108.72#
Duration of Neutral State	517.63± 173.32#

*Mean ± SD in sec ** Frequency (Mean ± SD), # Duration in sec (Mean ± SD)*

Table 3.4. *Measuring child's proximity to mother during social referencing observations.*

Variables	**Children (N=216)**
Within Mother's Square	243.9± 288.0#
2 Square Apart	436.9± 289.4
3 Square Apart	2.74 ± 20.8

 # Duration in sec (Mean ± SD

The child was mostly in a neutral state and smiled on an average of 4times during the session. Data on mother and child behaviors which included child's proximity to mother showed that most of the time the child was 2 squares apart from the mother (Table 3.4). Mother's affect measured during the entire session indicated that mother did not talk much to the child but initiated positive action and responded with positive comments in response to child's referencing (Table 3.5).

Table 3.5. *Measuring mother's affect during social referencing observations .*

Variables	**Children (N=216)**
Mom not talking	1.24 ±0.72** 671.27 ±31.88***
Mom talking	0.24 ±0.72 6.85 ±29.78
Mom initiate –ve action	0.03±0.31*
Mom initiate +ve action	5.30 ± 6.16*
Mom responds with –ve comment	0.01 ± 0.09*
Mom responds with +ve comment	5.24± 5.17*

*** Frequency (Mean ± SD), *** Duration (Mean ± SD)*

4 Conclusions

This study found that The Observer® 5.0 used on lap-top was useful in systematic and error free recording of the multiple events involved in collecting observational data on social referencing of children in a developing country community based setting. The data entry step is saved recorded and analysis functions produce statistical outputs including independent group comparisons to answer specific research questions. The backup project function of The Observer® 5.0 software is equipped to save the project (configuration, data files, analysis results) in just one file which can be retrieved easily for analytical purposes.

This research was supported by the grants from the New Zealand Milk Limited, Private Bag No. -12032, Auckland, New Zealand.

References:

1. Hornik, R., Gunnar, M.R. (1988). A descriptive analysis of infant social referencing. *Child Dev*., **59**, 626-634.
2. Feinman, S (1985). Emotional expression, social referencing and preparedness for learning in infancy: Mother knows best-but sometimes I know better. In: G. Zivan (Ed.), *The development of expressive behavior: biology-environment interactions*, New York: Academic Press, pp. 291-318.
3. Walden, T.A.; Ogan, T.A. (1988). The development of social referencing. *Child Dev*., **59**(5), 1230-1240.
4. Walker-Andrews, A.S. (1998). Emotions and social development: Infants' recognition of emotions in others. *Pediatrics*, **102**, 5 Suppl E, 1268- 1271.
5. The Observer®, Version 5 (2003). Professional System for Collection, Analysis, Presentation and Management of Observational Data, Noldus Information Technology, Wageningen, The Netherlands.

Using laptop based The Observer® 5.0 and Palm hand-held based Visual Basic (Appforge Add-On) software to assess social referencing skills among young children

P Verma[1], A Dutta[2], U Dhingra[2], B Lozoff [3], R M Angulo-Kinzler[4], VP Menon[1] , P Dhingra[1], A Sarkar[1], R E Black [2], M Osmany[1], R Juyal[1], S Sazawal[2], J Kumar[1], G S Hiremath[2]
[1] Center for Micronutrient Research, Annamalai University, Delhi, India, [2] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore USA.[3] Center for Human Growth and Development, University of Michigan, Michigan, USA; [4] Division of Kinesiology, Center for Human Motor Research, University of Michigan, Michigan USA.

Abstract
With the recent advent and increasing use of observational software, recording, storing and analysis of complex behavioral data have become efficient. We compared social referencing skills using The Observer® 5.0 software installed on laptop (CMROBS) and palm handled based visual basic software (CMRDEE) with multiple screens for recording of behaviors. Simultaneous assessments using laptop and palm handheld device were done for 11 minutes among 30 preschool children. Across all observed skills, score on The Observer® 5.0 (CMROBS) was higher than palm handheld software (CMRDEE) but the differences were not significant. The results suggest that CMRDEE installed in palm handheld device is a cheaper than and as efficient as CMROBS installed on laptop for assessing social referencing skills among children and can be used for behavioral studies involving continuous complex observations.

Keywords
Social referencing, CMROBS, CMRDEE, The Observer® 5.0, Palm handheld, Preschool children.

Introduction
Social referencing is a process of emotional communication in which the child uses mother's interpretation of the events to form an understanding of that event. Social referencing is hypothesized to operate primarily under conditions of uncertainty and requires children to have certain basic interpretive abilities to decode the socio emotional messages and relating them to appropriate objects and events. Given the fact that young children encounter a perceptual stream of novel or otherwise ambiguous events, social referencing, a secondary appraisal process, may play a significant role in their development [1]. Observational methods can be used for quantifying social referencing behaviors, activities and other physical aspects of the mother- child interaction in detail [2].

Traditionally, manual recording these observations were conducted using observation guides, recording sheets, checklists, field notes, pictures, or a combination of these methods. Recently, automated systems for collecting continuous observations of behavioral patterns have been developed to overcome the time consuming, labor intensive, expensive and error prone manual methods.

The Observer® 5.0 (Noldus Information Technology, Netherlands) [3] is a widely used software for the collection, management, analysis and presentation of observational data. We have previously developed a project (CMROBS) for recording social referencing in The Observer® 5.0 software. With improved versatility and reduced cost, the palm handheld device provides an attractive alternative for recording observational data. It has the benefit of low cost, ability to be mobile but may lead to loss of data in continuous observations due to small screen. We developed a social referencing recording system in palm handheld device and undertook the present study to compare CMROBS installed in laptop with CMRDEE installed on palm handheld device.

Methodology

Subjects and setting
This study was designed as add on to an ongoing clinical trial [4] where social referencing was already been done using CMROBS software developed in The Observer® 5.0 and installed on laptop. In a structured setting, overt observations of social referencing were carried out among 30 children aged 2- 4 years for 11 minutes, simultaneously using CMROBS and CMRDEE software. Social referencing observations were carried out in the development clinic of the Center for Micronutrient Research which was staffed with psychologists and child development specialists [Figure 1].

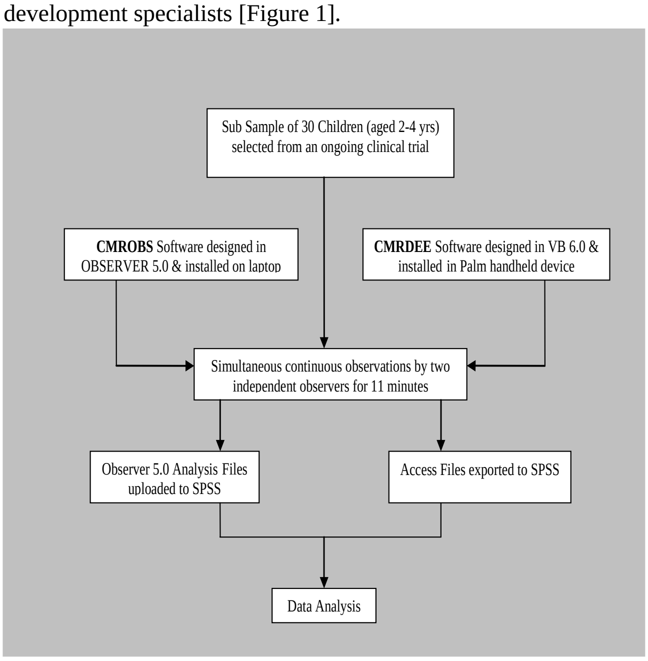


Figure 1. *Flowchart depicting the methodology adopted for comparing the social referencing skills using The Observer 5 and Palm handheld device.*

Stimulus toys used for social referencing observations

The study toys consisted of 1 free play toy, 4 familiarization toys and 3 stimulus toys. The free play toy was a 'pegboard' in which the child had to put all the pegs in the board. The familiarization toys included a 'piggy tail elephant' which needed to be keyed manually to take round turns, 'stacking barrels' that had 5 take apart barrels that could nest and stack, 'arranging rings' that required the child to stack rings of different sizes on a rod and lastly, a 'box' which had to be closed with a lid. The first stimulus toy was a 'Mini roll along puzzle train toy' which had detachable bogies and attractive Disney character passengers that could be manipulated by the child. The second stimulus toy was 'Block train with an engine driver'. The third stimulus toy was a 'Baby smartronics cookies shape surprise' which could teach shapes, numbers, sorting skills, entertained with music, fun phrases and had sounds and lights.

These three stimulus toys were selected following pre testing of a number of different age appropriate toys with children. All these stimulus toys met our criteria of "ambiguous stimulus" for the present study.

Observation procedure

The social referencing observations were carried out on a carpet (2m × 2m) which comprised of 16 squares (0.5m × 0.5m each). At the start of the observation, the child was placed two squares apart from the mother. The experimenter gave the uniform instructions to all the mothers before the start of observation, "We want to watch your child play with several toys. Please do not talk to your child unless he asks you questions or requests help". The entire 11 minutes assessment was continuous divided into three sessions named as free play, familiarization toy session for 5 minutes and stimulus toy session for 6 minutes. First of all, a short free play time was given to the children to help them develop some familiarity with the setting. It was followed by the second session during which all the familiarization toys were placed on the carpet at the same time and the child was allowed to explore the toys. Then, the stimulus toys were presented (each for duration of 2 minutes) in a box and were placed one square apart from the child and mother [Figure 2]. These stimulus toys were given one at a time.

Figure 2. *Recording the child and mother interaction (social referencing) in a structured play setting.*

Training and reliability

For both CMROBS and CMRDEE, training exercises were undertaken for two observers using video recorded social referencing observations on a pilot sample. Training was continued till there more than 80 % agreement for each device between the two observers.

Social referencing constructs and codes

Based on literature, social referencing constructs were identified and classified within 10 behavior classes. These were namely edit code (mistake as an event), stimulus toys (free play and 3 stimulus toys as 4 event codes), social referencing to mother, tester, student (Look, Show, Offer, Verbal, Bid as 19 event codes), contact with caregiver (degree of closeness to mother as 4 durational codes and 1event code), proximity with mother (closeness of child to the mother as 5 durational codes), mom affect (mother talk/ no talk as 2 durational codes and positive/negative initiation/ response as 4 event codes), child affect (mood/ general state of the child as 6 durational codes), latency to first touch (familiarization and stimulus toys as 2 event codes), contact with toy (stimulus toy, familiarization toy, no contact or multiple contact with toy as 4 durational codes), child's distance to stimulus toy (proximity to the stimulus toy as 5 durational codes). Abbreviated codes were developed for the recording of observations and the data was stored in the form of short sentences representing these codes.

CMROBS (The Observer® 5.0 installed on Laptop)

Social referencing project (CMROBS) was developed using the above mentioned constructs and codes in The Observer® 5.0 software and installed in laptop. Configuration of the project included setting, subjects, behavior and behavior classes which were specified in a spread view by which examiner could make sure that only one of those behaviors was active at a given time during the observation. The coder, child identification number, age and gender were short listed as the independent variables for this project. In the screen, there were two columns, one pertaining to subject and other to action which ensured subject wise continuous recording of the behavior. Behaviors that can occur simultaneously were placed in two different behavior classes. It was a continuous method of recording, case sensitive and had a fixed duration of 11 minutes observation. No modifiers were defined for the present project. In this software, each record was attached to a time code, so that computer could produce information about the location in the time line and the duration of any action being recorded.

At the start of the observation, the examiner entered the identification information of the child. Initial behavior for behavioral classes (proximity to mother, child's affect, mother's affect, distance to stimulus toy, contact with toy, contact with caregiver) was also recorded at the beginning of the observation session. As soon as the first familiarization toy was presented to the child, the examiner pressed the start key. During observations, the specific abbreviated key codes were pressed to record the mother and child behavior. At the end of observation, stop button on the toolbar was clicked. After the observation was complete, the data was stored as an ODF file which could then be retrieved and analyzed.

CMRDEE installed on Palm handheld

The same constructs were used to develop a social referencing project using Visual Basic 6.0 based software installed on palm. Visual basic program was converted to Palm operating system using AppForge Crossfire™ [5] and was uploaded to the Palm ™ m 125 series Handhelds [6] to record the social referencing data in a palm database format. The configuration ensured a pre- loaded coding scheme in order to recognize the input during observations. The software has multiple user interface screens. The main screen has the list of all 10 behavioral classes of social referencing and upon selection each screen displayed a set of behaviors for a particular behavioral class in the form of short sentences. Only one behavior from a particular behavior class could be active at any given point of time of observation. Alarm was set at 11 minutes to help the examiner to stop coding the behaviors at the end of the session.

The software was field tested and necessary modifications were done to adjust the look and feel of the recording screens (adjusting the size of the screen areas, adjusting the size of the characters, how the codes appear during scoring) to help the examiner use the software efficiently.

Identification information of scheduled children was uploaded into the palm database using custom designed software and the desired child was selected from the list of children at the time of observation.
From the applications launcher of the palm handheld, "CMRDEE" project had to be clicked for observing the social referencing skills of children live. The values for the independent variables were entered before the start of the observation. Initial behavior for each behavioral class was also recorded at the beginning of the observation session. The examiner needed to select the desired window for a behavioral class which represented a set of choices and upon selection of a particular behavior, the referencing behaviors were recorded against time into the database.

Data was downloaded in the personal computer at the end of the day by performing the HotSync operation facility provided by the Palm software and identification data for next scheduled children were uploaded at the base station located in the central office. A Separate desktop based software was also designed to view and edit the data collected using "CMRDEE". An in-house designed software was used to convert the palm database (PDB) files to populate an access database which could be read into SPSS 12.0 for analysis.
Time and date of both the laptop and palm handheld was set and checked daily for any discrepancy in the two observation systems.

Outcome measures and analysis

The outcome measures included the social referencing skills, child's and mother's affect, child's contact with stimulus and familiarization toys and child's proximity to mother.

Using the elementary statistic analysis tool of The Observer® 5.0 software, frequencies were calculated. Analysis files were exported to SPSS 12.0 and the proportion of children who displayed the social referencing and other mother child behaviors were calculated for both the methods.

Results

As compared to CMROBS, CMRDEE consistently recorded a lower proportion of children who displayed social referencing skills (Table 1). Similar trends were observed for scores of mother's affect, child's affect, child's contact with stimulus and familiarization toys and child's proximity to mother (Table 2 & 3). However, these differences were non significant.

Table 1. *Comparison of child's social referencing scores by using CMROBS installed on laptop and CMRDEE installed on palm handled device.*

Variables	CMROBS (n=30)	CMRDEE (n=30)
Child Shows Toys to Mother	14 (46.7)*	12 (40.0)
Child Verbalizes to Mother	18 (60.0)	12 (40.0)
Child Looks to Mother	28 (93.3)	24 (80.0)
Child Offers to Mother	7 (23.3)	5 (16.67)
Child Bids to Mother	3 (10.0)	3 (10.0)

**Proportion of children [Frequency (%)]*

Table 2. *Comparison of scores of child's affect and contact with toys by using CMROBS installed on laptop and CMRDEE installed on palm handled device*

Variables	CMROBS (n=30)	CMRDEE (n=30)
Child Cry	2 (6.7)*	1 (3.3)
Child Smile	20 (66.7)	17 (56.7)
Neutral Child	29 (96.7)	24 (80.0)
Wary Child	4 (13.3)	3 (10.0)
Child's Contact with Familiarization Toys	30 (100.0)	26 (86.7)
Child's Contact with Stimulus Toys	30 (100.0)	26 (86.7)

**Proportion of children [Frequency (%)]*

Table 3. *Comparison of child's and mother's scores by using CMROBS installed on laptop and CMRDEE installed on palm handled device.*

Variables	CMROBS (n=30)	CMRDEE (n=30)
Child clinging close to mother	3 (10.0)*	2 (6.67)
Child within arm length of mother	20 (66.7)	19 (63.3)
Child 1 square apart from mother	15 (50.0)	13 (43.3)
Child 2 square apart from mother	28 (93.3)	21 (70.0)
Mother not talking	30 (100.0)	25 (83.3)
Mother initiate positively	23 (76.7)	20 (66.7)
Mother respond positively	25 (83.3)	18 (60.0)

**Proportion of children [Frequency (%)]*

Conclusion

The results suggest that CMRDEE designed in VB 6.0 and installed in palm handheld device could successfully record the social referencing observations and performed as good as CMROBS designed in The Observer® 5.0 and installed on laptop.

The palm device is more convenient as it is simpler to operate, is portable, can be used independently without the need of a desktop computer in the field setting and provides mobility to the examiner to code the behaviors accurately. Furthermore, uploading the software and data transfer between the palm and PC is much easier. The CMRDEE project also enables the examiner to upload all the identification information of the children planned to be observed in the palm a day before and this way saves the time in the actual field set up. With the lithium- ion battery, palm handheld can record observations up to 2 hours at a stretch. The display of behavioral elements on the screen makes the palm device more user friendly.

A lag in recording the continuous and complex behavior observations due to small screen of the palm may have resulted in a minor loss of data, but the multiple advantages of palm make it a low cost preferable alternative in assessing the social referencing skills among young children.

Supported by New Zealand Milk Limited, Private Bag 12032, Auckland, New Zealand.

References:

1. Walden, T.A.; Ogan, T.A. (1988). The Development of Social Referencing. *Child Development*, **59**, 1230-1240.
2. Mumme, D.L; Fernald, A.; Herrera, C. (1996). Infant’s responses to facial and vocal emotional signals in a social referencing paradigm. *Child development*, **67**, 3219- 3237.
3. The Observer®, Version 5 (2003). Professional System for Collection, Analysis, Presentation and Management of Observational Data, Noldus Information Technology, Wageningen, The Netherlands.
4. Juyal R.; Osmamy M.; Black R.E.; Dhingra, U.; Sarkar A.; Dhingra P.; Verma P.; Marwah, D.; Saxsena R.; Menon V.P.; Sazawal, S. (2004) Efficacy of micronutrient fortification of milk on morbidity in pre-school children and growth - a double blind randomised controlled trial. *Asia Pac J Clin Nutr*. Aug; **13** (Suppl):S44.
5. AppForge Crossfire™ , Version 5.1 (2004). AppForge Extends and Simplifies Microsoft® .NET Development to All Mobile and Wireless Devices, AppForge, Inc., Georgia, USA.
6. Palm™ , Inc. (2001). California, USA.

Measuring nighttime parenting behavior using overnight infrared video recordings of mother-infant dyads

L.E. Volpe
and the Centers for the Prevention of Child Neglect
Department of Anthropology, University of Notre Dame, Notre Dame, IN, USA

Abstract

The present study explores sleeping and feeding patterns of 24 adolescent and 15 adult first-time mothers and their infants. The sample was drawn from a larger longitudinal study on transition to parenting across the first three years. When infants were four months of age, overnight infrared video recordings of nighttime parenting behavior were obtained at the Mother-Baby Behavioral Sleep Laboratory. Participants were allowed to maintain typical sleeping arrangements and routines, and performed all caregiver interventions at will. Videotapes were coded for infant feeding method, sleep location, presence of sleep-related risk factors, amount of physical contact, and mother-infant interactions using Observer 5.0. Sleep lab data were analyzed for group differences and were related to prenatal cognitive readiness to parent, daytime measures of parenting behavior, and child outcomes at one year of age. Implications for child development and for clinicians designing parenting intervention strategies are considered.

Keywords

Nighttime parenting, infant sleep, infant feeding, child development

Introduction

While research has explored the developmental implications of different patterns of infant care, relatively little is known about how parents care for their infants at night. Studies of nighttime parenting have become a topic of inquiry only within the last decade, and substantial data is needed to address the implications of nighttime parenting behavior for child development. Previous studies have shown a significant relationship between different sleeping arrangements and the form and frequency of particular behaviors, including breastfeeding, sleeping position, and presence of risk factors in the sleep environment. The present study attempts to determine whether these differences in nighttime parenting behavior apply to first-time mothers, particularly those in high-risk populations including adolescent mothers. The study explores whether nighttime parenting correlates with measures of caregiving behaviors and activities taken during the day, and examines the implications of particular parenting behaviors for physiological, cognitive, and socioemotional outcomes of children at one, two and three years of age. The current study is engaged in establishing new methods of research into nighttime parenting behavior and developing new taxonomic definitions of child neglect. By identifying and measuring risks as well as protective factors, the current project attempts to measure the multiple social ecologies and life circumstances that shape the parenting styles of adolescent and adult mothers.

Methods

Data on the sleeping and feeding patterns of 24 adolescent and 15 ethnically-matched adult primiparas were drawn from a larger longitudinal study on transition to parenting (N=684). Adolescent mothers were 15-18 years of age at childbirth and adult mothers were 22-37 years of age at childbirth. Participants were 56.4% European American, 43.6% African American, and 1% Latina. 62% of the infants in the sample were female.

Adolescent and adult mothers were brought into the Mother-Baby Behavioral Sleep Laboratory in order to observe nighttime parenting behaviors using overnight infra-red video recordings. The data were collected between July 2002 and May 2005. Sleep lab data for mother-infant dyads were obtained at two time points, once when infants were four months of age and again when infants were eight months of age. Participants were instructed to arrive shortly before they would normally begin preparing their infants for bed, and once in the lab were allowed to maintain typical sleeping arrangements and parenting routines. The lab was designed to resemble a home environment and enabled a variety of possible sleeping arrangements, including a standard double mattress with box springs for mothers or for mothers who bedshared with their infants, a bed-side bassinet for room-sharing, and a standard crib positioned in an adjacent room for those mothers wishing to sleep separately from their infants. Participants had minimal interaction with researchers, and performed all caregiving behavior for their infants at will.

Subjects were monitored using continuous infrared video recording. Infrared cameras were positioned above both the double bed and the crib. Participants were also monitored in a lounge area while watching TV and engaging in pre-sleep interactions with their infants. Mothers and infants sleeping in the separate rooms were monitored simultaneously, and recordings were combined on a split-screen format using a digital mixer [Panasonic Digital AV Mixer WJ-MX20]. Video recordings were reviewed in real time in their entirety. Videotapes were manually coded using Noldus Observer 5.0. Behavioral coding was completed using a taxonomy involving 10 independent variables, 11 behavioral classes and 14 modifier classes. Selected behaviors focused on maternal and infant sleep state, infant feeding method, sleep location, presence of sleep-related risk factors, mother-infant proximity, amount of physical contact between mother-infant dyads, and potential neglect issues. Infra-red video recordings were coded by pairs of researchers using continuous event recording.

Results

Sleep lab data were analyzed for group differences and were related to daytime measures of parenting behavior and infant outcomes, including prenatal cognitive readiness to parent, daytime measures of self-reported parenting behavior, and child outcomes at one year of age. Adult mothers were more likely than adolescents to breastfeed their infants, whereas adolescent mothers were more likely than adults to sleep in bed with their infants. Adolescent mothers were also more likely than adults to place their infants in the unsafe prone sleeping position, which is a risk factor for Sudden Infant Death Syndrome (SIDS). For both adults and adolescents, cognitive readiness to parent, including higher levels of knowledge of infant development, was related to a greater likelihood of breastfeeding and a lower likelihood of placing the infant to sleep in the prone position. Implications for child development and for clinicians designing parenting intervention strategies are considered.

Discussion

Our continued analysis will focus on the larger social ecology of nighttime parenting behavior, the social and biological context within which infant sleep takes place. More data is needed to better understand the personal characteristics and parenting experiences that shape individual mothers' nighttime parenting practices, and how factors like poverty, partner status, and household composition can further impact nighttime parenting behavior.

Finally, this ongoing study may have implications for designing parenting intervention programs. For example, breastfeeding may be a target of early prevention/intervention programs for adolescent mothers. Furthermore, intervention studies may need to include older adult female relatives in the program design for adolescents mothers, because grandmothers or other caregivers often provide a significant amount of childcare for the children of adolescent parents.

Research supported by NICHD HD 39456, Centers for the Prevention of Child Neglect 2001. Travel funding provided by the University of Notre Dame Graduate School.

References:

1. Ball, H. (2003). Breastfeeding, bed-sharing and infant sleep. *Birth* 30 (3), 181-188.
2. Blair, P.; Ball, H. (2004) The prevalence and characteristics associated with parent–infant bed-sharing in England. *Archives of Disease in Childhood* (89), 1106-1110.
3. McKenna, J. (2000). Cultural Influences on Infant and Childhood Sleep Biology, and the Science that Studies It: Toward a More Inclusive Paradigm. *Sleep and Breathing in Children: A Developmental Approach*, G. Loughlin, J. Carroll, and C. Marcus (Eds.). New York: Marcel Dekker, Inc.
4. McKenna, J. et al. (1999). Breastfeeding and Mother-Infant Cosleeping in Relation to SIDS Prevention. *Evolutionary Medicine*, W.R. Trevathan, E.O. Smith and J.J. McKenna, (Eds.). New York: Oxford University Press.
5. Whitman, T.; Borkowski, J.; Keogh, D.; and Weed, K. (2001). *Interwoven Lives: Adolescent Mothers and their Children.* Mahwah, N.J. : Lawrence Erlbaum Associates.

Bridging ethnography and engineering through the graphical language of Petri nets

Erwin R. Boer[1], Deborah Forster[2], Carrie A. Joyce[2], Pierre Fastrez[3], Jean-Baptiste Haué[4], Monal Chokshi[2], Elaine Garvey[2], Tayopa Mogilner[2], Jim Hollan[2]

[1] *LUEBEC, 5434 Renaissance Ave., San Diego, CA 92122, USA*

[2] *Distributed Cognition and HCI Laboratory, Department of Cognitive Science (0515), University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515 USA*

[3] *Département de communication, Université catholique de Louvain, Ruelle de la Lanterne Magique, 14, BE-1348 Louvain-la-Neuve, Belgium*

[4] *Centre de Recherche et d'Appui pour la Formation et ses Technologies (CRAFT) Ecole Polytechnique Fédérale de Lausanne (EPFL) CE 1 530, Station 1, CH - 1015 Lausanne, Switzerland*

Abstract

Both ethnographers and engineers deal with contextual and situational complexity in human behavior. For engineers, this complexity creates challenges in designing systems to support humans that are useful and predictable across a wide range of contexts because of the difficulty to analyze and model the contextual impact on observable behavior. The rich descriptive understanding found in the annotated case studies of ethnography could be of substantial use here, however, the ethnographer's language is considerably different from that of the causal systems and input/output models typical of engineering. Through a unique collaboration between automotive engineers and UCSD ethnographers, we employed a common graphical state-space based modeling language called Petri nets that enables annotated case studies to be represented in a computational framework that can be used in standard engineering practices. The ethnographer's goal was to explain and quantify the importance of context on driver behavior such that the engineers would be better able to design useful human-centered support systems and to assess whether those necessarily practically constrained support systems will most likely function as expected across a wide range of situations. This paper offers a discussion of the issues that arise in bridging the gap between engineering and ethnographic practice in the context of lane change behavior analysis.

Keywords

Driver behavior modeling, lane change behavior, interface design, ethnography, Petri nets.

1 Models

Models are the constrained embodiment of an analyst's understanding about a particular behavior and form a fruitful foundation for support system development. A model is the partial representation of a phenomenon of interest, the shape of which can range from linguistic descriptions on one extreme to logico-mathematical models on the other.

1.1 Purpose of computational models

Computational modeling of driver behavior requires clear mathematical definitions of input, output, and the processing logic that algorithmically transforms input to output. Such models force analysts to explicate the impact of different contextual elements quantitatively.

1.2 Top down and bottom up modeling

Ethnographers allow observable categories and patterns to emerge from their own interaction with the data, while engineers organize their data for model-based hypothesis testing. Engineers start with a model structure (top down) and use the data to identify its structure and associated parameters through a process of minimizing the discrepancy between model-predicted and observed behavior (monitoring, decision or control). Ethnographers use the data (bottom up) to arrive at a linguistic model structure but generally do not formulize the model. Both are searching for causal relationships (i.e. the model structure) at multiple levels of abstraction between discernable spatiotemporal categories, patterns, and measures.

1.3 World view of a model

The classical engineering models are very limited in the number of factors that influence behavior and often assume that any variation in observed behavior across situations and conditions is the result of input, process, or output noise. In other words, the world view adopted in these models is contextually very limited. The opposite approach is that any variation in behavior can be traced to some set of factors that explain why a particular behavior manifested as it did. This assumption of greater focus on the situated nature of behavior is adopted by ethnographers in their quest to establish an explanation of the data and analyses that ties together multiple environmental and experiential factors. In other words, the worldview adopted in those "models" is in principle contextually unbounded but constrained in the process of making meaning of the observed phenomena.

2 Petri Nets

Representing ethnographically mined behavioral phenomena in a graphical state space modeling language called Petri nets brings the behavioral analysis to a common middle ground that integrates both the bottom up (ethnography) and top down (engineering) approaches into one complete (i.e. well defined world bounds) predictive logically consistent computational modeling framework that can be used to assess the models' generality across drivers and contexts.

2.1 Definition of a Petri net

The Wikipedia [2] description of a Petri net is: *A Petri net is one of several mathematical representations of discrete distributed systems. As a modeling language, it graphically depicts the structure of a distributed system* (i.e. driver, vehicle, and environment) *as a directed bipartite graph with annotations. As such, a Petri net has place nodes, transition nodes, and directed arcs connecting places with transitions.* Figure 1 shows a high level lane change Petri net.

At any one time during a Petri net's execution, each place can hold zero or more tokens (i.e. a currently active state). *Unlike more traditional data processing systems that can process only a single stream of incoming tokens, Petri net transitions can consume tokens from multiple input places, act on them, and output tokens to multiple output places. Before acting on input tokens, a transition waits until the following two conditions are met: (i) a required number of tokens appears in every one of its input places, and (ii) the number of tokens in each of its output places falls below some threshold.*

Transitions act on input tokens by a process known as firing. When a transition fires, it consumes the tokens from its input places, performs some processing task, and places a specified number of tokens into each of its output places. Since more than one transition on a net can be firing at any one time, Petri nets are well suited for modeling concurrent behavior of a (geographically) distributed system (e.g. model behavior of driver's interaction with other road users, the dynamic environment, and the vehicle).

Since its inception in 1962 by Carl Adam Petri, Petri nets have been extended with characteristics such as colored tokens, stochastic firing, and hierarchy [8]. The values of tokens in colored Petri nets (currently our targeted modeling framework) are typed and can be tested and manipulated with a functional programming language. This extension offers sufficient flexibility to model conditional firings and to attribute each token with relevant information that gets propagated through the model and used differently by different places and transitions as tokens migrate (i.e. as behavior unfolds).

2.2 Petri nets embody dynamic structures

The common goal of all disciplines that describe behavior is their quest to characterize stable structures in their observations. Where the disciplines differ is the manner in which they find these stable structures and how they label them. Petri-nets essentially embody the stable structures. The labels associated with the states and transitions that make up the graph can either be linguistic or formulaic (Figure 1).

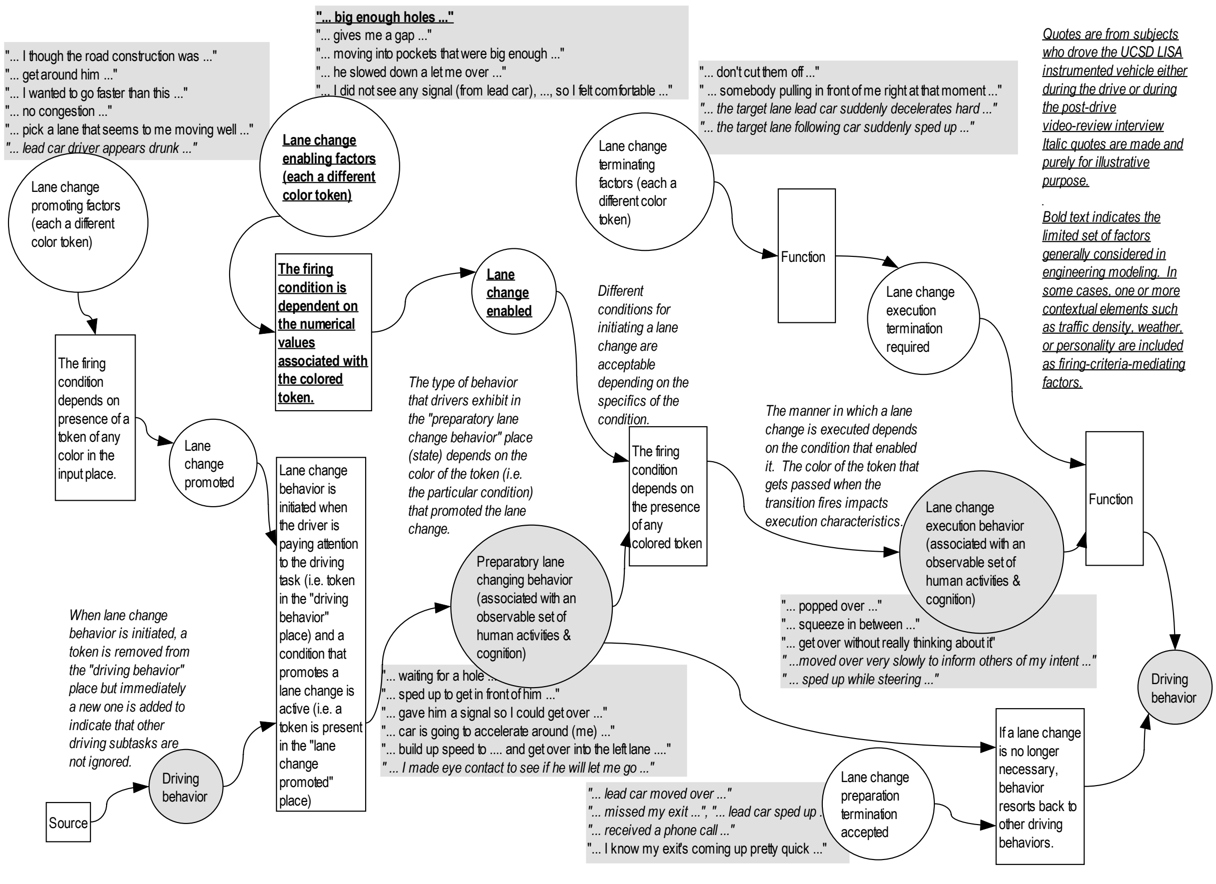


Figure 1. *High level Petri net of lane change behavior. Circles are place nodes (behavioral states). Rectangular boxes are transition nodes (behavioral transitions). The quoted phrases in shaded boxes are actual comments made by drivers during their drive; they were used as one of the data sources to organize observed behavior. They are associated with the place node that partly overlaps the shaded text box. The bold underlined text refers to the often adopted lane change view in engineering analysis. See text for further details.*

3 Distributed cognition

Distributed Cognition (Dcog) provides a particular theoretical systems-based orientation that influences the way a cognitive ethnography will be carried out. The strong theoretical claim of Dcog is that cognition is inherently a distributed process in a complex system (e.g. human, vehicle environment) at any level of description and analysis. Cognition from this view is situated, embodied and is manifest across traditional boundaries that tend to stuff cognition inside the skull of individuals.

3.1 Utility of driver goals and needs

The main challenge in establishing mutually beneficial collaboration between automotive engineers and UCSD ethnographers was only partially resolved through the use of Petri nets; the remaining difficulty proved to be the importance and singularity of human needs and goals in explaining observed driver behavior. Engineers and many human factors practitioners are deeply engrained in the perspective that driving is highly goal driven and that driver's goals and needs generate the majority of observable behavior. This is clearly evident in the hierarchical knowledge, rule, skill or strategy, tactical, operational driver models [1,7]. On the other hand, the UCSD ethnographers start with a theoretical orientation that emphasizes the distributed nature of cognitive processes in complex setting that usually involve coordinated interaction dynamics between humans, technology and structure in the environment [3].

3.2 Centralized motivation vs Dcog

In recognizing that much of human cognition is organized and coordinated with structure in the work environment rather than centralized in the brain, we are encouraged to look to our environment as representing our cognition rather than vice versa. As designers, by emphasizing the influence of complex structures in the environment on our behavior, we are better able to support the driver by modifying the environment and thus the opportunities and constraints on unfolding behavior. The goal is to structure the environment such that it promotes and enhances safety, satisfaction, and success.

3.3 The role of drivers' conscious deliberation

Even though lane-change analysis does not necessarily require human cognition per-se to explain or describe it, driver's verbal protocol does offer important insights into the way drivers make meaning from their interaction with the work environment. Verbal protocol offers a complementary stream of data that aids in mining for commonalities in behavior across different contexts, drivers, and cultures.

Often, the discrepancy between cultural (driver centered) models of how a system works and the logical engineering of the system, explains a common error in operation [4]. How drivers make meaning of a driver support system may be highly predictive of whether it will be accepted or not.

4 Lane changing

Our initial focus in bridging engineering and ethnographic analysis practices through Petri nets was on lane change behavior (see [5] for an engineering approach to the analysis of naturalistic lane changes). The lane change model in Figure 1 is one of the many levels of Petri nets that embody the analysis results in a form that both parties can easily understand and work with.

4.1 The stages of a lane change

Ethnographic analysis of behavior surrounding lane changes resulted in a number of hierarchically organized Petri nets of which a high level example is shown in Figure 1. This Petri net shows that lane changing behavior can be characterized in terms of three main states (place nodes) and three main conditions (transition nodes). Each place node (behavioral state) is a Petri net thereby creating a hierarchical network that reflects patterns and commonalities observed at different levels of behavioral analysis.

The classical engineering approach focuses primarily on mathematically characterizing the condition that enables the execution of a lane change (shown in Figure 1 as bold underlined text). In most cases these criteria are not explored extensively in terms of situational dependencies (i.e. the firing conditions are functions of the color of the token in the source places) and generally ignore the fact that the speed with which a triggered lane change is executed is affected by a number and type of contextual factors (e.g. tightness of the spot, confidence about accommodating behavior of drivers in target lane). Furthermore, the engineering approach largely ignores the other stages and thereby many opportunities for additional types of driver support.

The observable data from the highly instrumented LISA vehicle at UCSD were the starting point for the ethnographic investigation [3]; see our CoBeX data mining and visualization tool paper in these proceedings for a complete list of the data sources used that included video plus audio of each drive and post-drive reflective interviews. Examples of speech uttered by our subject drivers are shown in Figure 1 to exemplify the contextual diversity in behavior mediating processes.

5 Mining for meaning

Ethnography is based on a participant-observer stance of the researcher which means that the ethnographer is interactively immersed in the phenomena, examining it through their own experience as well as collecting observational data and interacting with 'informants' (i.e. is passenger in data collection drives with drivers). Interactive-immersion continues during the analysis phase is which the ethnographer is examining data in detail. The first pass is to select a few cases of interest (limited set of lane changes in our case), the second to scour all the sources of data on those cases, etc. [6]. The resulting discourse attempts to portray the phenomena (i.e. the lane change) from the subject's point of view.

5.1 Mining for general behavioral phenomena

The ethnographic analysis of lane changes can be characterized as a process of mining for meaning. Video transcriptions are the primary entry point in the development of a model that structures the spatiotemporal interactions between driver, vehicle, and environment based on an integration of analyses that include ethnographer participation, behavioral encoding, contextual/situational encoding, verbal output, reflective interviews (i.e. interview drivers with videos of their drive), etc.

Whether verbal protocol or observed driver activities are used as a starting point is less important than the fact that all available data are used to identify general patterns that can be characterized by behavioral states and transitions. These behavioral patterns are either labeled based on drivers' own verbal references or are categorized by the ethnographer to either capture a set of verbal references different drivers used to refer to the apparently similar behavior or to capture a meaningful label that arose from the ethnographer's own deep-rooted participation in the data collection and analysis.

5.2 Satisfying logical constraints

Once these states and transitions have been established, they can be organized in a Petri net to construct a relational network that characterizes the situated conditions that cause behavior to flow down a particular state trajectory. The resulting Petri net is almost by definition incomplete in that some states may not be connected and some transitions may not be associated with states, or some states may not have any transition that assure that behavior does not get stuck in one state. In other words, the logical requirements of the Petri net force the analyst to formulate hypotheses to fill the gaps or to go back to the data and perform a directed search for evidence about the missing transitions or states.

5.3 Integrating formulistic representations

Once a Petri net has been established, the behavior associated with a state (place node) can be characterized by some form of a dynamical system such as a continuous controller, an open or closed loop maneuver generator, a spatiotemporal head movement generator, etc.

Similarly, a transition can be characterized by a mathematical decision or presence criteria based on a number of input variables. For example, a transition to the lane change execution state may be defined in terms of time headway and time to collision to lead and following vehicles in the target lane (i.e. the typical engineering focus on lane change behavior).

This process of identifying and employing a dynamical systems characterization of behavioral phenomena essentially bridges the gap between a high level state-transition-representation of human behavior to a low level behavior-generation-representation. The latter is needed for model based hypothesis testing to assess the generality of the model across situations and subjects.

5.4 Verifying generality

The resulting Petri net is generally a reflection of the behavior (lane changes in our case) of a few instances of a single driver or a few instances of several drivers. The next step is to confront the model with the data to converge on a representation that generalizes across situations and drivers. This process leads to an extended Petri net that is a reflection of the "population".

The fact that states and transitions are mathematically defined enables us to automatically search for instances of a particular state in a pre-defined temporal or spatial window around an observed lane change. It can also be used to look for instances of a state and its associated behavior across all the available data to gain access to terminated or aborted lane changes (i.e. those that show preparatory behavior but not a lane change execution).

6 Support systems

Each identified interaction with contextual elements offers an opportunity for a support system to improve the efficiency and effectiveness of the interaction by creating a modification to the driver's work environment.

6.1 Support demands a computational model

To help a driver, the support system must be able to recognize a particular situation that is associated with a behavior (activity) that can benefit from a particular support element. This means that the situation including driver behavior itself needs to be uniquely characterized by an observable set of variables (i.e. in order to avoid application of support in situations where it is not beneficial, or vice versa to avoid not providing the support in situations where it would be beneficial); in other words, a computational Petri net model of that situation is needed to guide the development and assessment of driver support systems.

6.2 Diminishing return on support

The presence of any type of support system alters the work-domain of the driver and thus necessarily his behavior. Predicting the adaptation of behavior following the introduction of a support system or the modification of a work environment is one of the most challenging processes that interface designers face. In principle, a subsequent ethnographic analysis should be conducted with the support system in place to assess the manner in which drivers integrate the support system's functionality into their overall activity patterns. Until we can make reliable predictions about human adaptation to modifications in a work environment, our quest for understanding human behavior generation remains incomplete.

7 Conclusion

In this paper, we showed how Petri nets provide a collaborative computational framework for meshing diverse analysis techniques. Through a detailed analysis of driver's natural lane changing behavior, measurements were represented that span multiple levels ranging from body movements (hands, feet, eyes) to verbal descriptions (categories) associated with situations and contexts. The structure of interactions imposed by the Petri nets directed and expedited the search for meaning and understanding, illustrating a seldom recognized, yet powerful, side effect of a model-based approach. The common language of Petri nets provided the automotive engineers a view into the world of ethnography and with that a deeper understanding about the sources of apparent complexity in driver behavior and a structured approach to mining for and developing support systems.

This research was supported by Nissan Motor Co. LTD, Japan.

References

1. Boer, E.R.; Hildreth, E.C.; Goodrich, M.A. (1998). A driver model of attention management and task scheduling: Satisficing decision making with dynamic mental models, *XVIIth European Annual Conference on Human Decision making and Manual Control*, Valenciennes, France, Dec. 14-16.
2. http://en.wikipedia.org
3. Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA. The MIT Press.
4. Hutchins, E.; Holder, B. (2000) Conceptual models for understanding an encounter with a mountain wave. *Proceedings of HCI-Aero 2000*, Toulouse, France, September, 2000.
5. Lee, S.E.; Olsen, E.C.B.; Wierwille, W.W. (2004). A comprehensive examination of naturalistic lane changes. *NHTSA*, Washington, DC.
6. McCall, J.; Achler, O.; Trivedi, M.M.; Fastrez, P.; Forster, D.; Haue, J.B.; Hollan, J.; Boer, E. (2004). A collaborative approach for human-centered driver assistance systems, *7th IEEE Conference on Intelligent Transportation Systems*.
7. Michon, J.A. (1985). A critical view of driver behavior models: what do we know, what should we do? In: Evans, L. and Schwing, R. (Eds.) *Human Behavior and Traffic Safety*. Plenum Press, New York.
8. Murata, T., (1989). Petri Nets: properties, analysis and applications. *Proceedings of the IEEE*, Vol. 77, No 4, April, pp. 541-580.

How can we measure goodness of conversation with robots?

E. T. Harada [1] , S. Suto [2], M. Nambu [3]
[1] *Faculty of Social Sciences, HOSEI University, Tokyo, Japan*
[2] *Department of Psychology, CHUO University, Tokyo, Japan*
[3] *Department of Media Architecture, Future University, Hokkaido, Japan*

Abstract

The problem how to evaluate communication robots was investigated by a psychological experiment, comparing a human (a confederate) and three kinds of conversation designs for a robot, i.e., (a)planned utterances with a robot-like synthesized voice, (b)planned utterances with a human voice (recorded), and (c)emergent utterances with human voice. Experimental participants evaluated the communication partners after a few minutes interaction in three given situations each. During conversation, the distance between the participant and the partner was measured on every stopping position of the participant. Every index showed general preferences to a human agent against robots. With 3 different robots designs, behaviors during talking and receiving gifts from robots showed higher evaluation to emergent utterances beyond planned utterances. However, memory performance of conversations was best in the condition of planned utterances with synthesized-voices, implying importance of integrity as a robot. Dimensions of evaluations to communication robots were discussed.

Keywords

Communication robots, subjective evaluation, memory test, personal space, emergent vs. planned conversation

1 Introduction

Robots are now moving from factories to home and public spaces, and expanding their functions from exact-and-fast manufacturing to communicating with people. Therefore, the problem how to evaluate robots as communication agents is becoming an important issue. They are special because their objectives are doing “good communication”, and also they have physical bodies, which make them different from software agents in computers.

Basically usability evaluations are based on measuring costs and benefits for getting goals and, therefore, the objective ambiguity of “good communication” make evaluations difficult. Most popular methods for this “goodness” evaluation might be subjective evaluations, e.g. semantic differential methods. These methods are intuitive and easy to understand, however, their reliability and/or sensitivity are not clear yet. In order to compensate these problems, previous studies were using objective changes of conversation analysis as indices of usability evaluation on communication systems (e.g., video-phones) [2, 4]. Unfortunately however, communication levels of robots are still not so high that the human utterances during human-robots conversations seem too much affected by limitations of robots talk.

However, the second characteristics of communication robots may bring us another possibility of objective measurements; i.e., the existence of physical bodies extracts many natural responses by human. For example, users of a message system showed different speech behaviors when they used robots as the interface, compared with usual PC-based message interfaces [1]. Studies of human-robots interactions showed that eye-contacts and co-behaviors with robots were frequently observed and correlated with subjective evaluations [3, 5]. There seems to be some possibilities to evaluate robots as communication agents using these objective measures.

As an exploratory study, we conducted a psychological experiment to collect multiple indices which may be used to evaluate communication robots. In the experiment, participants were instructed to evaluate the partner of communication after a few minutes interaction as a role playing in 3 kinds of social settings, i.e., talk to an information desk, transact something with a clerk at a counter, and do some trials of a simple game. We compared a human agent (a confederate) and three kinds of agent robot varying conversation designs. Three conversation designs were (a) typical robot-like conversation, i.e., planned utterances recorded with a robot-like synthesized voice, (b) planned utterances with recorded human voice, and (c) human-like emergent utterances with human voice via speakers in a robot. To keep equality of conversational contents with all agents, planned utterances were controlled by an experimenter as a fake system, and utterances in the emergent conversation were restricted by scenarios.

One of the most interesting indexes in the experiment was distances between a participant and a robot while talking, based on analogy of the concept of personal space. Personal space is a sensitive measure of inter-personal relationship, conversational objectives and context, and also characteristics of communication participants. In the previous studies, the average distance between robots and participants did not show any correlates with subjective evaluations of robots[3]. However in that study, contents or objectives of conversation were not considered as factors, and the target of evaluations were a robot itself, not the conversation with that robot. Therefore in this study, the distances during conversation were measured and analyzed about the relationship with conversation designs of robots and contents of conversation.

Another measurements focused here was human reaction to requirements/ gifts from robots. Previous studies reported that instructions from robots were not followed by human, which was attributed to lack of social representability with robots [7]. However, there is another result that people did accept the robot’s requirement, when they have had interactions and already made up relationships [6]. In fact, replying to a partner’s speech act is an important component of natural conversation, and it might be a measurement of conversational goodness. Therefore, we included requirements and gift offering tasks, and investigated these reply.

2 Method

2.1 Design

The experiment was executed by mixed two factors design. One was the factor of dialogue partner, or agents, which had 4 levels as *between* factor; a confederate, a planned conversation-synthesized voice, a planned conversation-human recorded voice, and an emergent conversation-human voice. The other factor was conversational setting of 3 levels, which was *within* factors: explanatory as an information desk, transaction as a window clerk, and playing games.

2.2 Participants

Forty-two undergraduate students participated in the experiment as a volunteer, two of whom were omitted from data analysis because of data-equipment troubles.

2.3 Settings and tasks of conversations

Three kinds of social settings as operating conversational contents were prepared. The explanatory was set in "the exhibition of Japanese hobgoblin" with a confederate (human agent), and as an information desk of an aquarium with a robot, both of whom were asked from a guest (participants) about the contents of exhibition. As a window clerk, the task of buying a train ticket was set with a confederate, and of buying a movie ticket was set with a robot. The game partners were situated as a tournament of "rock-paper-scissors" game, with both a human and a robot. Scenarios were made and utterances were designed and recorded for the planned-conversation conditions.

Additional conversational tasks were prepared for observing reply to speech-act of a confederate/ a robot. One was the requirement, asking the participants to place the pen on the desk before going out from the 'stage', or to hand the trashcan to the person outside the laboratory, one of which was asked by the confederate and the other by the robot, or vise versa. The other was presentation of candies, by the robot after 'all experimental session was done' and experimenters were out of the laboratory, saying that "please take candies as many as you like, as my gratitude for the participation".

2.4 Robots and their interface designs

PaPeRo robot made by NEC was used as an agent robot (http://www.incx.nec.co.jp/robot/english/robotcenter_e.html). Even though PaPeRo has its own natural language processors and can move around voluntarily, all utterances and small movements were remote controlled from the next room of the laboratory in this experiment.

Figure 1. *The PaPeRo robot (NEC co. Ltd); the height is* 440mm.

There were two dimensions of conversational design; the types of utterances and voices, and three types of conversation were made; (a) planned utterances with a recorded robot-like synthesized voice, (b) planned utterances with a recoded human voice, and (c) emergent utterances by the confederate in the next room, via speakers in a robot. Even though the human-like emergent utterances with robot-like synthesized voice might be the ideal conversation by robot, it was not included in this experiment, because it was not possible by current technology in a perfect fashion.

2.5 Laboratory and devices

Figure 1 showed the sketch of the laboratory. The size of the room was 6m by 6m. The participants were instructed to think that the area separated by screens was the stage where they were supposed to play a role with a person (confederate) or a robot. Before the first screen, there was a desk and a chair for the participants to answer some paper-and-pencil type questionnaires and the space functioned as a dressing room, where experimenter was standing as a prompter. Behind the second screen, there were two experimenters to measure and recode the distance to the participant and the angle between the participant and the partner during sessions: because of difficulty to measure the distance between participants and a partner directly, the distance was calculated by the angle of the two and two distances, one of which was fixed as the sitting point of the confederate/ robot.

Three video cameras were set for recording a) the figure of the participants from their front, b) the overall view of the laboratory including the participants and the conversation partner in the same frame, and c) recoding of the distances and angles of standing points by participants, by filming the display of the instrument (LS-401 by MAX co. Ltd.). The distance was measured by laser-beam pointing to the heel of the participant's sandals which were instructed to wear in the laboratory. In order to make a fixed line-orientation between the participants and the conversation partner, there was a white tape on the floor, even though no instruction about how to use this line.

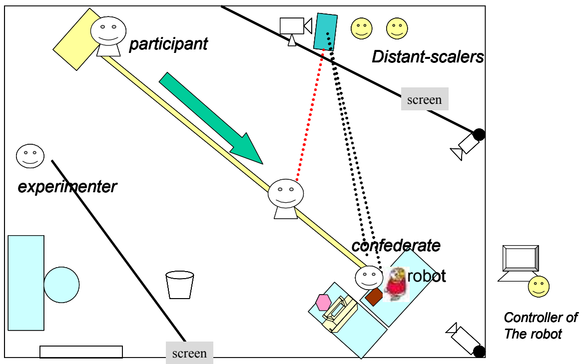


Figure 2. *The bird-view of the laboratory setting.*

2.6 Measurements and tests

There were two kinds of subjective evaluations, both of which were rated by 6 points scales. One was the evaluation of conversations with 6 dimension; enjoyable, easy to talk, understandable, reliable, natural, and "want to do the same conversation again". This rating was executed after every conversation session finished, making every participant evaluated 6 times. The other was Semantic Differential evaluations of conversation partners, a confederate and a robot, which was executed after all of 6 sessions finished. It included 11 items based on previous studies of robot evaluation [3] and added items about weightiness; i.e., easy to talk, human likeness, understandable, familiar, relax, preferable, good, fast, interesting, smart, and heavy.

After 6 sessions, there was a surprising memory test about contents of two exploratory situations, with a human agent and a robot. Recognition of names of explanted items (fishes and hobgoblins), and paired recognition of names with contents were tested by paper-and-pencil task.

2.7 Procudures

The experiment was executed individually. When participants came to the laboratory, they were instructed to change their shoes to sandals 'for cleanliness of the lab', and explained the objectives of the experiment to evaluate conversations with robots, after playing short skits of several situations in accordance with rough scenario, which was demonstrated by the experimenter once. The description of the situation and objectives of 'the person in the play' was written on the card. During preparation by participants, the conversation partner (a confederate or a robot) was placed at the fixed position on 'the stage'. After understanding the task (scenario), the participants entered the stage from the starting point, and walked to approach to the partner, while the partner talked to the participants before they stopped (e.g., "Welcome the HOSEI Aquarium!, how can I help you?"). During the conversation, the angle and the distance was measure any time participants stopped. Participants were not told about measuring positioning and distances, and it was confirmed that no one was aware of the measurement at experiments were executing after experiments.

After a conversational session ended, the participants came back to the 'dressing room', and evaluated the conversation with 6 dimensions. All participants started with a human agent with 3 sessions, order of which was counterbalanced by participants, and then did the 3 sessions with a robot, as conditions of the robot was randomly assigned to the participant. After 6 sessions, subjective evaluations of two partners, memory tests were executed, followed by debriefing and interviews. The Experiment lasted about a half hour.

3 Results

3.1 Subjective ratings

Eleven items of subjective evaluations for conversation partners were analyzed by factor analysis (varimax rotation) and extracted two factors (cummurative variance 62.82%), named as the *FAMILIARITY* factor and the *CONVERSATIONAL ABILITY* factor. One way ANOVA with two factor scores by agents (4; human and 3 types of conversation design) showed that the *CONVERSATIONAL ABILITY* was significantly higher with a human agent compared with three kinds of robots *($F(3, 73)=7.10$, $p<.01$)*. Because utterances in the emergent conversation condition were the same as in the human agent condition, the results that there were no differences within 3 robot conditions was surprising. The *FAMILIARITY* did not show any significant effects by one way ANOVA(4), however it showed that human agents were higher evaluated than averages of 3 robots when doing planned comparison within participants.

Factor analysis of six items of conversation evaluations showed 1-factor structure (67.63%), and 2 way ANOVA (4x3) with the factor scores showed (a) explanatory conversation was higher evaluated than other two tasks, and (b) in addition to human agents were higher evaluated than 3 kinds of robots, the emergent-conversation robots were higher evaluated than other two planned-conversation robots, which implied that emergency of communication was important for conversation.

3.2 Memory tests

Recognition tests for names and contents of items which were explained in the exploratory session, were scored and analyzed by one way ANOVA of agents(4). The results showed significant differences between conditions *($F(3, 76)= 4.72$, $p<.05$)*. There were no differences between human agents and planned-conversation with robot-voice condition, and planned-conversation with human-voices were the worst (Figure 3).

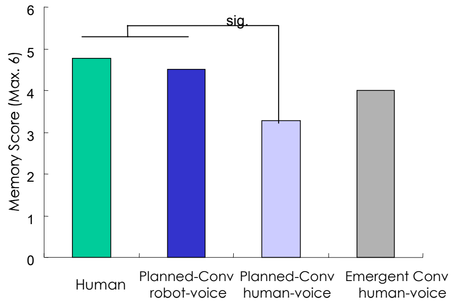


Figure 3. *Average scores of memory scores for recognition tests.*

3.3 Distances and movements

The distances between participants and the conversation partner were calculated by every stopping position by participants during conversation session, and analyzed by agents (4) X settings (3) ANOVA. The results indicated the interaction of two factors *($F(6,152)$ $=1.95$, $p<0.1$)*, and human agents condition and planned- conversation robot-voice conditions showed the differentiation between settings.

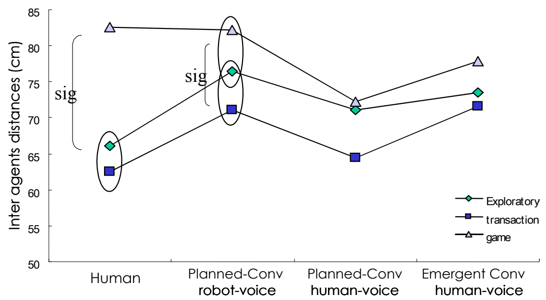


Figure 4. *Average distances between participants and the partner by function of agent conditions and settings.*

In addition, the frequencies of moving by participants were counted and analyzed, which resulted that significant tendencies of agents both with approaching and leaving (approaching, *$F(3,76)$ $=2.26$, p $<.10$*; leaving, *$F(3,76)$ $=2.58$, $p<.10$*). Main effects of settings ($F(2,152)$ =20.38, $p<.01$), and interaction of agents and settings ($F(6,152)$ = 2.79, $p<.05$) was significant only with approaching.

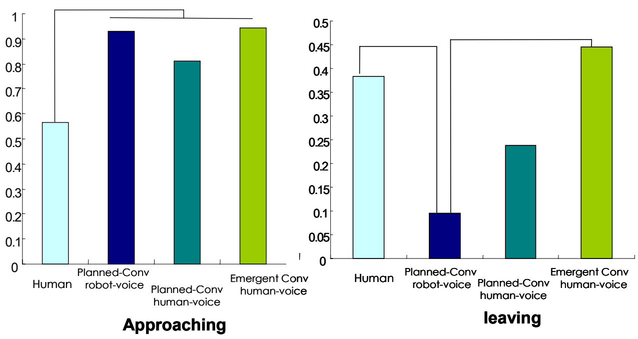


Figure 5. *Average frequencies of participants moving by function of agent conditions and settings*

Because some conversational settings included objects that to be handed over or showed, approaching movements were mostly affected by these physical necessity, producing frequent occurrence of approaching to robots compared with human agents, who could reach their own hands and change postures. In contrast, leaving, or distancing from the partner was more frequent with a human and the emergent conversation condition, implying that more human-like conversation made participants to aware that they had approached too much closer while talking.

Other behavior measurements during conversation were gestures and/or paralinguistic indices by participants. In this study, only nodding frequencies were counted with the explanatory setting. There was significant effects of agents *(F(3,71) =20.95, p<.001)* and Figure 6 showed. that human and emergent utterances from robots extracted more nodding than planned utterances.

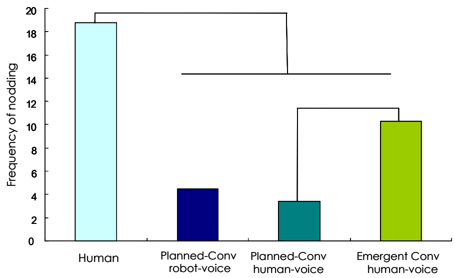


Figure 6. *Average frequencies of participants nodding during conversation.*

3.4 Accceptance of requirement / gift

When the conversation partners asked to participants to place or hand some objects to other places/ people just after their conversation finished, all of these requirements were accepted and executed immediately by participants. These results agreed with results which showed the importance of preceding interactions [6], not with results in which people did not follow robots' instruction. [7].

Sharply Contrasted with this results, receiving of gifts showed large differences between agents (Figure 7). Even though the offers were after 3 sessions of conversations, the offer in the planned- utterances with robot voices condition were rare to accept, whereas these were mostly accepted in the emergent conversation conditions. It seemed that human-like utterances was important to get some candies in accordance with utterances from robots, implying responsibility had effects [7]

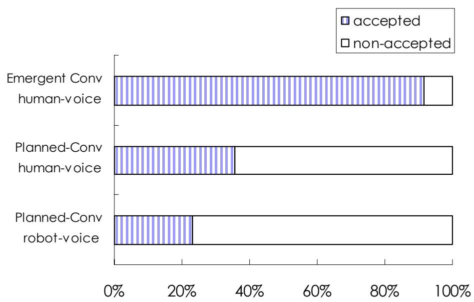


Figure 6. *Rates of receiving of gift by function of agents.*

Discussion

Results of the experiment showed overall superiority of human agents beyond robots as communication partners. However, the evaluation data to robots were also effected by conversational design, and their trends were not in a single fashion; (a) in direct subjective evaluations of robots showed little differences between designs, whereas evaluations of conversation showed some sensitivity of design changes, (b) behavior during conversation, i.e., nodding while listening and leaving (taking distances) movements, showed similarity to inter-human behaviors in the emergent-utterances condition, and this might be related to receiving of gift offers, and (c) similarity to a human partner in aspects of understanding of conversational contents and also patterns of taking default distance which varied with communication settings was observed with a robot of robot-like voice and planned utterances.

These results indicated that there were two dimension of communication goodness, one is human resemblance, or human-like intelligent conversation, and the other might be the total integrity as a robot. Unfortunately, subjective evaluation of robots themselves by SD method was not sensitive enough to distinct these dimensions, even though the methodology are most popular and easy to use in evaluation for communication agents. These results implied the acute necessity of development of easy to use and sensitive methods of objective measurement of communication goodness, for supporting wholesome social expansion of robots for communication uses.

The authors thank J. Osada, Y. Fujita, and NEC Personal Robot Research Center for supporting this study.)

References:

1. Harada, E.T. (2002) Speech as a user-interface: Verifying Effects of existence of agency and social contexts. *19th Annual meeting of Japanese Cognitive science Society,* Japan (In Japanese).
2. Harada, E.T. (2004) Usability for communication systems: from Cognitive Engineering points of views. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 16(2), 18-24.* (In Japanese).
3. Kanda, T.; Imai, M.; Ono, T; Ishiguro, H (2003). Numerical Analysis of Body Movements on Human-Robot Interaction. *IPSJ Journal*, 44(11), 2699-2709 (In Japanese).
4. Nambu, M.; Harada, E.T. (2002) Interactive processes in video-mediated referential communication. *Japanese Journal of Psychology, 73(3), 219-226.* (In Japanese).
5. Ono, T; Imai, M.; Ishiguro, H; Nakatsu, R. (2001) . Embodied Communication Emergent from Mutual Physical Expression between Humans and Robots. *IPSJ Journal*, 42(6), 1348-1358 (In Japanese).
6. Ono, T; Imai, M.; Nakatsu, R. (2000). Reading a robot's mind: A model of utterance understanding base on the theory of mind mechanism, *Advanced Robotics, 14 (4), 311-326.*
7. Yamamoto, Y.; Sato, M; Hiraki, K.; Yamasaki, N.; Anzai, Y. (1992). A request of the robot: An experiment with the human-robot interactive system HuRIS. Proceeding of *IEEE International workshop on Robot and Human Communication (ROMAN'92), pp.204-209.*

Virtual Reality technology - a tool for behavioural studies involving disabled people

G. Johansson, R. C. Davies, J. Eriksson, M. Haake, J. af Klercker, C. Magnusson, K. Rassmus Gröhn, M. Wallergård
Department of Design Sciences, Lund University, Lund, Sweden

Abstract
The increased computer capacity and new software from the game industry have created powerful and inexpensive Virtual Reality systems. By combing realistic virtual environments and virtual characters a useful platform for measuring behaviour is achieved. The method offers ecological valid environments, controllable and repeatable situations with no disturbances and gives a usable method to measure cognitive procedures. Often the Virtual Reality technique is combined with other measurement techniques such as tracking of motion or measurement of different physiological parameters. In this paper the method is applied in studies involving people with visual or cognitive impairment.

Keywords
Virtual Reality, virtual characters, disabled

Background
During the last decade, due to increased computer capacity and development of the 3D computer game industry, small and inexpensive Virtual Reality systems have been made available and the usage and application of virtual environments for simulation and visualisation has dramatically expanded. This means that it is becoming commonplace, and the true potential of being able to use a person's natural perception and interaction metaphors can be realised. In particular, virtual environments including virtual characters offer new possibilities to study and measure behaviour.

What is Virual Reality?
The term, Virtual Reality (VR) has been in use since the early 90s when authors such as Rheingold [12] and Kalawsky [7], began to write about how people could be immersed in a computer-generated environment (or Virtual Environment (VE)) with devices that matched the senses and captured body gestures. In those days, small screens in front of the eyes, called Head Mounted Displays (HMD) and headphones to allow the users to perceive the VE, with Data Gloves to allow interaction with the virtual objects were the main components.

However, there are many more technologies available today from single user, networked systems so several people can cooperate in the same VE at the same time, to large screen systems together with special glasses and audio-speakers that allow a group of people to share and control an experience with 3D perception. It is even possible for people to feel (with the fingers) virtual objects through so called haptic devices, often used in surgical training and design and even to experience virtual smells. Nevertheless, the most common systems are also the simplest, and much can be achieved with a normal PC and a couple of projectors. There are many books on the subject, one that can be recommended that covers most of the technology and application areas is Stanney [14].

This technology also has the ability to offer the user not only a role as a passive observer but as an active participant performing a variety of tasks in the environment. The user communicates with the computer system using different devices tracking for instance the movement and the position of the hand, the head or the eye. The realism of the virtual environment can be chosen by varying the wealth of details when adding colours, patterns, textures, lighting and behaviours in the model.

VR systems aim to give the users a feeling of presence, of actually *being in* the VE. This is important for the success of the application, as explored in the recent series of books edited by Riva and Davide [13], where the focus is on how communication is supported through virtual technologies. When this is combined with knowledge on human perception and natural interaction metaphors, VR becomes a very powerful tool.

Interacting with a virtual environment
When performing a task in a virtual environment you normally want to move around (navigate) and interact with the objects in the environment.

Navigation in a VE can have two meanings; travel and wayfinding. Travel is the movement of the viewpoint from one location to another, whereas wayfinding can be described as the cognitive process of determining a path through the environment to the desired destination

There are basically two types of self-controlled navigation; fly-through and walk-through. In the former case the user can move freely in 3D space and may also tilt and pan the viewpoint. Walk-through navigation is a simplification of real world navigation since it's not possible to raise or lower the viewpoint.

To be able to perform a task a natural interaction with the objects in the environment is important. Interaction techniques for manipulation of objects should support: object selection, object positioning and object rotation.

Interacting with virtual characters/ agents
Human models (mannequins) have been used in 3D environments for a long time. They have been used for testing the environment in different ways. A main topic has been in ergonomic evaluations of an environment.

The last years in computer development, have brought us closer to animated interactive educational and social virtual characters in real-time virtual reality environments. With better and more powerful algorithms for intelligent behaviour, computer generated speech, communicational gestures, etc., we are soon on the verge to encounter them in various circumstances – and we have already encountered a precursor in the Microsoft Paper Clip pedagogical agent, not to speak about many of the educational CD-ROMs for children.

The state of the art in virtual pedagogical agents' research can be seen in the agents Laura and REA by Thimothy Bickmore, which he has presented in his doctoral thesis at MIT-Media lab [2] .

In training contexts the applications with virtual characters can grow into inhabited virtual environments where one can simulate and train not only how to handle artefacts, but also how to interact with people under especially risky, stressful, and even dangerous scenarios. For example, one can imagine banking employees, training the procedures around bank robbery or rescue party handling panicked people under severe stress. At the department we have started a project where we have surveyed the area of virtual pedagogical agents [5], and we are now moving into a phase where we are investigating how people react and interact with virtual pedagogical characters, with a special focus on the visual and social qualities within the agents [6].

The Virtual Environment method

Using virtual environments and virtual characters as a tool for behaviour measurements has both advantages and disadvantages in the research design. The method affords ecological valid environments, controllable and repeatable experiments with no disturbances and a usable method to measure cognitive procedures. On the other hand the computer interface adds an extra threshold for the user and most VEs do not provide motor feedback when performing a task.

A valuable and common way of working with VEs is to use a scenario technique and performing different tasks. The activity of the people are registered using video cameras and the activities in the virtual environments are logged. Often other measurements are performed simultaneously i.e. tracking of motion or measurement of different physiological parameters. Another interesting approach is to combine virtual environment/objects and real life environment/objects (augmented reality.

Research areas

Within a recently founded centre – RE-FLEX - at Lund University (www.reflex.lth.se/reflex), diverse multi-disciplinary projects using Virtual Reality technology are performed. The department of Design Sciences is one of the key members of the centre and our research group is involved in several projects using virtual environments in the field of rehabilitation and habilitation.

In this research field virtual reality is used as a tool for: planning and design with active participation of the disabled people using their tacit knowledge, training and rehabilitation, assisting in the activities of daily living, diagnosis and treatment, experiencing different environments and activities difficult to access due to the disability. In all these application areas the measurement of the behaviour of a user performing a special task or using a special environment is important.

Currently, we are mainly working with people with visual impairments and various cognitive disabilities (such as brain injury).

Visual impairment

The research group has explored haptic and auditory environments for persons with visual impairments since 1995. The type of environments explored include desktop type applications such as memory games, a battleship game, mathematics applications, applications for drawings and images etc. Recently we have performed a set of studies focusing more on what we call non-visual virtual reality – i.e 3D environments involving the senses of touch and hearing. Two studies of navigation within a virtual 3D traffic environment have been performed [9]. The second study [10, 11] involved ten severely visually impaired users using a virtual haptic-audio traffic environment. The virtual environment was a model of a real traffic environment in Lund and included 484 static objects and 35 dynamic objects (cars and bicycles).

Cognitive impairment

We are studying the use of virtual reality in the rehabilitation process as a training and assessment tool [8, 15]. There is a growing body of evidence that information collected from patients' behaviour in a virtual environment provides grounds for valid assessments of people with brain injury [2, 16]. We are also developing methods for involvement of cognitively disabled users early in a planning process of public environments such as transport systems. Traditional planning tools, such as 2D drawings, may be hard for able-bodied people to understand [1, 4] and can therefore be expected to be inaccessible for people with cognitive problems. We are currently conducting a study that aims to find out if a VR based method, which makes it possible to interact with an environment that does not yet exist in reality, might provide means for this population to communicate their tacit knowledge and experiences regarding accessibility issues.

References

1. Bailey K., Brumm J., Grossardt T. (2002). Integrating visualization into structured public involvement - Case study of highway improvement in central Kentucky. *Transport Planning and Analysis 2002: Transportation Research Record* **(1817)** 50-57.
2. Bickmore, T. (2003). Relational Agents: Effecting Change through Human-Computer Relationships. *Ph.D. Thesis,* Massachusetts Institute of Technology (MIT), Cambridge, MA.
3. Brooks B.M., Rose F.D. (2003). The use of virtual reality in memory rehabilitation: current findings and future directions. *NeuroRehabilitation.* **18(2):**147-57.
4. Bucolo, S., Impey, P., Hayes, J. (2001). Client expectations of virtual environments for urban design development. *Proceedings of the International Conference on Information Visualisation*, pp 690-694.
5. Gulz, A. and Haake, M. (2005). Design of animated pedagogical agents – a look at their look and visual form. *International Journal of Human-Computer Studies*, in press.
6. Gulz, A. and Haake, M. (2005). Social and Visual Style in Virtual Pedagogical Agents. *Workshop Proceeding: Adapting the Interaction Style to Affective*

Factors at at the 10th Interantional Conference on User Modeling (UM'05), in press.

7. Kalawsky, R. (1993). The Science of Virtual Reality and Virtual Environments : A Technical, Scientific and Engineering Reference on Virtual Environments, Addison Wesley.
8. Lindén, A., Davies, R. C., Boschian, K., Minör, U., Olsson, R., Sonesson, B., Wallergård, M, & Johansson, G. (2000). Special considerations for navigation and interaction in virtual environments for people with brain injury. In P. Sharkey, A. Cesarani, L. Pugnetti & A. Rizzo (Eds.), *Proceedings of the 3rd international conference on disability, Virtual Reality and associated technologies* (pp. 299-304). University of Reading, Maidenhead, UK.
9. Magnusson C., Rassmus-Gröhn K. (2003). Non-visual Zoom and Scrolling Operations in a Virtual Haptic Environment, *EuroHaptics 2003*, Dublin, Ireland.
10. Magnusson C., Rassmus-Gröhn K. (2004). A Dynamic Haptic-Audio Traffic Environment, *Eurohaptics 2004*, Munich, Germany June 5-7 2004.
11. Magnusson C., Rassmus-Gröhn K. (2005). A Virtual Traffic Environment for People with Visual Impairments, Accepted for publication in *Visual Impairment Research.*
12. Rheingold, H. (1992). Virtual Reality, The Revolutionary Technology of Computer-Centred Artificial Worlds – and how it Promises to Transform Society, Touchstone Books, ISBN: 0-671-77897-8
13. Riva, G., Davide, F. (2001-2004). Emerging Communication; Studies in New Technologies and Practices in Communication, IOS press.
14. Stanney K. (2001). The Handbook of Virtual Environments, Lawrence Erlbaum Associates, ISBN: 0-8058-3270-X.
15. Wallergård, M., Lindén, A., Davies, R. C., Boschian, K., Minör, U., Sonesson, B., & Johansson, G. (2005). Initial usability testing of navigation and interaction methods in virtual environments – developing usable interfaces for brain injury rehabilitation. Submitted to *Presence - teleoperators and virtual environments*.
16. Zhang, L., Abreu, B. C., Seale, G. S., Masel, B., Christiansen, C. H., & Ottenbacher, K. J. (2003). A virtual reality environment for evaluation of a daily living skill in brain injury rehabilitation: Reliability and validity. *Archives of Physical Medicine and Rehabilitation*, **84**, 1118-1124.

Driver Behavior Interface Test Equipment – D-BITE and the RoadSense Project

A.V. Thomas[1] , M.H. Martens[2]
[1] *Jaguar & Land-Rover Research, Jaguar Cars Ltd., Coventry, United Kingdom*
[2] *TNO Human Factors, Soesterberg, The Netherlands*

Abstract

The EC co-funded RoadSense project was established to provide the framework for the Human-Machine Interface of innovative Advanced Driver Assistance Systems (ADAS) during development and evaluation for safety, comfort and support.
Critical driving scenarios based in accidentology were collected and correlated to existing or new human factors indicators and metrics. Engineering parameters associated with these parameters were identified and hence sensing systems defined for tests in simulators or vehicles to validate the ergonomic and electronic concepts.
D-BITE (Driver Behavior Interface Test Equipment) open-architecture modular systems were developed and demonstrated with a range of ADAS's, integrating inputs from video, vehicle data systems, radar, infra-red, specialist sensors, audio, and keyboard inputs, all synchronized for real-time analysis.
Laboratory post-processing can then isolate events of interest, measure exact timing for stimuli and responses of the driver and vehicle, and extract behavioral patterns. Data can be further processed by third-party systems for detailed ergonomic usability studies, etc.

Keywords

Driver behavior, metrics, HMI, ADAS.

1 Introduction

The growth of road transport Telematics and Advanced Driver Assistance Systems raises the question of operational suitability of the Human-Vehicle Interface (HVI). Installation of systems such as adaptive steering, navigation and infotainment services has grown in the last decade from almost zero to about 6% of new vehicles, set to reach 50%, with growth rates of up to 24x during this decade [2]. The concerns over distraction [5] are exemplified by legislation prohibiting hand-held mobile phone use while driving, enacted while this project was still under way. On the other hand, traffic efficiency and safety, driver stress and fatigue, could and should be improved by wise application of systems such as dynamic navigation taking into account live conditions, adaptive cruise control, following-distance or lane-keeping support, blind-spot monitoring; night vision, voice control, and adaptive emergency brake assistance. Assessment guidelines based on valid Human Factors metrics are consequently desirable for manufacturer and consumer assurance, taking into account safety, psychological comfort and usefulness of support to the driving task.

1.1 Project Outline

The EU "RoadSense" (Road Awareness for Driving via a Strategy that Evaluates Numerous SystEms) project was therefore created under the European Commission's Fifth Framework 'GROWTH' program to research a set of metrics, methods, and tools on driver workload for widespread adoption, and developed D-BITE to appraise HVI strategies during system development as a contribution to global research and evaluation activities. This paper focuses on technical and HVI aspects of test and measurement.

The consortium comprised five European vehicle manufacturers, three research and technology organizations and three universities across five countries: Centro Ricerche FIAT, PSA-Citroën, Renault, Porsche, Cranfield University, Université Blaise Pascal, Université Technologie Compiègne, TNO, CNRS LASMEA and CNRS HEUDIASYC, with the EC contributing 50% of the funds. The project ran from 2001 to 2004, with a budget of €4.4M. Besides road transport, there are other potential transport operative HMI applications.

1.2 Research Approach

RoadSense started by investigating the available Human Factors (HF) Indicators and Metrics relevant to driver behavior, setting stringent criteria for adoption. In parallel, road traffic accident databases were surveyed to find which could give definitive information on cognitive failure in driver performance as a cause, and hence define representative scenarios in which to test the critical HVI. Emerging ADAS applications were also identified, and these three strands framed the specification of a suite of Indicators and Metrics, and D-BITE requirements.

HF Testing in simulators, on test tracks and on public roads was used to validate and refine the parameters. Prototypes of D-BITE were created and tested in the laboratory and in vehicles, further refined, and analysis tools developed.

Finally, four ADAS-equipped vehicles were provided, and D-BITE was installed in each to demonstrate the use of the Indicators, Metrics, and D-BITE, against criteria for data quality, user-friendliness, and reliability; one vehicle replicated the major simulator tests which had included a virtual model of the real road route. In total, over 170 drivers participated in over 210 runs.

2 Definition of Methods & Tools

2.1 Critical Scenarios & Accidentology

Hard data on the influence of ADAS on driver performance in relation to involvement in accidents was sparse, due to their low market penetration to date and arguably because some may actively help reduce involvement in accidents. Only one database, limited in geographic scope, contained the in-depth type of information needed and was available to the project: LAB (Laboratoire d'Accidentologie, de Biomécanique et d'étude du comportement humain, a joint research facility of PSA and Renault) classifies driver activity before the accident as assessed by a team of technicians and a cognitive psychologist, following the INRETS model [8]. The RoadSense analysis demonstrated that critical driving scenarios in the gestation of an accident can be associated with driver functional failure in the transitions from normal, through marginal, to unsafe driving, and thence into collision. The cognitive functions of perception, evaluation, interpretation, decision, and action are employed continuously and interdependently. A functional failure occurs if any of these are performed incorrectly, inaccurately, too late, or are missed, that is regardless of consequences. This is easier to understand with examples:

- perception - failure to see, hear, or notice something significant for safety;
- evaluation - mistaken estimation of relative position, velocity, curvature or grip;
- interpretation - incorrect appraisal of signals, markings, or a road-user's intention;
- decision - choosing unsafe direction, speed, or gear;
- action - braking or turning too hard or not hard enough, missing a gear-change, or hitting the wrong pedal.

Perception was the leading functional failure (46% overall), then defective action (27%), but comprehension failures (evaluation, interpretation and decision together) also accounted for 27% of cases. Action failures followed other primary functional failures in 87% of the accidents. Thus, when devising driver information systems, it is important to help drivers not only to detect something that they would not have perceived otherwise but to understand something that they have detected but might not understand or anticipate correctly.

The generic scenario covers the concerns raised by accidentology, guiding the experimenter to assemble a test routine combining the elements of road configuration, speeds, time of day, traffic, duration, primary and secondary tasks, driver profiles, etc. which are relevant to the HMI technology being assessed.

2.2 Human Factors Indicators and Metrics

Sufficient consensus was achieved on the following objective indicators for Primary (essential driving) task and Secondary (discretionary or minor) task performance:

- lateral control
- longitudinal control
- visual management
- interactions with other vehicles
- situation awareness
- system suitability

Similarly, subjective indicators were confirmed for drivers' own judgments on

- mental workload
- system usability

Selection was made on the basis of

- sensitivity to changes in workload
- diagnosticity in isolating demands on specific driver resources
- intrusiveness of measurement technique upon the primary-task
- implementation requirements such as specific equipment or operator training
- operator acceptance of the technique
- selective sensitivity to mental workload (independence of such factors as physical load)
- bandwidth and reliability, both within, and across tests.

The subjective methods identified to support these indicators were:

- NASA-TLX Task Load Index
- Situation Awareness Rating scale Technique (SART)
- Bedford Workload Scale
- Risk Assessment
- Overall Workload
- Rating Scale Mental Effort (RSME).

Risk Assessment is the product of the driver's verbal assessments of (a) risk of an accident on a scale between 1.0 and 10^{-7} and of (b) severity of outcome - from superficial car damage only, to fatal injuries. It is very subjective, but may actually be more relevant to the drivers' behaviour than an objective technical calculation such as time-to-collision and equivalent crash test speed that an accident reconstructionist would apply.

Attributing changes in driver behaviour directly to use of new technology needs care, since the reliability of measurement also needs to be considered. Many of the HMI assessment techniques have never been evaluated for their reliability and validity [6] to the standards of other areas of psychometrics. The measures most closely related to the new equipment (the usability of the interface and the internal cognitive states of the driver) are more valid as measures of the effects of the new technology, but they are also the least reliable [7]. Many physiological indicators of driver status were unsuitable because there was no consensus on their significance, reliability, or repeatability in a traffic situation, or they were too intrusive and require medical personnel to implement them. However, D-BITE could be used in future studies to investigate physiological responses in a traffic context.

2.3 Instrumentation - D-BITE specification

It is not always obvious what information might be relevant until post-test analysis is under way. Therefore the data acquisition system must be flexible, comprehensive and have a very high capacity. Many of the metrics use the same elementary control and dynamics data from the normal vehicle CAN (Controller Area Network) data bus, plus video of the road ahead, the driver's hands, feet and eyes, and the input and output of the ADAS. Operator text comments and driver/operator conversation should also be recorded. CAN includes wheel speed, gear selection, brake application and level, throttle pedal position, steering wheel angle and yaw rate. Radar or stereo optical image processing establishes the position and velocity of the host vehicle relative to the road/lane and other traffic. Specialized test equipment such as various gaze trackers, Peripheral Detection Task (PDT) [9], and Lane Position Sensing (LPS) [1] also needs control and record interfaces with D-BITE. PDT, devised by TNO, generates random (in location and time), pre-determined stimuli in the driver's peripheral visual field, presented either on the outer screen (driving simulators) or with LEDs attached to the head (field experiment); the driver responds on detecting the stimulus by pressing a button attached to a finger. PDT measures and records the reaction time lapse and percentage of missed signals.

Redundancy in data can be used to cross-validate, and much data of no intrinsic interest, e.g. between manoeuvres or ADAS interventions, can be used to confirm or update 'baseline' conditions

Most importantly, every data channel must be synchronized to a common time base regardless of the latency or internal processing delay of the channels, to ensure that corresponding stimuli, driver reactions, and vehicle responses are correctly understood in sequence. Finally, it would be unforgivable if D-BITE's own user interface were not ergonomically sound in ensuring that the correct systems are in place and operational throughout, and that subsequent analysis is efficient and effective.

2.4 Validating Indicators & Metrics

Additional tests were performed for the effects of road type and geometry, familiarity with the test route and vehicle, workload, driving sequence, and test facility on metrics and methods, and for equipment compatibility with in-vehicle conditions [10].

Sensitivity tests for lateral control metrics were performed by TNO using the LPS on straight and curved track with 12 experienced drivers undertaking a secondary task, with three different steering characteristics. Objective outputs included Time-to-Line-Crossing (TLC), with subjective evaluations of handling behaviour and preferences, and RSME. High-frequency control, driving speed, steering reversal rate, steering entropy error, TLC, and standard deviation of steering wheel rotation were all sensitive to steering feel. Target values for these parameters were identified, but RSME did not yield significant results.

Sensitivity of metrics to workload and familiarity was explored firstly by superimposing a sustained visual secondary task on the primary driving task for five experienced drivers, on a straightforward low/medium speed route, repeated over a few days, using ISCAN equipment [4] for gaze target detection. In general, the visual management metrics were sensitive to the primary task demand but not to route and car familiarization [3]. Secondly, sensitivity of NASA-TLX, lateral and longitudinal control were combined with assessment of intrusiveness and validity by PSA. This compared 18 experienced drivers at spontaneous speeds on an autoroute, then while making simple adjustments to a navigation system, and finally during destination input. The subjective scores were sensitive to task difficulty, but the objective metrics were not: drivers tended to apportion resources between primary and secondary tasks e.g. by avoiding overtaking other traffic while performing a secondary task. In this respect, the subjective measures may be more indicative. 18 younger drivers undertook validity testing for intrusiveness, auditory and short-term-retention demand sensitivity, and reaction-time-threshold. Destination selection and phone number entry (memory and manual task) were performed on an autoroute. A PDT was triggered by the experimenter during spontaneous lane changing with or without a vocal task. NASA-TLX was again applied. The superimposed task increased late detections by a factor of more than 4, omissions by a factor of 10, and reaction time by a third; lane changing was associated with the majority of missed or late responses. TLX scores confirmed low, but definite, PDT intrusiveness on subjectively experienced temporal, effort and mental demands.

Testing on the Cranfield STI SimDrive™ simulator was undertaken to validate and correlate objective and subjective metrics for driver behaviour, with 40 female and 44 male subjects aged 18 to 75. The virtual test route comprised a 25.75km mix of motorway, rural trunk, minor rural and urban roads with complex intersections, traffic lights and pedestrian crossings. Participants performed a second run after a controlled interval of 3-9 weeks to avoid route learning effects, while retaining simulator familiarity. Objective data was captured at 10m intervals, and subjective data at fifteen selected sections along the route (typically one or two test types per location), or at the end. Objective and subjective data at these sites were compared with each other, between test and re-test, and across driver age groups. The aggregated test/re-test correlation was less than 0.50 in all except five instances. Overall, 34% of test/re-test measures were below 0.1. Repeatability of these objective measures in the controlled situation of a simulator was thus not observed.

3 Implementation

RoadSense built on previous investigations by RoadSense academic partners in the fields of optical image processing, telemetry, and high data volume mobile instrumentation and control.

D-BITE is modular to allow for ever-increasing types of ADAS applications and their interactions. This is combined with distributed processing of data at several levels from raw sensor data to finished recordings, the ergonomist using a laptop for off-board configuration or onboard configuration and control, and a bespoke onboard industrial computer; many sensors have integral smart sensor chips or a dedicated PC. Multiple streamed video cameras, vehicle dynamics CAN data, the specialized sensors and the ADAS HMI under test can create over 100 GB of data per hour. The high data rates required a network with a peak bandwidth such as IEE 1394a at 400 Mb/s, with additional Mindready software (SedNetTM 1394 Low-level API (LLA)) to handle bulk alphameric data as well as images. Bench-top systems were used to confirm D-BITE concept feasibility and examine compatibility of different subsystems, from which a core list of hardware and software devices was generated for implementation on the vehicles. It is also possible to mix IEE1394 with Ethernet and USB 2.0; all have been demonstrated, as have Windows 98, 2000 and XP, and Linux platforms. IEEE 1394 retains advantages of connective topology (peer-to-peer, tree, or chain), power distribution, and high compatibility with digital video devices.

A key feature of D-BITE is full synchronization of the records of events and data regardless of the process time. This makes it possible to identify the correct sequence of inputs and discriminate between primary and secondary causes, reactions, and effects. For instance, a traffic signal changing from green to amber can be detected on forward-view cameras, while driver gaze can be linked to the observation of the light a moment later and the movement of the foot from throttle to above, but not immediately touching, the brake pedal as can be seen on another camera and linked to engine, brake and car speed data, even while the driver may have been using a visual display. Each data item is time-stamped. Additionally, the observer can monitor much of the output data and insert time-stamped comments. Statistical analysis of some activities within a trial can be computed cumulatively in real time, or await post-processing. Two demonstrator cars, each equipped with their own D-BITE units, recorded two-car tests synchronously with each other via a wireless network.

The Jaguar D-BITE system implementation was configured to monitor the touch-screen driver interface for Voice-controlled GPS navigation, climate, audio and hands-free car phone systems (although the final experimental HF programme did not utilize all these facilities). Renault, PSA and FIAT installations were similar in concept, but Renault demonstrated a haptic-throttle headway support system with audio-visual warnings, PSA used a night vision plus pedestrian detection and warning installation, and FIAT used a multi-function driver display. Alternative host ADAS systems were also short-listed and feasible: GPRS/GSP dynamic route guidance/floating car data system, blind-spot monitoring and warning, and lane-keeping support.

D-BITE enables user configuration for each experiment first in terms of HF indicators, then metrics and creates a 'shopping list' of the hardware and software modules using the laptop alone or in CTMU (Central Task Monitoring Unit) mode on-line to the industrial PC. On-line, D-BITE checks module presence on the installation, displaying alternatives where they exist, verifies and activates them, and prepares output files. The user can select and view the application channels 'live' and monitor them while calibrating, or starting the recording and the test scenario. D-BITE integrates all the data continuously onto a single storage device, e.g. a removable 120GB hard drive, a CD-RW or DVD, for post processing with D-BITE PLAYER on another PC while the next experiment starts. The user can however use the CTMU in situ to review and process data with D-BITE PLAYER.

D-BITE PLAYER allows the user to select a test for replay, selecting individual data channels as required, or grouped with the relevant indicators and metrics, to be re-played for analysis, e.g. to find a particular sequence of events. Having selected what is to be displayed on the monitor(s), fully synchronized re-play is available at 1/30 to 30x real time or frame-by-frame. A scalable chronogram shows a cursor over moving markers representing every data item recorded on the selected channels, with time traces of parameter values as the replay proceeds. All data can be tabulated on a spreadsheet and exported in various formats for further analysis, summarizing, or reporting. Images and data can be exported for specific processing using existing HF tools such as Noldus Observer products.

The user can jump to events using bookmarks created during the test, the cursors in the chronogram, spreadsheet, or player control, or manually input a time to the search box, to access a particular interval, then zoom in to inspect or replay that sequence. Newly developed pattern-recognition routines allow quick comparisons of one driver's behaviour between different conditions, between groups of drivers under the same test, or aggregated drivers between different conditions.

D-BITE's modularity allows the user to add new hardware and software by means of simple control files which can be written or edited manually, or managed from the IPU and CTMU tools.

4 Conclusions

RoadSense has provided the methods and tools to conduct further HVI research, and to appraise new ADAS technologies in a more consistent and efficient manner than hitherto.

Questions for investigation remain over correlating simulator and equivalent road test measures, and the discrimination in marginal or critical overload and underload.

The authors are grateful to the contribution to the work of their colleagues in the partner organizations, principally : A.Hutton, J.Richardson, C.Pickering, D.Cousen, L.Andreone, E.Deregibus, N.Stelter, L.Krauß, LTrassoudaine, C.Thompson, S.Fu, W.Wang, A.Dugarry, Y.Huang, S.Barnes, D.Harris, R.Stewart, F,Nathan, E.Chin, A.Priez, Y.Page, R.Verschuren, J.Hogema, D.Meizel, P,Bonnifait, P.Crubille, M.Shawky, V.Cherfauoi, G.Brissot, and O.Bezet, and for the support of EC Directorate General – Research (Contract No. G3RD-CT-2000-00254). Views are those of the co-authors and do not necessarily reflect the views of their employers or of other partner organizations.

References:

1. Hogema, J.; *Lane Position Sensor,* internal document, TNO Human Factors, Soesterberg, The Netherlands.
2. Magney, P. (2004). *Automotive Interior: The Next Consumer Electronics Battleground.* http://www.telematicsresearch.com/PDFs/TMPress070804.pdf Telematics Research Group Inc.
3. Martens, M.H. (2003). Road familiarity and eyes-off-the-road: an exploratory study. *Experimental results driving study* (Report TNO TM - 03 - M033). TNO Human Factors, Soesterberg, The Netherlands.
4. Martens, M. (2005). Measuring Visual Scanning Behavior. http://www.tno.nl/defensie_en_veiligheid/militair_optreden/skilled_behaviour/measuring_visual_scanning/index.xml TNO, Soesterberg, The Netherlands.
5. Otzelberger, B.; Schmotzer, Ch. (1998). Nebentätigkeiten beim Autofahren unter besonderer Berücksichtigung des Telefonierens – eine Literaturstudie. *Kuratorium für Verkehrsicherheit*, Vienna.
6. Stanton, N.A.; Young, M. (2000). Modelling driver behaviour with adaptive cruise control. *3rd International Conference on Engineering Psychology and Cognitive Ergonomics*, Edinburgh.
7. Stewart, R.J.; Harris, D. (2003). Assessing the Effect of New Technology on Driver Behaviour – a Theoretical Model. in Harris D.; Duffy V.; Smith M.; Stephanidis C. (Eds.). *Human – Centred Computing: Cognitive, Social and Ergonomic Aspects.* publ. Lawrence Erlbaum, Mahwah, N.J.
8. Van Elslande P.; Alberton L.; Nachtergaele C.; Blanchet G. (1997). Scénarios-types de production de l'erreur humaine dans l'accident de la route, problématique et analyse qualitative. *Rapport INRETS* ***n° 218***, INRETS, Arcueil, France.
9. Van Winsum W.; Martens M.; Herland L. (1999). The effects of speech versus tactile driver support messages on workload, driver behaviour and user acceptance. *TNO-report TM-99-003* TNO Human Factors, Soesterberg, The Netherlands.
10. Chin, E.; Nathan, F.; Verschuren, R.M.A.F.; Hogema, J.H.; Martens, M. (2004). Tests on Measures for HVI evaluation (road and track). *Roadsense deliverable* **2.1 part II b**.

Temporal pattern analysis and its applicability in soccer

G.K. Jonsson[1], J. Bloomfield[2], R. Polman[2], K. Houlahan[3] & P. O'Donoghue[3]
1 Human Behaviour Laboratory, University of Iceland & Department of Psychology, University of Aberdeen, UK
2 Department of Sport Science, University of Hull, UK
3 School of Sport, University of Wales Institute Cardiff, UK

Abstract
It is important to employ the specificity principle in physical conditioning and coaching in order to enhance sports performance. To this end the specific requirements of the performance must be investigated to elicit a high degree of transfer into competitive scenarios. Time-motion analysis has been used as a method to detail the performance though previous studies have been limited in their approaches and have also failed to provide information concerning interaction of motion. The Bloomfield Movement Classification is recognized as the most comprehensive time-motion analysis method which includes a combination of 17 modes of motion, 14 directional categories, 4 intensity types and other specific instantaneous movement and sport-specific events including turns, swerves and on the ball activity. Together with T-pattern analysis an alternative method is offered as to detailing the movements performed by the players with an objective of re-producing the specificities of performance.

Keywords
Soccer, time-motion analysis, temporal patterns

1 Introduction
Skilled behavior is fundamental to all sporting activities. This often requires a dynamic co-coordinated process of perception, cognition and action with evidence identifying that skill acquisition and skilled performance share underlying mechanisms across the perceptual, cognitive, and motor domains [6, 22]. In order to produce skilled behavior it is important to identify the discrete requirements of sporting activities. However, it has been shown that coaches are unable to observe and recall all of the discrete incidents and activities that are required for a complete understanding of performance [9]. Skilled sporting behavior is enhanced through processes of learning, acquisition and physical conditioning. There are several recognized principles for these processes with one of the most important being specificity. Evidence from exercise physiology studies indicates that specific neurophysiological adaptations to physical conditioning are in direct relation to specific exercises performed [1, 10, 25]. Specificity of patterns of movement producing motor skills has been expressed as the fundamental constituent of any movement related to speed and power [15, 17, 23, 24]. Many sporting skills require a very precise moment and direction of force application and the neural pattern for the precise force application is acquired only by repetition of the desired skill [17]. Thus, precise movement patterns of specific motor skills are needed to effectively entrain appropriate motor programs for efficient co-ordination of muscular contraction.

Physical conditioning effects are limited to the physiological systems used and overloaded [10, 26]. It is recognized that each sport favors specific biological strategies for success based on the rules of the game and its physiological and biomechanical demands [13]. However, specific movement patterns must be first identified in order to apply the specificity principle in physical conditioning. This can be easier to accomplish in simple actions than in more complex actions. Although the physiological responses to performance and biomechanical analysis of several critical skills have been extensively investigated, it can be argued that there is insufficient knowledge regarding specific detail concerning the physical demands. The aim of the paper is to offer an alternative method of detailing the movements performed by the players, and finally the identification of hidden patterns within the movements through use of T-pattern detection with an objective of re-producing specific patterns of movement which can be used to enhance physical conditioning and coaching practices.

1.1 Time-Motion Analysis
Time-based analysis of match-play is a methodology which has been used to assess the physical and physiological demands of sport [7, 8, 28]. The main advantages of the paradigm are that it offers a non-intrusive method of analyzing performance during match-play, and that distances covered provide a crude measure of energy expended [18]. Motion modalities are classified according to type, intensity (or quality), duration (or distance), and frequency [20]. This data can then be applied to the construction of research models and in physical conditioning. However, it is critical to continue this form of research in order to re-assess the demands of the modern game due to the rapid evolution of the sport [21, 27].

As the game of soccer has evolved, the methods of analyzing performance have also developed (although perhaps not at the same rate). On review of the current studies performed there appear to be large discrepancies between studies in the distances covered by players largely due to methodological differences [20]. Methods have evolved from the simple use of dictated notation of players' movements and stop-watches [5, 18], to the current utilization of video recordings and computerized analyses [3, 14]. However, there is a need for a much more rigorous approach to produce the high level of specificity that is required to go about enhancing performance, in particular the considerable differences in the approaches used to classify motion modes.

A limitation of most of the previous work on the physical demands in soccer is that in general only crude measures have been used to index activity patterns. This is surprising as it has been discovered that there are 1000-1200 discrete movement changes (incorporating rapid and frequent changes in pace and direction) in a game, with the mean duration being 4.5-6s per movement [19]. With the use of computer-based time-motion analysis, it is now possible to control video images and to 'enhance sport specific analytical procedures' [11]. With several time-motion studies analyzing the time spent or distance covered walking, jogging, cruising, sprinting and backing it is true to say that this only provides a general analysis of the performance demands in soccer. Information regarding a complete set of movement categories, directions, intensities, turns and playing activity is needed to thoroughly understand and evaluate the performance requirements. In addition, it is necessary to gain an understanding into the interaction of these motions as this knowledge would make it possible to objectively design training that is highly specific to performance.

1.2 The Bloomfield Movement Classification

The 'Bloomfield Movement Classification' (BMC) [3] is a time-motion, computerized video analysis method involving a detailed account of motions, directions, intensities and events (turns, swerves, contact and on the ball activity). The BMC supplies codes for 14 modes of timed-motion, 3 ‘other’ non-timed movements, 14 directions, 4 intensities, 5 turning categories and 7 'On the Ball' activity classifications (see Table 1).

The Observer Version 5.0 (Noldus Information Technology, The Netherlands) was chosen as the software to perform the collection, management and presentation of the BMC as observational data could be collected, reviewed and edited with synchronized display of the corresponding video images [16]. Furthermore, the system requires a configuration to be independently composed consisting of states (continuous) and events (discrete) to define how observed behavioral data is to be notated. The Observer 5.0 configuration is comprised of behaviors (state or event) and allows for two further modifiers to be added used to describe the behavior. Once the initiation of a behavior was observed, a representative key pressed on a QWERTY (AT Enhanced) keyboard followed by the keys for the appropriate modifiers signify the entry and recording in a Event Log (Figure 1).

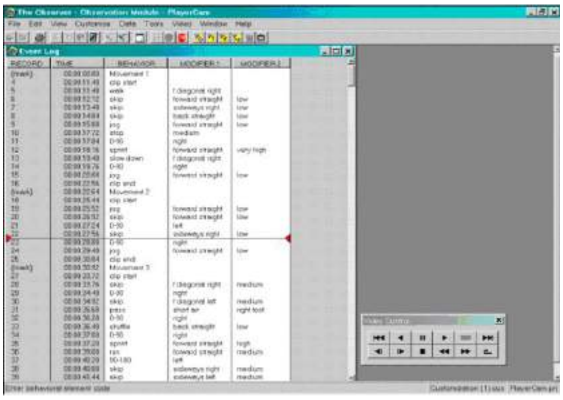


Figure 1. *The Observer 5.0 Event Log*

Footage from FA Premier League soccer players were collected during FA Premier League matches televised by Sky Sports Interactive 'PlayerCam' Service (British Sky Broadcasting Group, UK). For reliability of the method, 8 observers repeated three 5min observations each of 4 FA Premier League players This provided 32 sets of 3 observations for the purpose of assessing intra-observer reliability and 3 sets of 112 pairs of observations for the purpose of assessing inter-observer reliability. Each observer was presented with the BMC and the list of definitions and interpretations, which were verbally explained. Short sections (usually <15s) of the observation were initially viewed at 1x normal speed and replayed to perform data entry using a frame rate of 0.04s with video paused and scrolled for an accurate perceived start and finish of movement. All behaviours and modifiers were classed together in the tests of reliability. Kappa (κ) values for intra-observer agreement ranged from 0.79-0.92 and from 0.64-0.78 for inter-observer agreement. This is interpreted as a good to very good strength of intra-observer agreement and a good strength of inter-observer agreement [2]. Observers also followed a manual process of video playback to assess the quality of their data entry and make amendments if deemed necessary.

Table 1. *The 'Bloomfield Movement Classification' Behaviors and Modifiers.*

BEHAVIORS (Modifiers in parenthesis)	**MODIFIERS**
1. TIMED **Motion** Sprint (A+B), Run (A+B), Shuffle (A+B), Skip (A+B), Jog (A+B), Walk (A), Stand Still, Slow Down (A+B), Jump (C), Land, Dive (D), Slide (D), Fall, Get Up (B)	**Direction (A)** Forwards, Forwards Diagonally Right/Left, Sideways Right/Left, Backwards, Backwards Diagonally Right/Left, Arc Forwards Left to Right/Right to Left, Arc Backwards Left to Right/Right to Left, Arc Sideways Right/Left **Intensity (B)** Low, Medium, High, Very High **Jump (C)** Vertical, Forwards, Backwards, Sideways (E) **Dive (D)** Feet first, Head first **Turn (E)** Right/Left **Type (F)** Push, Pull, Pushed, Pulled, Other **Control (G)** Right/Left foot, Head, Chest, Thigh, Other **Pass/Shoot (H)** Long Air, Short Air, Long Ground, Short Ground, Other **How (I)** Right/Left Foot, Header, Backheel, Overhead, Other **Dribble (J)** Start, End **Touches (K)** Start,1-3, 4-6, 7-10, >10
2. INSTANTANEOUS (NON-TIMED) **Other Movement** Stop (B), Swerve (E), Impact(F+B) **Turns** 0^0-90^0 (E) 90^0-180^0 (E) 180^0-270^0 (E) 270^0-360^0 (E) >360^0 (E) **On the Ball Activity** Receive (G), Pass (H+I), Shoot (H+I), Dribble (J+K), Tackle, Trick, Other	

2 Temporal Pattern Analysis

A Temporal pattern (T-Pattern) is essentially a combination of events where the events occur in the same order with the consecutive time distances between consecutive pattern components remaining relatively invariant with respect to an expectation assuming, as a null hypothesis, that each component is independently and randomly distributed over time. As stated by Magnusson ‘that is, if A is an earlier and B a later component of the same recurring T-pattern then after an occurrence of A at t, there is an interval [*t*+d1, *t*+d2](d2≥ d1≥d0) that tends to contain at least one occurrence of B more often than would be expected by chance’ [12]. The temporal relationship between A and B is defined as a critical interval and this concept lies at the centre of the pattern detection algorithms.

Through use of the Theme 5.0 software package, pattern detection algorithms can analyze both ordinal and temporal data however, for the algorithms to generate the most meaningful analyses the raw data must be time coded i.e. an event must be coded according to time of occurrence as well as event type. The method of time-motion computerized video analysis therefore lends itself to the use of T-Pattern detection and through use of the

Bloomfield Movement Classification detailed and highly complex patterns that are specific to the performance of competition can be identified.

2.1 Soccer Match Analysis

T-patterning has been already been used to establish playing patterns in soccer [4]. The data show that a high number of temporal interactive play-patterns exist in soccer with the number, frequency and complexity of the detected patterns indicating that sport behavior is more synchronized than the human eye can detect. A typical within-team event pattern from the soccer analysis is shown in Figure 2. This figure displays a T-pattern that occurred three times during the first half of a European Championship qualifying match (1998). The pattern describes how player A (Zinedine Zidane) moves the ball towards the opponents goal by receiving the ball in, and then passing it out of, pitch zones 8, 11 and then 14 consecutively. Player A then completes the sequence by passing it on to player B who receives it in zone 15. The pattern describes an attacking movement through the middle of the pitch. Traditional frequency analysis of passing would have identified the ball reception and subsequent pass from each zone as discrete events but would not have linked the consecutive actions in the four zones.

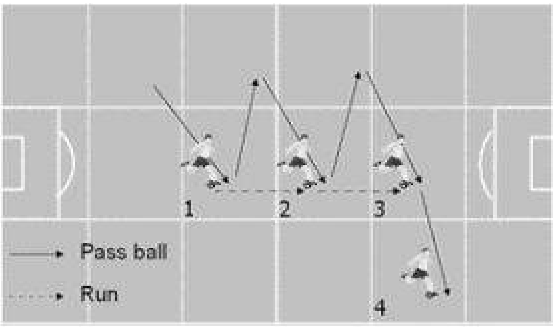


***Figure 2**. A temporal pattern relating to attacking movement of the ball through the centre of the pitch.*

Figure 3 also show a T-pattern from a club match (Liverpool – Sunderland, English Premiership) in which the pattern involves players from both teams and relates to the critical incident of shots on goal. The pattern occurred on three occasions during the second half of the match and includes two shots on goal within each pattern.

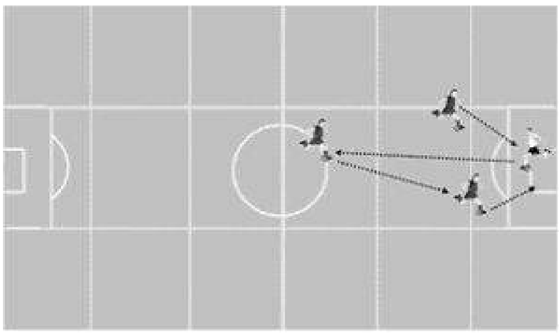


***Figure 3**. A T-pattern incorporating regularity in shots on goal in a F.A. Premiership match. The pattern includes 6 of Team A's 8 shots. Pattern occurrences 2 and 3 resulted in a goal.*

The total time period covered the three patterns therefore includes six shots by Team A which represents 75% of their total shot on target during the second half. Even more significantly two of the three pattern occurrences resulted in goals. It is difficult to conceive of coaching situations in which the type of information identified in Figures 2 and 3, and the further analyses that they may stimulate, would not be of value in enhancing coach knowledge. At the very least the analyses shown provide a perspective on team performance that is unattainable using traditional frequency counts of discrete events within a match. Another issue addressed in the study was to investigate the potential interrelationship between performance rating by coaches and the degree of structure in team performance. Several soccer coaches were asked to rate the performance of every player (on both teams) on a simple ten point Likert type scale. The data show that the coaches' ratings of team performance were significantly correlated to the number of patterns identified for each team ($r=0.81$, $p<0.05$).

3 Exemplar Data

Purposeful movement (PM) of the FA Premier League players (defined as any perceived purposeful and deliberate movement made by the player to influence match-play or change location on the field, regardless of intensity), was first assessed. A high degree of reliability existed with Kappa (κ) values ranging between 0.91-0.98 for intra-observer reliability and 0.85-0.96 for inter-observer reliability (with a pre-test mean of 0.89 and post-test mean of 0.92) each of which are interpreted as very good strength of agreement [2]. Subsequently, the PM was analyzed using the BMC. The efforts of one player were then further assessed for temporal patterns in the data through use of Theme 5.0. This resulted in two complex and highly significant t-patterns detected which are displayed in Figures 4 and 5.

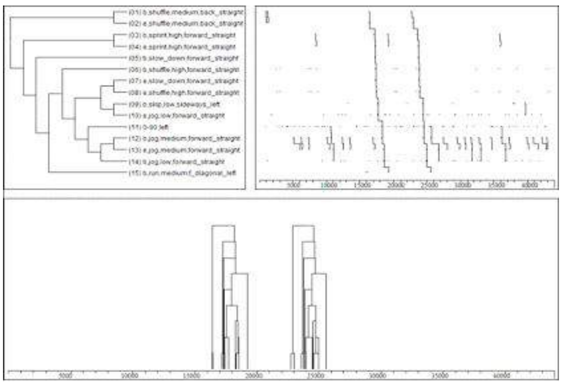


***Figure 4**. A T-pattern incorporating regularity of movements by a centre forward in a F.A. Premiership match.*

Figure 4 displays a pattern which identifies a varied range of motions but uniformity within the ranges of intensity, beginning with medium, progressing to high, and finally ending in low and medium intensity activity. This correlates strongly with the intermittent nature of the sport. The pattern begins with the player shuffling backwards and then sprinting forwards and slowing down in which he shuffles at high intensity suggesting he slows quite abruptly from his sprint. From this point he skips sideways left at low intensity perhaps preparing for another high intensity burst, turns left and jogs forward at low intensity and gradually increases pace into a run changing direction by moving diagonally left to complete the complex pattern.

Figure 5 displays a pattern which begins at high intensity and ends with low intensity movement. The pattern is simpler than that in Figure 4 and begins with the player sprinting forwards submaximally, slowing down and shuffling at high intensity, again suggesting he slows quite abruptly from his sprint. This deceleration phase should be commonly adopted into physical conditioning.

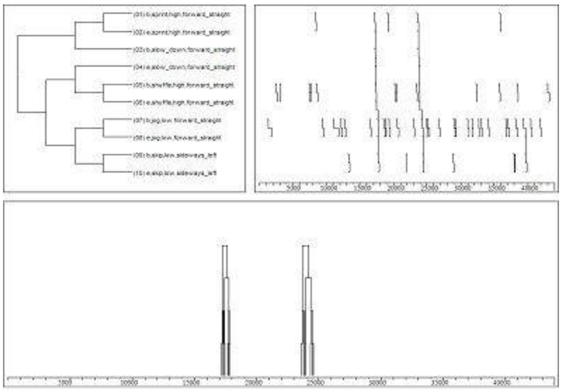


Figure 5*. A T-pattern incorporating regularity of movements by a centre forward in a F.A. Premiership match.*

4 Conclusions and future directions

The potential for use of T-pattern detection in soccer is tremendous. The identification of patterns that are not identifiable through simple observation has great benefit not only in matchplay but also in establishing the physical demands through time-motion analysis. With regards the latter, highly specific physical conditioning practices can be employed through the use of the Bloomfield Movement Classification and Theme 5.0 which will enhance the condition of the players and optimize time spent in training. A comprehensive study is therefore required investigating the demands between the different positions played in soccer as well as different time phases in a match. This study is ongoing using the BMC and the 'PlayerCam' facility provided by Sky Sports Interactive Service (British Sky Broadcasting Group, UK) with televised FA Premier League matches. It is aimed to observe a total of 54 players - 3 from each position (defender, midfielder, forward) and time period (0-15, 15-30, 30-45, 45-60, 6-75, 75-90) and to investigate the differences within hidden pattern analysis. Preliminary investigation of pattern complexity between player positions suggests that a higher number of different patterns and pattern occurrences are detected for Defenders than Forwards and Midfielders. The same also seems to apply for length of patterns. These findings, and their significance, need further examinations.

Once these physical demands have been identified, and the complex hidden patterns that occur it becomes possible to perform further research into establishing the physiological and biomechanical demands which will further assist the enhancement of coaching and physical conditioning practices.

References:

1. Ackland, T.R. and J. Bloomfield (1996). Applied anatomy. In: (Edited by) J. Bloomfield, P.A. Fricker, and K.D. Fitch, *Science and Medicine in Sport*, 2nd Edn. Victoria, Australia, Blackwell Science. pp. 2 - 30.
2. Altman, D.G. (1991). *Practical Statistics for Medical Research.*, pp. 404.
3. Bloomfield, J., R.C.J. Polman, and P.G. O'Donoghue (2004). The 'Bloomfield Movement Classification': Motion analysis of individuals in team sports. *International Journal of Performance Analysis in Sport,* **4**(2), 20 - 31.
4. Borrie, A., G.K. Jonsson, and M. M.S. (2002). Temporal pattern analysis and its applicability in sport: An explanation and exemplar data. *Journal of Sports Sciences*. **20**, 845 - 852.
5. Brookes, J.D. and J.E. Knowles (1974). A movement analysis of player behaviour in soccer match performance. *British Proceedings of Sport Psychology*. 246 - 256.
6. Carlson, R.A. and R.G. Yaure (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16**, 484 - 496.
7. Drust, B., T. Reilly, and E. Rienzi (1998). Analysis of work rate in soccer. *Sport Exercise and Injury*, **4**, 151 - 155.
8. Ekblom, B. (1986). Applied physiology of soccer. *Sports Medicine*, **3**(1), 50 - 60.
9. Franks, I.M. and G. Miller (1986). Eyewitness testimony in sport. *Journal of Sport Behavior*, **9**, 39 - 45.
10. Holloszy, J.O. and E.F. Coyle (1984). Adaptations of skeletal muscle to endurance exercise and their metabolic consequences. *Journal of Applied Physiology*, **56**, 831 - 838.
11. Hughes, M.D. (2003). Notational Analysis. In: (Edited by) T. Reilly and A.M. Williams, *Science and Soccer*, 2nd Edn. London, E & FN Spon. pp. 245 - 264.
12. Magnusson, M.S. (2000). Discovering Hidden Time Patterns in Behavior: T-Patterns and their Detection. *Behavior Research Methods, Instruments and Computers*, **32**(1), 93 - 110.
13. McCall, G.E., et al. (1999). Acute and chronic hormonal responses to resistance training designed to promote muscle hypertrophy. *Canadian Journal of Applied Physiology*, **24**(1), 96 - 107.
14. Mohr, M., P. Krustrup, and J. Bangsbo (2003). Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences*, **21**(7), 519 - 528.
15. Morrissey, M.C., E.A. Harman, and M.J. Johnson (1995). Resistance training modes: Specificity and effectiveness. *Medicine and Science in Sport and Exercise*, **27**(5), 648 - 660.
16. Noldus, L.P.J.J., et al. (1999). The Observer Video-Pro: new software for the collection, management and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments and Computers*, **32**, 197 - 206.
17. Poliquin, C. (1990). Theory and methodology of strength training: Part 4. *Sports Coach*, **13**(3), 29 - 31.
18. Reilly, T. and V. Thomas (1976). A motion analysis of work-rate in different positional roles in professional football match-play. *Journal of Human Movement Studies*, **2**(2), 87 - 89.
19. Reilly, T. (1990). *Physiology of Sports.* London, E & FN Spon.
20. Reilly, T. (2003). Motion analysis and physiological demands. In: (Edited by) A.M. Williams and T. Reilly, *Science and Soccer*, 2nd Edn. London, E & FN Spon. pp. 59 - 72.
21. Rienzi, E., et al. (2000). Investigation of anthropometric and work-rate profiles of elite South American international soccer players. *Journal of Sports Medicine and Physical Fitness*, **40**, 162 - 169.
22. Rosenbaum, D.A. (1987). Successive approximations to a model of human motor programming. In: (Edited by) G.H. Bower, *The psychology of learning and motivation*. San Diego, Academic Press. pp. 1 - 52.
23. Sale, D.G. and D. MacDougall (1981). Specificity in strength training: A review for the coach and athlete.

Canadian Journal of Applied Sport Science, **4**, 56 - 59.
24. Sale, D.G. (1992). Neural adaptations to strength training. In: (Edited by) P.V. Komi, *Strength and power in sport*. Oxford, Blackwell Scientific Publications. pp. 249 - 265.
25. Saltin, B., et al. (1977). Fiber types and metabolic potentials of skeletal muscles in sedentary man and endurance runners. In: (Edited by) P. Milvy, *Marathon: Physiological, Medical, Epidemiological, and Psychological Studies*. New York, New York Academy of Sciences. pp. 3 - 29.
26. Schmidtbleicher, D. (1992). Training for power events. In: (Edited by) P.V. Komi, *Strength and power in sport*. Oxford, Blackwell Scientific Publications. pp. 381 - 398.
27. Strudwick, A., T. Reilly, and D. Doran (2002). Anthropometric and fitness profiles of elite players in two football codes. *Journal of Sports Medicine and Physical Fitness*, **42**, 239 - 242.
28. Tumilty, D. (1993). Physiological characteristics of elite soccer players. *Sports Medicine*, **16**(2), 80 - 96.

SocioXensor: Measuring user behaviour and user eXperience in conteXt with mobile devices

I. Mulder [1], G.H. ter Hofte [1], J. Kort [2]
[1] *Telematica Instituut, Enschede, The Netherlands*
[2] *Mobile: User Centered Innovation, TNO Information and Communication Technology, Groningen/Delft, The Netherlands*

Abstract
Mobile devices tend to travel along with people wherever they are and whatever they are doing, and consequently enter various social contexts of that person. This literally puts these devices in an ideal position to capture several aspects of social phenomena. We are currently designing and implementing SocioXensor, an extensible toolkit that exploits the hardware sensors and software capabilities of contemporary mobile devices like PDAs and smartphones to capture objective data about human behaviour and social context (e.g., proximity and communication), together with objective data about application usage and highly subjective data about user experience (e.g., needs, frustrations, and other feelings). Thus, we provide the social sciences with a research instrument to gain a much deeper, detailed, and dynamic insight into these phenomena and their relations, which in turn can inform the design of successful context-sensitive applications.

Keywords
In-situ measurements, context-sensitive applications; mobile devices; logging; experience sampling

1 Introduction
Man is a social being, continuously and dynamically adapting to his social context, and increasingly supported by advances in mobile technology. Current research into context-sensitive applications stresses the relevance of using context information in applications in order to improve desirable properties such as social translucence (see for example, [3,7]). The massive success of context-mediating applications such as Presence and Instant Messaging applications [8] is a further testimony to the importance of using context information in context-sensitive applications. Mobile applications typically operate in very dynamic contexts of end-users, which makes it even more relevant for mobile applications to be sensitive about context.

Despite occasional design successes such as Presence and Instant Messaging applications, researchers are still lacking a systematic understanding *which* context information is relevant in *what* kind of situation and *which* kind of applications. At the same time, designers of context-sensitive applications (also referred to as context-aware computing and communication, ubiquitous computing and ambient intelligence) face design issues like: selecting which context information should be conveyed or aggregated to other human users (who then interpret that information), and selecting which context information is predictive enough such that it can be interpreted by applications. Although many methods exist to study social phenomena, including interviews, focus groups, surveys, laboratory experiments, ethnography, diary studies, logging and experience sampling, obtaining the right answers to design context-sensitive applications proves to be rather complicated [6].

In this article we describe "SocioXensor", a research instrument for field trials in experience and application research in the area of context-sensitive applications. In particular, SocioXensor aims to strengthen logging and experience sampling by combining them with contemporary mobile and wearable devices such as smartphones and PDAs. Such devices are personal in nature and stay and travel together with one person most of the time and consequently enter various contexts of that person (e.g., home, work, and mobile context). The hardware sensors and software capabilities of such devices provide ample opportunities to capture objective data about application usage, human behaviour and the context in which this takes place, together with sampling of subjective user experience. In this way, SocioXensor allows scientists to gain a much deeper, quantitative and dynamic insight into the relations between user experiences, human behaviour, context, and application usage. SocioXensor can be applied for formative evaluation, which results in insights that can be used by designers and developers to create successful context-sensitive applications. It can also be applied for summative evaluation, for instance to evaluate a specific application.

In the remainder of this article, we first describe SocioXensor in more detail. Then, we briefly describe how SocioXensor can be applied. We conclude with a brief summary.

2 SocioXensor
The SocioXensor research instrument is an extensible software toolkit for capturing objective data about application usage, human behaviour and the context in which this takes place, together with sampling of subjective user experiences, at any time, in any location. The core idea of SocioXensor is not to bring the people to the lab, but to bring the lab to the people by using wearable, personal mobile devices like smartphones as the primary data capturing device, supported by sensors and beacons in other infrastructures where possible and appropriate (e.g., in mobile testbed networks). Put differently, SocioXensor seeks to maximize the validity of data collected by focusing on in-situ data collection, and therefore, avoiding or minimizing retrospective recall present in other self-report techniques such as surveys and interviews. As illustrated in Figure 2, SocioXensor can be more obtrusive than logging, but is typically less obtrusive than direct observation methods such as ethnography (which allow for very rich data capturing) or lab experiments.

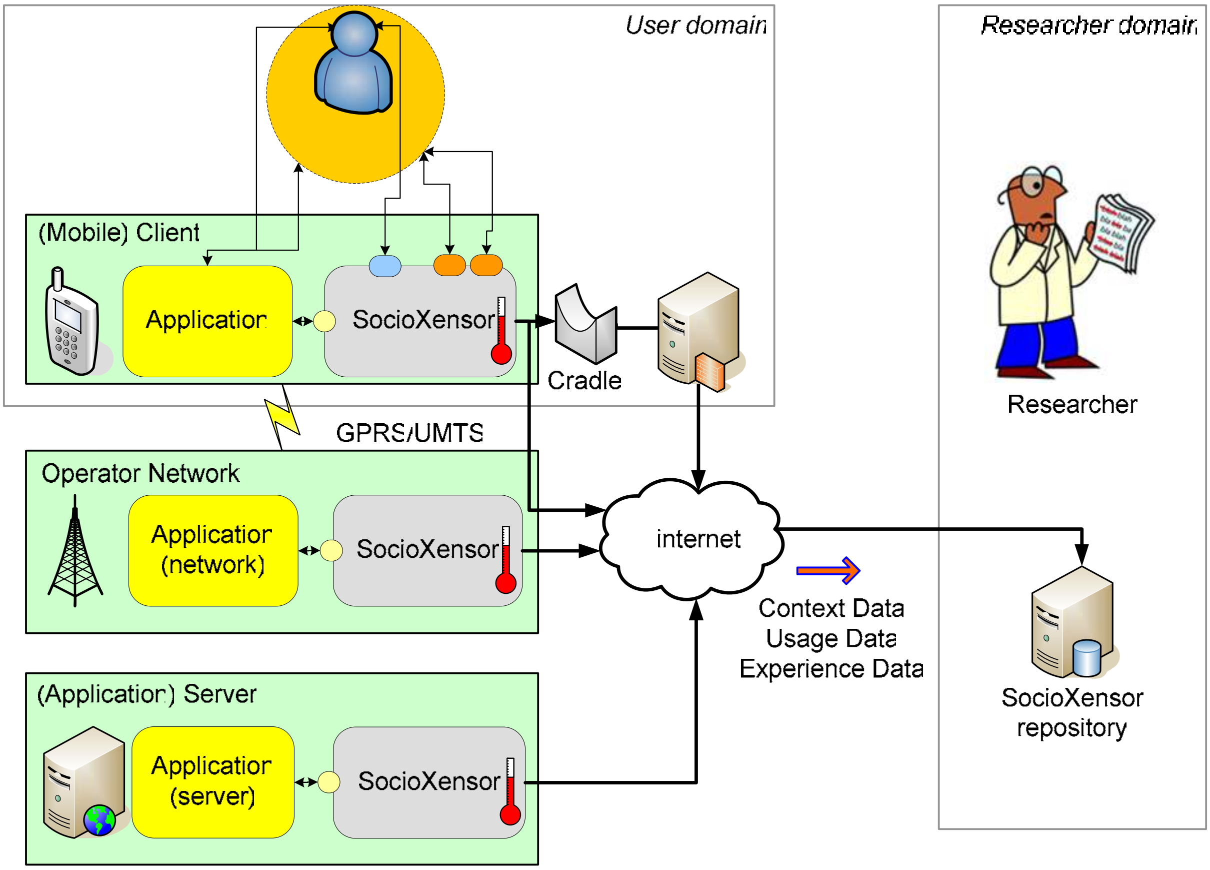


Figure 1. *SocioXensor high-level architecture.*

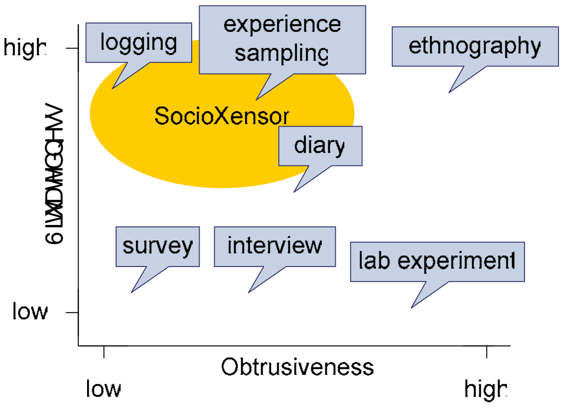


Figure 2. *SocioXensor target area in terms of obtrusiveness and situatedness, compared to other methods.*

More specifically, SocioXensor typically can collect data at times and locations that would be impractical or very costly with ethnography and lab studies, while maximizing the chance that subjects exhibit their natural behaviour in their natural context.

2.1 Type of data collected

The SocioXensor instrument and method focuses on capturing three types of data:

- *User Experience data*: subjective information such as opinions and feelings, which can be obtained using an experience sampling procedure [2]. For example, by notifying the user of a sample (a survey with several closed questions defined by a researcher), according a pseudo-random schedule defined by a researcher (e.g., an inter-sample time uniformly distributed between 45 minutes and 1 hour and 45 minutes, with the earliest sample of a day not before 8 AM and the latest sample of a day not after 10 PM). It is also possible to notify the user on a schedule based on human behaviour and context data (see below); for example, notify the user of a sample shortly after a telephone conversation is completed.
- *Human behaviour and context data*: raw, objective data about human behaviour and context (e.g., location, proximity, activity and communication) that is captured unobtrusively through device technologies on contemporary mobile devices such as PDAs and smartphones (e.g., GSM Cell-IDs, GPS location data, Bluetooth device detection, audio microphone, call logs, contact data, and calendar data). This raw behaviour and context data can be used in later analysis to find relations and predictiveness with user experiences: for example, which raw context data predicts low tolerance for interruptions? Which raw context data predicts the relevance of other colleagues that might be able to help you given your current context?
- *Application usage data*: raw, objective data about the usage of the application that is being studied. The raw data may range from low-level keystrokes and screens to high-level application events. Note that in formative evaluation usage of SocioXensor, this type of data is typically not collected.

2.2 Architecture

The SocioXensor architecture prescribes what client, network and server elements are involved in SocioXensor, and how they interact (see Figure 1). The SocioXensor architecture also prescribes how (third-party) plug-ins for context sensors, experience samplers, and application usage sensors should interact with the SocioXensor data manager, which takes care of local storage of captured data and uploading data to a central SocioXensor repository via appropriate media at appropriate moments (see Figure 3 for an illustration of this plug-in architecture on a mobile client).

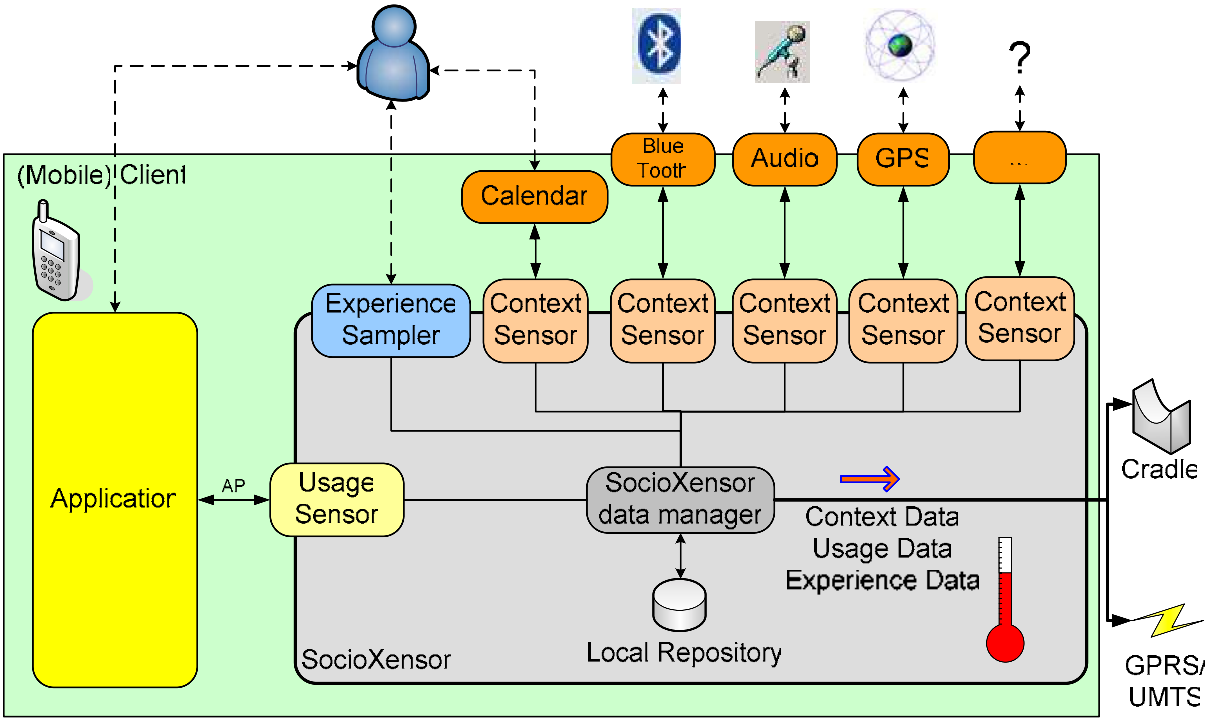


***Figure 3**. SocioXensor mobile client architecture.*

2.3 Approach

SocioXensor fits into en evolutionary prototyping research and design strategy, and can be used to get answers in field trials to formative and summative evaluation questions, such as:

- In which contexts do information and communication needs arise, and how often?
- Which (combination) of context information is relevant for an application?
- In which contexts are application features actually used?
- Which (combination of) context information predicts a user experience of an application?
- Did the user experience of an application improve? In which contexts?

The SocioXensor method guides researchers in their choices which questions to ask, according to which schedule to obtain user experience data, which human behaviour and context data sensors to use and which application usage data to log.

3 An example

In this section we argue how SocioXensor could have be applied, based on an earlier exploratory study we did into context factors that predict availability at home, at work, and on the move, in which we only used experience sampling (for details, see [4]).

In previous research, we explored how technology can help people to communicate with the right person at the right time at the right place via the right communication channel (as opposed to communication with "anyone, anytime, and anyplace", as often promised in marketing of mobile and ubiquitous communication technologies). In an exploratory study, we hypothesized that availability for interruption depends on:

- Conversation and type of conversation medium;
- Being together with others;
- Location (work, home, on the move).

It is hard for humans to remember when and under which circumstances one was available for interruption. Therefore, we decided to study availability for interruption with an in-situ research method, which minimizes the problems of retrospective recall. In this study, we used the experience sampling method to get insight in: "Which context factors are most useful in context-aware communication applications that convey availability of people for communication, not only at work and at home, but also on the move?".

Each sample of our experience sampling instrument consisted of 4-6 questions that could be answered by an experienced participant in less than 15 seconds. The questions for each sample and multiple-choice answers can be found in Table 1. All questions and answers were presented originally Dutch; all respondents understood Dutch; although two subjects were non-native speakers.

***Table 1**. Questions and answers of the FRUX ESM interruptability instrument.*

nr	Question	answers				
1	How interruptible are you now?	1 entirely not	2	3	4	5 entirely yes
2	Are you in conversation?	yes	no			
3	How are you in conversation?	face to face	via fixed telephone	via mobile telephone	via Instant Messaging	otherwise
4	Where are you now?	at home	at the <company> office	in transit	somewhere else	
5	Where at the <company> office?	My own office room	office room of a colleague	hallway/ hall/ stairs	meeting room	somewhere else
6	How are you in transit?	On foot	on a bicycle	in a car	with public transport	otherwise
7	With how many people are you? (incl. yourself)	1	2	3	4-6	7 or more

All questions were presented on a PDA, as illustrated in ***Figure 4***.

Figure 4 Example of a question used in the experience sampling study.

Each subject participated for 7 days. For all subjects and for all days, both weekday and weekend, samples started at the earliest on 8 AM and ended the latest at 10 PM, which corresponds to 14 hours per day. Samples were scheduled to be at least 45 minutes and at most 1 hour 45 minutes (1h45) apart, according to a uniform random distribution.

Examples of results we obtained from this study include the following. Respondents reported 37% of their time to be in conversation, amounting to 36h10 on average of the 97h40 we sampled per respondent per week. Most conversation (34% of all samples) concerned face to face conversation, amounting to 33h12 on average per respondent per week. Moreover, there were some medium correlations [1]: namely between a person's availability for interruption and respectively being in conversation (r = - .369), face-to-face communication (r = - .343), and being in a meeting room (r = - .293).

Such results can be used to decide which context sensors seem to be the best predictors and hence should be applied in a context-sensitive application. For example, audio sensors might be appropriate to capture "being in a conversation". In the exploratory study we did, we were especially interested in those variables that are good predictors and at the same time have low costs and are easy to be implemented in a context-sensitive tool.

With SocioXensor, we could have obtained the same data with even less effort of our users, e.g., by asking only the first question about availability for interruption and using context sensors to capture data about the other six questions. Moreover, SocioXensor can provide more reliable answers about the predictiveness of particular sensors (e.g., How predictive is a bluetooth-based proximity sensor for interruptability? How predictive is a audio conversation detector for interruptability?).

4 Summary

SocioXensor provides the social sciences with an instrument to gain a much deeper, detailed and dynamic insight into these phenomena and their relations, which in turn informs the design of successful context-sensitive applications. In addition to such formative evaluation for design, SocioXensor can also be extended with modules allowing summative evaluation of application usage in context. SocioXensor not only strengthens this crucial reciprocal link between evaluation and design, but also provides benefits to a broader audience of scientific communities including medical/biological sciences such as epidemiology.

Acknowledgements

This paper has been written in the project "Freeband User Experience" (FRUX), which is part of the research programme Freeband Communication (www.freeband.nl), which aims at the generation of public knowledge in advanced telecommunication technology and applications in the Netherlands. FRUX is a joint effort of Ericsson, ISC (Dutch Police ICT service organization), Telematica Instituut, TNO ICT, Delft University of Technology, VU University medical center, Waag Society, and Web Integration.

References

1. Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*. 2nd Ed. Hillsdale, NJ: Erlbaum.
2. Csikszentmihalyi, M.; Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, **175**, 526-536.
3. Erickson, T.; Kellogg, W.A. (2000). Social translucence: an approach to designing systems that support social processes. *Transactions on Computer-Human Interaction (TOCHI)*, **7**, 59-83.
4. ter Hofte, G.H.; Ebben, P. (2004). *Experience Sampling: Method, Tool, Pilot and Experiences*. Report TI/RS/2004/072 (Freeband/FRUX/D1.5.1). Telematica Instituut, Enschede. Available from https://doc.telin.nl/dscgi/ds.py/ViewProps/File-45237
5. Intille, S.S.; Rondoni, J.; Kukla, C.; Ancona, I.; Bao, L. (2003). A Context-Aware Experience Sampling Tool. In *Proceedings of CHI2003*, pp. 972-97.
6. Kort, J.; Steen, M.G.D.; de Poot, H.; ter Hofte, G.H.; Mulder, I. (2005). Studying usage of complex applications. In this issue.
7. Schilit, B.N.; Hilbert, D.M.; Trevor, J. Context-aware communication, In *IEEE Wireless Communications*, **9**, 5, (Oct. 2002), 46-54.
8. de Vos, H.; ter Hofte, G.H.; de Poot, H. (2004). IM [@Work] - Adoption of Instant Messaging in a Knowledge Worker Organisation. In *Proceedings of the 37th Hawaii International Conference on Systems Sciences (HICSS'04)*.

Kinetic T-patterns detection in conversation: comparison between Icelandic and Italian style of communication

A. Vescovo, A. Agliati, L. Anolli
Department of Epistemology and Learning Ermeneutics, University of Milan Bicocca, Milan, Italy

Abstract
Human behavior measuring represents a "challenge" for psychologists, since it calls for management of the plurality of behavior systems and their variability. This paper focuses on observation, as the specific device used for detection and measurement of micro-units of interactive nonverbal behavior. Observation was performed by Theme, the software used for nonverbal behaviors and T-patterns detection. To study the cultural influence on nonverbal behavior, participants were twenty Italian and twenty Icelandic couples, involved in a cooperative task. Different kinds of gestures emerged according to culture. Italians displayed gestures like "bag-hand", while Icelanders manifested "ball-hands" and "pianist". Moreover, Icelandic interactions were characterized by a more complex structure of repeated events. On the whole, this methodological approach seems to be functional for nonverbal behavior "measurement" in conversation. Furthermore, it seems to provide an analysis procedure that considers cultural variables, thus representing an attempt to face the challenges coming from behavior complexity.

Keywords
Observation; cultural differences; gestures; Theme; T-patterns.

1 Introduction
Human behavior can be considered a complex system, for its variable nature and its multicomponent character, as it is deeply influenced by different factors and it is defined by a plurality of systems (verbal, vocal, nonverbal). Firstly, cultural standards, social rules, situational elements form the context to which human behavior should adapt. Moreover, human behavior also change in function of individual variables that are related both to stable features like personality traits and dispositions, and to contextual factors like actual mood.
Behavior complexity is also due to the plurality of systems: human behavior, in particular, is displayed by verbal, nonverbal, and vocal features, that all together contribute in synergy to meaning definition [1].
In the light of this complexity, human behavior measuring may be considered a sort of intriguing "challenge" that has to consider and manage all these variability sources. Consequently, behavior measuring should allow to detect cues coming from these different systems and to integrate them in a meaningful process. Different measurement devices can be used to detect human behavior features: observation, psychometric tests, self-reports focus on different aspects of behavior, at different levels (micro/macro). In particular, this paper will focus on observation, as the specific device used for detection and measurement of micro-units of interactive nonverbal behavior.

1.1 Human behavior rhythm and regularity
Behavior complexity is not synonymous of chaos: on the contrary, different behavioral modalities (verbal, non verbal, vocal, etc.) are displayed in a well organized and coordinated way, to give origin to meaningful behavior. Moreover, behavior in interaction has a defined temporal organization. As Eibl-Eibesfeldt states, "behavior consists of patterns in time. Investigations of behavior deal with sequences that, in contrast to bodily characteristics, are not always visible" [3].
As these behavioral patterns are hidden in interaction time flow, they can be detected only thorough specific technical procedures for interaction analysis. Theme software [11,12] is the device used in the study described in this paper to detect hidden time patterns of nonverbal behavior units in a collaborative interaction.
Hidden time patterns recur during interaction flow and give rise to a behavioral regularity that is the origin of interaction rhythm. It is a sort of "behavioral dance" [8] that constitutes the interaction background and characterizes it, as behavior time structure and rhythm remains quite stable during interaction flow.

1.2 Culture and time structure of behavior in conversation
Conversation may be defined as the best place to detect and analyze behavior regularities, since it can be considered as a universal communication system, characterized by a deep and intrinsic organization, which follows given social and cultural standards [1,5]. Consequently culture, should be described as a strong behavioral organizer, that provides individuals with standards and resources to follow for their participation to social life, and to everyday conversations in particular. Cultural influence is displayed in verbal and nonverbal behaviors' variability; different experimental studies have outlined cultural specificities for gestures, both in typology and in production frequency during conversation [1,6,7,16].
In addition, culture seems also to influence interaction's time-structure, as revealed by researches about chronemics [9,10].
Nonverbal behaviors' time patterns detection and comparison in two different cultures' conversations represents the main object of this study. Observation was the elective measuring device used for nonverbal behavior recognition and coding, while Theme software was used for the analysis of the nonverbal behaviors observed and for time patterns detection.

2 Nonverbal behavior observation in conversation
Nonverbal behavior micro-analytic observation has been performed through the preliminary definition of some nonverbal behavior units (minimal coded action units), portioned in categories, that became the constituents of a nonverbal behaviors grid of detection. This grid, named

VVT (Variable Value Table) is an input for Theme Coder, the software used for nonverbal behaviors coding.

2.1 T-patterns detection through Theme software

The Theme Software [12,13] is a tool to facilitate the study of the structure of face-to-face human interactions, searching for a particular type of repeated time patterns, called T-patterns (Type-patterns).

Theme package is constituted by Theme Coder, a specific software used to perform interaction observation. It collects and memorizes the beginning and ending of every occurrence of the behavior units detected by a human coder, on the base of the specific behaviors grid utilized (VVT). In this study Theme Coder has been used to perform observation of nonverbal behavior units in interaction

Once behavior detection is completed, it is entered in Theme software, that generates two different data as output: the frequency of every behavioral unit observed in the interaction (IR data) and the number of T-patterns, defined as sequences of behavioral units occurring in the same order and with approximately the same time distance between consecutive elements.

Figure 1 and 2 show recurrent hierarchical organized patterns of elements with a statistically significant tendency of the sub-patterns and/or primitives to recur in the same order and with similar temporal distances between them.

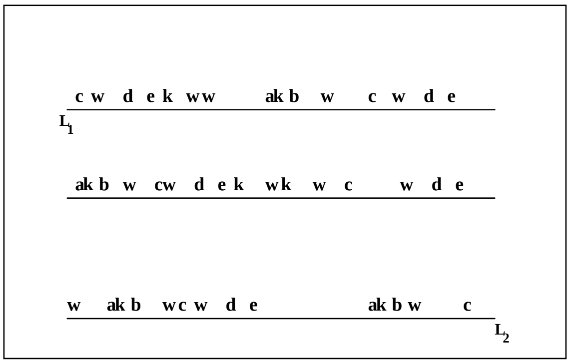


Figure 1. *Stream of observed behavioral units [13, modified].*

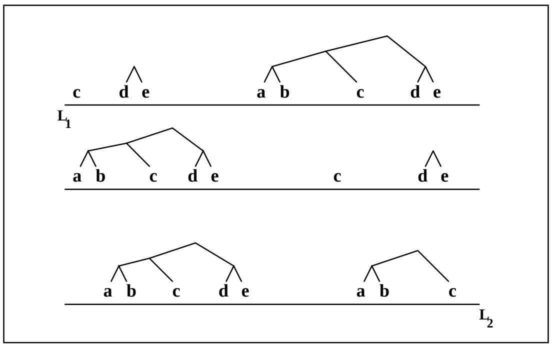


Figure 2. *Repeated hierarchical T- patterns [13, modified].*

3 Current study

The present study's general was is the observation at a micro-analytic level of nonverbal behaviors displayed in conversation. In particular, cultural influence on nonverbal behavior production and on conversation rhythm has been analyzed. To this end, participants were couples belonging to two different nationalities, specifically twenty Italian and twenty Icelandic couples. The half of individuals were males and half of them females; to verify if individuals' knowledge could have an influence on nonverbal interactive behavior, twenty couples were formed by friends, while twenty couples were formed by unknown participants; all of them were between twenty and thirty years.

3.1 Setting and procedure

To ensure a good control on contextual variables, this study was carried out in a laboratory setting, partly in the Communication Psychology Lab at Catholic University in Milan and partly in the Human Behavior Laboratory of Reykjavik University. Participants were involved in a cooperative task: each couple was asked to think up and to talk about an advertising of a bubble bath, as if they were creative and advertising experts.

Their conversations had a fifteen minutes fixed duration and were audio and video recorded.

Each video sequence was then coded frame by frame using Theme Coder software; interaction observation was carried out by two different human coders, in order to obtain a high reliability. A reliability of 85% was found according to McGrew formula [14]. The coding was done separately for each behavior unit, and both the beginning frame and the ending frame of each behavior unit was marked.

The grid used for interactions observation is represented in figure 3.

Category	**Behavior units description**
Hands	rotation, hand contact, crossed hands, hand on face, hand on body, hand on chair, hand on hair, palm inward, palm forward, palm down, palm upward, palm back, palm outside, fist, bag-hand, mirror-hands, rubbing hands, ax-hand, hitchhiking, ring, hands away, ball-hands
Fingers	pointing, counting, negation, pianist
Arms	arms upward, crossed arms, arm forward, open arms, dangling arms, arms back, bend arms
Head	head forward, head down, head back, tilt head l/r, tilt head up/down, negation by head
Trunk	trunk forward, tilt trunk l/r, trunk back, swinging
Face	smile, laugh
Gaze	avert gaze
Shoulders	shoulders shrug
Turn	Speak

Figure 3. *The nonverbal behaviors grid used in this study.*

The grid is formed by forty-eight behavior units, portioned in nine categories, and describes different types of body movements related to hands, arms, head, fingers, trunk, shoulders, gaze direction and face mimics; also turn-taking was considered as a behavioral unit. The mostly behavior units were taken from grids used in previous researches on nonverbal behavior [2,4,8,15,16].

Behaviors coded were then analyzed with Theme, in order to detect T-patterns.

Only a segment was taken into consideration for each interaction, *i.e.* the three central minutes; this limit was due to the long amount of time necessary for behavior coding.

3.2 Results

Two different analysis were carried out on data collected. A first multivariate analysis of variance was performed on IR data for each behavior unit coded. A significant main effect was found for culture variable ($F_{1,29}$ = 14.01, p < .0001); in particular, culture seems to influence gestures production, both about frequency and typologies. In particular, Italian participants displayed a higher gestures frequency than Icelandic ones.

Moreover, different kinds of gestures emerged according to cultural differences. Italian participants displayed specific gestures like “bag-hand” and “ring” that were never used by Icelanders. On the other hand, Icelandic participants manifested “ball-hands” and “pianist”, that were never present in Italian participants. The complete analysis output is not reported here due to the lack of space.

The second step consisted in a qualitative analysis of the T-patterns detected in interactions. By way of example, two different T-patterns that represent two typical gestures that seem to be culturally discriminators will be reported.

Figure 4 shows a T-pattern detected during one Italian couple interaction in which the “bag-hand” gesture is displayed. The correspondent interaction frame is shown in figure 5.

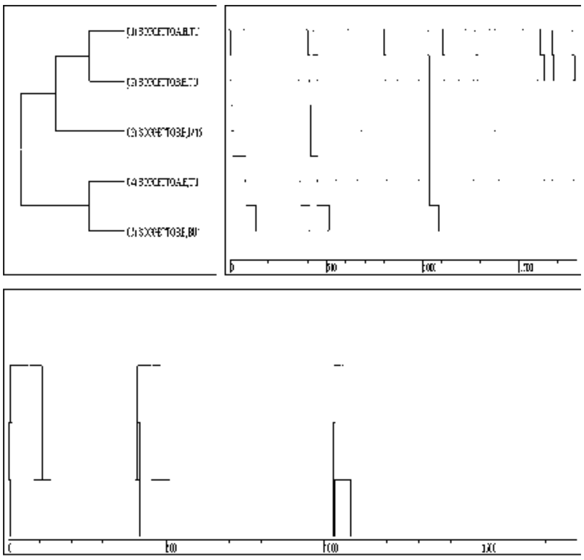


Figure 4. *T-pattern detected in an Italian interaction between friends: (1) subjectA, b, turn; (2) subjectB, b, turn; (3) subjectB, b, bag-hand; (4) subjectA, e, turn; (5) subjectB, b, trunk forward.*

Figure 5*. Interaction frame of an Italian couple representing the “bag-hand” gesture, performed by the participant on the left.*

Figure 6 shows a typical gesture detected in Icelandic people, “pianist”. T-pattern’s length and levels analysis reveals that the sequence is quite complex.

Figure 7 represents the correspondent interaction frame; it can be noticed that the “pianist” gesture is performed by both interactants at the same time.

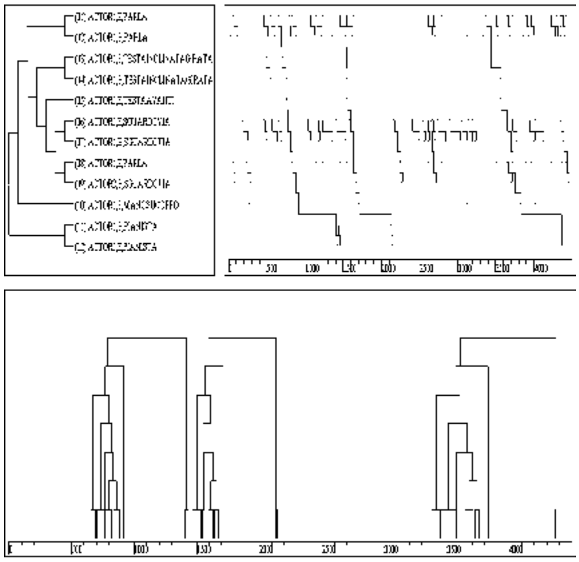


Figure 6. *T-pattern detected in an Icelandic interaction between friends: (1) subjectB, e, turn; (2) subjectA, b, turn; (3) subjectA, b, tilt head; (4) subjectA, e, tilt head; (5) subjectA, b, head forward; (6) subjectA, e, look away; (7) subjectB, b, look away; (8) subjectA, e, turn; (9) subjectB, b, look away; (10) subjectA, e, hand on body; (11) subjectA, b, pianist; (12) subjectA, e, pianist.*

Figure 7. *Interaction frame of an Icelandic couple; it is represented the “pianist” gesture, performed by both participants at the same time.*

T-patterns length and levels analysis outlined that Icelandic couples interactions were characterized by a more regular rhythm than Italian ones, because their T-patterns were characterized by a more complex structure of repeated events. In addition, Icelandic style of conversation seems based on a smoother rhythm, shown by a larger repetition of the same behavior units in the critical interval of time.

3.3 Discussion

The study described in this paper suggests a methodological approach to nonverbal behavior measurement based on observation, and provides an attempt to detect the hidden time structure of conversation. This measuring device on one hand allowed behavioral units detection with a minimum interference on behavior production’s spontaneity. This should be particularly interesting if applied to nonverbal behavior, which is itself less voluntarily controlled.

On the other hand, it introduces a way towards behavioral structure description, strictly linked with interactive rhythm.

In such a way, behavior regularity emerged as a stable feature of human interactions.

Moreover, behavioral structure seems in turn to be affected by cultural standards. In fact, as our results pointed out, Italians and Icelanders use a different conversational rhythm, highlighting a different general organization of nonverbal behavior.

In addition, cultural comparison outlined some nonverbal behavior related specificities. In particular gestures confirm themselves as culturally influenced features, in agreement with the social pragmatic perspective [5].

On the whole, the methodological approach followed in this study seems to be functional for nonverbal behavior detection and "measurement" in conversational settings. Furthermore, it seems to provide an analysis procedure that takes into account cultural variables, thus representing an attempt to face the challenges coming from behavior complexity.

Needless to say that some experimental limits are to be underlined. They refer primarily to the artificiality of laboratory setting and the consequent ecological validity diminishing of the study.

Furthermore, only nonverbal behavior was considered: for example there are no connections between specific gestures and correspondent linguistic topics.

A multimodal interaction analysis employing the same methodological approach could be an interesting suggestion for future studies.

References:

1. Anolli, L. (2002). *Psicologia della comunicazione*. Bologna, Il Mulino.
2. Argyle, M. (1992). *Il corpo e il suo linguaggio: Studio sulla comunicazione non verbale*. Bologna, Zanichelli.
3. Eibl-Eibesfeldt, I. (1970). *Ethology: The biology of behavior*. New York, Rinehart & Winston.
4. Ekman, P.; Friesen, W.V. (1978). *Manual for Facial Action Coding System*. Palo Alto, Consulting Psychology Press.
5. Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York, Academic Press.
6. Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, **32**, 1489-1522.
7. Graham, J.A.; Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, **10**, 56-67.
8. Grammer, K.; Kruck, K.B.; Magnusson, M.S. (1998). The courtship dance: Patterns of nonverbal synchronization in opposite-sex encounters. *Journal of Nonverbal Behavior*, **22**, 3-29.
9. Hall, E.T. (1966). *The hidden dimension*. Garden City, Doubleday.
10. Hall, E.T.; Hall, M.R. (1990). Monochronic and polychronic time. *The Nonverbal Communication Reader: Classic and contemporary readings*, 237-240.
11. Magnusson, M.S. (1993). *Theme, Behavior Research Software. User's manual, with notes on theory, model and pattern detection method.* Unpubblished manuscript. University of Iceland, Reykjavik.
12. Magnusson, M.S. (1996). Hidden real time patterns in intra- and inter-individual behavior: Description and detection. *European Journal of Psychological Assessment,* **12**, 112-123.
13. Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments and Computer*, **32**, 93-110.
14. McGrew, W.C. (1972). *An ethological study of children's behaviour*. New York, Academic Press.
15. Mehrabian, A. (1972). *Nonverbal communication*. Chicago, Aldine-Atherton.
16. Morris, D. (1990). *L'uomo e i suoi gesti*. Milano, Mondadori.

Smiling behavior in employment interviews - searching for T-patterns in FACS-data with Theme

Frank Schwab, Dirk Henry
Media & Organizational Psychology, Saarland University, Germany

Abstract

Studies examining employment interviews often address smiling behavior as behavioral aspect because of its relevance and multi-functionality in interaction regulation. But most of the studies lack an ecologically valid setting. In order to overcome this disadvantage, we conducted a study about the role of smiling behavior in employment interviews in a more ecologically valid setting. From a methodological point of view, two aspects of this study are mentionable. We used a round robin design where each interviewer talks - in turn - to all applicants and all applicants meet all interviewers; this produces different behavior samples of one person in different dyads. Nonverbal smiling behavior was coded with FACS (Facial Action Coding System) as a systematic, reliable and objective method. Results show that subjects produced a large amount of smiling behavior. Smiling behavior varies between different dyads, indicating that adjustment processes take place within the dyads. Contradicting to previous studies we found a negative correlation between smiling of the applicants and the chance to get selected for the job. A detailed analysis of smiling behavior is conducted by the means of T-pattern-analysis with THEME.

Keywords

employment interviews, smiling behavior, FACS, round robin, THEME

1 Introduction

Smiling is the most important and the most used nonverbal signal in human communication [1]. It has many different functions as well in the *intrapsychological* as in the *interactive* level of communication, for example affective attachment, signal of emotional relatedness (cp. [3]), relationship (like – dislike, [11]), readiness to act, behavior regulation (cp. social referencing, [4]) or signal of resonance (For an elaborated description of smiling functions see [1]).

Because of its numerous functions it is likely to suppose that smiling plays an important role in the context of job employment interviews. This thesis is provided by former empirical investigations. A lot of research reports a positive correlation between smiling behavior and a positive impression of the job applicant resp. his selection for the job (cp. [7, 8, 9, 10, 14]). However, most of the studies lack an ecologically valid setting in favor of experimental control (pre-set content, pre-set nonverbal behavior with extreme instructions "as much as possible" vs. "never") and neglect the interactive nature of a communication process . There is no study that uses an objective, systematic coding system for nonverbal behavior. Studies neglect the intra-individual variability of the nonverbal behavior

2 Hypothesis

H1: During the employment interview the subjects adapt their smiling behavior.
H2: There is a correlation between smiling behavior and selection of the subjects.
H3: There are differences in the kind of smiling of interviewer and applicant.

3 Methods

The aim of the study was to develop an ecologically valid setting and to use an objective and precise methodology to report facial behavior. Concerning the experimental design we used the “round robin” design according to Warner, Kenny and Stoto [13]. The subjects interacted in dyads and performed a fictitious employment interview (role play). Students of business studies were chosen as interviewers (N = 5) while psychology students (N=7) act as applicants (35 dyads) for an internship. No guidelines for content or nonverbal behavior were given. These interviews were recorded and afterwards coded with FACS (cp. [6]).

3.1 Round Robin

The round robin design from Warner, Kelly and Stoto [13] can be compared with a tournament system in which all participants are playing at least one time against every other participant. Assigning this example in a psychological context a round robin design can be described as a design in which all possible pairs of subjects from some set of subjects interact. For each person paired with every other person, an observation of some social behavior is made (e.g. speech pattern, nonverbal behavior,…). Thus the “treatments” to which each subject reacts are the behaviors of other subjects.

Subject	Subject			
	1	2	3	4
1	-	X_{121} X_{122} X_{123}	X_{131} X_{132} X_{133}	X_{141} X_{142} X_{143}
2	X_{211} X_{212} X_{213}	-	X_{231} X_{232} X_{233}	X_{241} X_{242} X_{243}
3	X_{311} X_{312} X_{313}	X_{321} X_{322} X_{323}	-	X_{341} X_{342} X_{343}
4	X_{411} X_{412} X_{413}	X_{421} X_{422} X_{423}	X_{331} X_{332} X_{333}	-

Table 1. *Round robin design* [13].

Table 1 shows the principle of a round robin design. The term X_{ijk} is an observation of some social behavior of person i toward person j at a time k. The diagonal cells are

empty, because a person can't be paired with himself. The matrix is asymmetric, since the behavior of person i toward person j is ordinarily different from that of j to i.
In order to use the round robin design for the context of employment interviews some adjustments of the design must be made. In employment interviews two groups of subjects are interacting with each other: the interviewers and the applicants. It is not necessary that every subject interacts with every other subject but every interviewer with every applicant and vice versa. Also it is not necessary to investigate the same dyad several times, because in real life there is only one employment interview in most cases. The round robin design has some important advantages: Observing real social interaction, it is possible to study interaction along the criteria of [2]. More than one behavior sample of a person is produced. The employment process is reproduced with high ecologically validity.
In our study we got 5 male interviewers (students of economics) and 7 female applicants (students of psychology). The applicants should apply for an internship. We made five appointments. At every appointment the interviewer conducts 7 employment interviews, one with every applicant. The order of the applicants was varied. There were no special instructions neither for the interviewer nor for the applicant.

3.2 Technical experimental setup

During the interviews the interviewer and the applicant were sitting right-angled to each other so that there faces could be recorded frontal by two remote dome cameras hanging on the wall beyond the direct visual field of the subjects. A microphone was placed at the edge of the table.

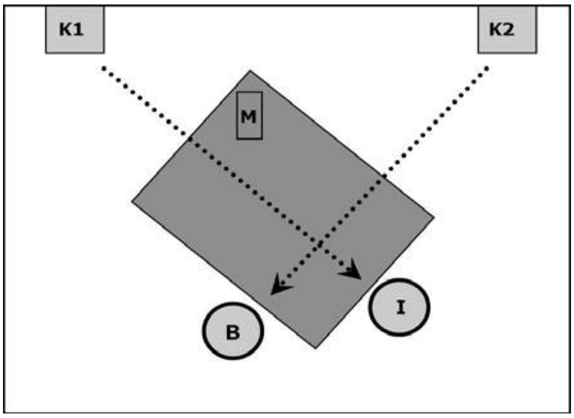


Figure 1. *schematic figure of the setting: I, B = subjects, K1, K2 = cameras, m = microphone*

The video signal of the dome cameras was transmitted to a mixer in the next room that mixed together the pictures of interviewer and applicant.

Figure 2. Screenshot *from the interviews*

A time code generator faded in the time in hours, minutes, seconds and hundredths. Finally the video signal was recorded by a digital video recorder that enabled us to control the video frame by frame (25 fps). With the aid of a monitor and a control panel we kept the record under surveillance and checked that the cameras focused directly the faces of the subjects.

3.3 Analyses of smiling bahavior

The smiling behavior was coded with the Facial Action Coding System (FACS, [6]). FACS is a system for coding the most slightly facial movements, the so called "action units" (AU). It is based on the anatomy of the face and it focuses on visible changes of the face. The appliance of FACS guarantees an objective coding of the nonverbal behavior because the description of the nonverbal behavior is clearly separated from the interpretation. The appliance of FACS enabled us to differentiate between different kinds of smiling behavior, especially between "happy felt" and "happy unfelt". "Happy felt" is later interpreted as the facial expression of "happiness" that isn't entirely under conscious control. In contrast "happy unfelt" is interpreted as some kind of social smile and not as the expression of an emotion.
In contrast to Ekman and Friesen [6] we used another definition of the observation period. Originally a new observation period begins with a change of intensity of an action unit by two steps. In the context of smiling behavior the question arises whether a subject realizes a change of intensity in smiling as a totally new smile. To anticipate the risk of an overestimation we changed the observation period: a new observation period begins as soon as the criteria for coding AU 12 (= raising the lip corners) are fulfilled and lasts as long as these criteria are met.

3.4 T-pattern-analysis

The dynamics of interaction between subjects making up an organized system is a complex matter where behavioral events are constantly happening within ever changing temporal contexts that determine the meaning, effect and function of each event. For the understanding of such systems quantification alone is not enough, matters of pattern and structure must be considered. Most of the behavioral sciences have traditionally accepted repeated patterns among their central concerns. However, tools based on relevant and adequate models and methods have been hard to find in this area. The T-pattern-detection considers not only order but also real-time when searching for hidden interaction structure [15]. It provides a real-time model, called the T-system, and is advanced together with specifically created detection algorithms. A particular pattern type, called T-pattern, is the core of the model and its detection is based on the definition of a so called critical time interval relation among series of behavioral events. THEME is used as software package to conduct an explorative study on the hidden interaction structure of smiling behavior in employment interviews.

4 Results

First of all it is to mention that the subjects showed a very high nonverbal expressivity in comparison with everyday dyadic discussions of strangers. Merten [12] found in his studies a value of 18.5 expressions of happiness per ten minutes. The interviewers in our study smiled 27.2 and the applicants even 38.8 times (per 10 min.). These numbers underline the importance of nonverbal behavior, especially smiling, in the context of employment interviews.

In H1 we investigated whether the subjects showed a constant level of smiling behavior with each partner or whether they varied their smiling behavior. Interestingly we got different results for the different kinds of smiling. Concerning "happy felt" the interviewer as well as the applicants didn't vary their smiling behavior over the different employment interviews. For "happy unfelt" we found for every single subject a significant difference in its smiling behavior over the different employment interviews.

H2 focuses on the effect of smiling behavior on the subject's selection. In contrast to former results we did not find a positive correlation between smiling behavior and the chance to get a job. For "happy unfelt" we even found a significant negative correlation with the selection decision. For "happy felt" we found neither a positive nor a negative correlation with the selection decision.

The goal of H3 was to investigate whether the applicants and the interviewers show differences in the proportion they show "happy felt" and "happy unfelt". The results evidence that there are indeed significant differences between the smiling behavior of the applicants and the smiling behavior of the interviewers. This significance is caused by the fact that applicants show much more "happy unfelt" then interviewers. "Happy unfelt" may be the product of some sort of "display rule" (cp. [5]). The applicants try not to show any negative affects and blend or mask them with a social smile.

5 Conclusion

The study showed that in an ecologically more valid setting the correlation between smiling behavior and the selection in an employment interview isn't positive. In contrast by differentiating smiling behavior in "happy felt" and "happy unfelt" the results show a negative correlation with the selection decision for "happy unfelt". This study is only a first step into the analysis of nonverbal behavior in employment interviews. Further analysis of our data will look into the following questions:

- Which person initiates what kind of smiling behavior and how is the other person reacting?
- Are there repeating nonverbal patterns? If so, is there a correlation between these patterns and the selection decision.
- Is there commonness between interviews that are rated as very successful?

In future studies it would be interesting not only to investigate smiling behavior but also the whole spectrum of nonverbal behavior. Another important point would be to investigate different sorts of dyads: same sex and mixed sex interactions. Because of the high expression of "happy unfelt" it would be interesting to look for the meaning of it's signaling. Perhaps one can find hints by putting nonverbal behavior in the context of the verbal content.

References:

1. Bänninger-Huber, E. (1996). *Mimik – Übertragung – Interaktion. Die Untersuchung affektiver Prozesse in der Psychotherapie.* Bern: Huber.
2. Bänninger-Huber, E. & v. Salisch, M. (1994). Die Untersuchung des mimischen Affektausdrucks in face-to-face Interaktionen, *Psychologische Rundschau*, **45**, 79-98.
3. Bänninger-Huber, E., Moser, U. & Steiner, F. (1990). Mikroanalytische Untersuchung affektiver Regulierungsprozesse in Paar-Interaktionen. *Zeitschrift für Klinische Psychologie*, **2**, 123-143.
4. Campos, J.J. & Stenberg, C. (1981). Perception, appraisal and emotion: The onset of social referencing. In M.E. Lamb & L.R. Sherrod (Eds.), *Infant social cognition: Empirical and theoretical considerations*. Hillsdale, NJ: Erlbaum
5. Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on motivation* (pp. 207-283). Lincoln, Neb.: University of Nebraska Press.
6. Ekman, P. & Friesen, W.V. (1978). The Facial Acting Coding System: *A technique for the mesurement of facial movement*. Palo Alto: Consulting Psychologists Press.
7. Forbes, R.J. & Jackson, P.R. (1980). Non-verbal behaviour and the outcome of selection interviews. *Journal of occupational Psychology*, **53**, 65-72.
8. Imada, A.S. & Hakel, M.D. (1977). Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Journal of Applied Psychology*, **62 (3)**. 295-300.
9. Keenan, A. & Wedderburn, A.A.I. (1975). Effects of the non-verbal behaviour of interviewers on candidates' impressions. *Journal of occupational Psychology*, **48**, 129-131.
10. Liden, R.C., Martin, C.L. & Parsons, C.K. (1993). Interviewer and applicant behaviors in employment interviews. *Academy of Management Journal*, **2**, 372-386.
11. Mehrabian, A. (1972). *Nonverbal Communication*. Chicago: Aldine-Atherton.
12. Merten, J. (2001). *Beziehungsregulation in Psychotherapien. Maladaptive Beziehungsmuster und der therapeutische Prozeß*. Stuttgart: Kohlhammer.
13. Warner, R.M., Kenny, D.A. & Stoto, M. (1979). A new round robin analysis of variance for social

interaction data. *Journal of Personality and Social Psychology,* **37**, 1742-1757.

14. Young, D.M., Beier, E.G. & Beier, S. (1979). Beyond words: Influence of nonverbal behavior of female job applicants in the employment interview. *Personel and Guidance Journal*. 346-350.
15. *Magnusson, M.S.* (2005). Understanding Social Interaction: Discovering Hidden Structure with Model and Algorithms. *L. Anolli, S. Duncan Jr., M.S. Magnusson* and *G. Riva* (Eds.) *The Hidden Structure of Interaction: From Neurons to Culture Patterns.* Amsterdam: IOS Press

Measuring the image quality of a car based sight effectiveness enhancement system

H.H.K. Andersen, K.D. Hansen
Systems Analysis Department, Risø National Laboratory, Roskilde, Denmark.

Abstract
In the present study we assessed the perceptual qualities of video recordings based on infrared images. Evaluations of the images were based on raw eye movement data obtained with a remote eye tracker. One CCD (normal) video recording and three types of infrared: Long wave, short wave and data fusion of the two were compared. The videos were mixed using a quad mixer giving us four synchronously running displays in one recording and the objective was to determine of which of the displays was preferred by the participants. The videos were recorded during nighttime driving. Data includes dwell time on areas of interest. Each display was defined as an area of interest. We used two different types of tasks: 1) detection tasks (detection of high emissive objects e.g. humans) and 2) maneuver tasks (low emissive terrain features e.g. road signs) for measuring the participants' preferences. The data fusion is used in combination with long wave in maneuver tasks and short wave is the least preferred.

Keywords
Eye movement tracking analysis, enhanced vision system, infrared images, image quality analysis.

1 Introduction
Every year numerous accidents happen on European roads due to bad visibility (fog, night, heavy rain). At least two important factors distinguish nighttime driving from daytime driving and are limiting factors to the visibility: 1) visibility is limited by the range of the car's headlights [2] and 2) glare from the headlights of oncoming cars or other light sources reduces visibility. [2; 12; 13].

A sight effectiveness enhancement system (SEE) based on infrared images has the potential of making objects more distant than the headlights' range visible to the driver. The infrared images are converted to a monochrome display placed somewhere in the car. Infrared displays installed in cars can aid or augment the driver's sight during nighttime driving by making the 'invisible visible'. Thereby the driver is enabled to detect and react earlier to potentially dangerous situations. This is supported by Hollnagel and Källhammer [9]. They found in a driving simulator experiment that drivers using a night vision enhancement system (NVES are similar to SEE systems) had better control on braking and swerving compared to drivers with no NVES. The drivers in the NVES condition gained time to assess the situation and respond to it. However, many factors influence the effectiveness of infrared displays and may introduce risks as well as benefits. Some of the issues currently debated and investigated concerning SEE systems in the form of infrared displays are:

The importance of the field of view, size, brightness contrast etc. of a display. Grönqvist [6] showed that a broader field of view improves driver's anticipatory abilities and supports better driving decisions. Karlsson [11] studied the effects of size and brightness contrast in displays on driver performance. He found that drivers using vision enhancement systems gained higher anticipation control than those without a night vision enhancement system. Results also indicated that drivers with the largest and brightest display were preoccupied with the display and paid less attention to the environment through the windshield.

Position of the display. There are several possible locations for displaying infrared information. A head-up display (HUD) is placed in front of the driver so that the driver does not have to move his head in order to see the image. The driver still has to shift between fixating directly on the external environment and the display. A head-down display (HDD) usually indicates that the driver must move his eyes or head further to see the display, increasing the workload of the driver.

The risk of cognitive capture, meaning that drivers may focus too much on the display and consequently reduce the chances of detecting obstacles occurring outside the display in the direct view through the windshield [13].

Familiarity is also considered as a factor influencing the effectiveness of infrared images on driver performance. Infrared images/recordings are thermal representations of the surroundings. Differences in thermal emission are represented in (in this case) grayscale resembling a black and white movie but with important qualitative differences [4]. Heat emissive objects such as people or animals stand out in the images in contrast to low emissive objects, e.g. roads, buildings, signs etc. [Ibid.]. Information about shadows and shading may not be present. Objects may appear different according to time of day, temperature, season etc [Ibid.].

2 Method
The purpose of the study was to evaluate which of the four available displays the participants preferred. The study was conducted as a within-subject design involving 8 participants.

2.1 Participants
3 Females and 5 males from the age of 25 – 40 have participated (mean age: 31). They have owned a drivers license between 6 – 22 years (mean: 13 years). Their degree of driving experience varies but none of them is novices.

2.2 Materials
In this study we used a CCD video recording and three types of infrared recordings: Long Wave, Short Wave and a data fusion of the two. The recordings were mixed using a quad mixer resulting in four synchronously running displays in one recording as seen in figure 1. The long wave display (LWIR) is seen in the top left corner. In the top right the short wave (SWIR) is shown. The normal

CCD display (Driver's sight) is shown in the bottom left corner and at last the data fusion display in the bottom right corner.

Figure 1. *Screen shot from video sequence.*

Qualities of infrared images differ according to wavelength [12]. Long wave images are based on heat emitted from objects whereas short wave images are based on infrared light both emitted and actively reflected from objects. Each of the infrared recordings has inherent limitations and strengths regarding the type of information they show. Short wave images bear the highest resemblance to images based on visible light, as it lies closer to the visible spectrum of light. Long wave is further away from visible light and may thus appear more unfamiliar to the subject.

The 8 videos used in the study have been chosen out of 19. The selected videos are all recorded during nighttime in both rural and urban areas. The remaining videos have been used as practice videos shown to the participants in the beginning of a session. The order of the videos was varied across participants to limit learning effects. In order to limit recognition in the second run, the time distance between the first and the second viewing of a sequence was as long as possible. Thus the order of the video sequences in the second run of each session was identical to the first runs

2.3 Tasks

Participants were asked to look at 8 different mixed video recordings two times with two different purposes in mind: detection and maneuver. All 8 videos were first used in detection tasks and secondly in maneuver tasks. Detection relates to being aware of and noticing high emissive objects like humans or wild life. Detection tasks also included vehicles even though they are not always high emissive. A significant part of the maneuver tasks was related to the paying attention to terrain features – the surroundings of the road, the course of the road etc. That is, all factors that have implications for maneuvering the car safely. Terrain features are mostly of a low emissive nature. Participants were free to use and switch between any of the views available to them and they were instructed to continuously verbalize what they saw. They had to give verbal notice as soon as they detected something in relation to the categories described above and should not wait until they were able to identify precisely what it was. They were at all times allowed to correct any wrong observations.

2.4 Technology and Laboratory Settings

Two connected standard PCs placed on an angle table were used in the present study. In that way it was possible for the researcher to be placed out of sight, behind the participant during the session. The eye tracking data was recorded using a SMI remote eye tracker at a 50Hz-sampling rate. This type of eye tracking is non intrusive since there is no physical contact between the eye tracking system and the participant. The system was calibrated between almost every video sequence. Eye tracking data was recorded in ASCII files containing information about the coordinates for the position of the eyes. A 9-point calibration procedure was used. We did a recalibration between every video sequence and tested accuracy after each session.

3 Analysis

Eye-movement tracking has been used in many areas aiming at evaluation of behavior, gathering meaningful information, for example in aviation [14], plant control operations [8]; and ship navigation [7]. These studies show the importance of using dwell time analysis to point out where people look and not look. Recently eye movement tracking has come to play an important role in computer interface evaluation. We do not se much methodological difference in evaluating moving sequences of pictures compared to evaluating conventional computer user interfaces. Both are 'maps' that contain ordered spatial symbolic information.

Several studies within other areas of research (on picture perception) have focused on the use of eye movement tracking. Wolf [15] for example showed that more complex visual stimuli require more fixations than a simple one. Yarbus [16] showed that when people were asked to look at pictures they would concentrate on the areas of the pictures that would give them the most information and when perceiving stationary visual points the eye is either fixating or changing to a new fixation, making a saccade.

Buswell [3], in his early studies, found two types of eye movement patterns: The first is a survey of the picture where the eye moves quickly with short pauses over the entire picture. The other is a set of long fixations meant to examine the image with fixations ranging from 100-1300 milliseconds. He also pointed out that longer fixation times are related with more complex perceptual processing. Also found was a strong relationship between the number of fixations in an image and the apparent amount of information in that area of the image. Just and Carpenter [10] claim that fixations are relatively stable and will last until all information at a certain fixated visual point (a word, for example) has been processed. These fixations will vary depending on the task and will range from 70 to 1200 milliseconds for cognitive tasks, and will leave out all other types of cognitive processes during the fixation.

Most of these studies focus on viewing static objects while we need to study moving images. Goldberg and Kotval [5] have been evaluating dynamic user interfaces based on eye-movement tracking. Among other things they found it valid to use eye-tracking measures like the ones discussed above. They proposed a list of content independent quantitative eye-movement measures divided into temporal and spatial dimensions. For our purpose we find that two of the temporal measures could be used as a first starting point for the evaluation of the infrared image quality.

Fixation duration (unit: milliseconds). Longer fixations could indicate that a person is spending more cognitive processing time to interpret symbolic information. Thus, longer fixation duration means lack of meaningful objects, or the eye has met difficult information. Average fixation time can be used as unit for analysis, and may serve as useful indicator of the comparative complexity of the image viewed. A measure of fixation duration is either an overall average, or is indexed within stages of the viewing sequences (average per task).

Scanpath duration (unit: milliseconds). Time spend in area of interest (AOI). Can be used to compare norm (optimal) visual behavior with observed behavior, i.e. does the person follow a sensible and rational strategy? Both measures have to be analyzed in relation to the tasks given a subject during the test.

Many eye movement-tracking studies have found large individual differences in fixations patterns among the participants. The problem could be that if the information acquisition is associated with one fixation only, the fixation definition will also depend on context variables, and consequently vary both within subject and tasks and between subjects and tasks. Any fixation definition must take into consideration that the analyzed fixation pattern depends on the definition of one single fixation, i.e. the smallest cluster of sampling points and the sampling duration.

Comparing fixation definitions that vary in milliseconds can give qualitatively different fixation patterns, or scan paths. In stead we in this study based our analysis on the notion of dwells [1]. By definition, a "dwell" requires an area, i.e. an interface parameter, upon which the line-of-gaze is directed. A dwell consists of several single fixations, the target is an area rather than the exact location of an eye point of gaze marker. Dwells are conceptually defined here as several single fixations within an area of interest (AOI). In this study the operational definition of a dwell was that the eye movement cross hair must stop for a minimum of 120 milliseconds inside the border of an AOI (or be within 0.5 degrees visual angle from the AOI center, to allow for small calibration inaccuracies). As soon as the cross hair leaves the border of an area of interest, the dwell is terminated.

In the present study we defined each of the displays as an area of interest (see figure 2). In addition we defined all four displays as one large area of interest (not shown in fig. 2) in order to get an overview of visual activity between the displays. The area acted as a control to ensure usable data. All visual activity on each of the areas was recorded and analyzed.

Figure 2. *The white boxes represent areas of interest.*

The eye tracking data (measured in milliseconds) was extracted from our SMI iView data analysis software module exported to a spreadsheet for further statistical analyses (ANOVA and t-test).

4 Results and Discussion

Results show that the participants have used the long wave infrared display for a significant longer time compared to the other three in connection with detection tasks - as seen in table 1.

Table 1 – detection (only significant results are shown)

t-test	Mean (seconds)	P(T<=t)two-tail
LWIR vs. SWIR	22,04; 12,80	0.01
LWIR vs. CCD	22,04; 11,11	0.01
LWIR vs. Fusion	22,04; 10,78	0.01

Short wave is the least preferred of all displays in maneuver tasks – see table 2. Apart from that the participants do not seem favor one display over the rest as in the detection tasks.

Table 2 – maneuver (only significant results are shown)

t-test	Mean (seconds)	P(T<=t)two-tail
LWIR vs. SWIR	16,82; 7,61	0.01
CCD vs. SWIR	16,11; 7,61	0.02
Fusion vs. SWIR	17,46; 7,61	0.03

The long wave recording is based solely on heat emitted from objects. It shows the thermal signature of the objects in a scene. It is on the one hand possible to see very far ahead and high emissive objects stands out very clearly in contrast to the background - e.g. the dog on the curve on the right in figure 1. On the other hand the amount and details of other types of information is limited. Low emissive objects or objects that adopt the temperature of the surroundings are not easy to see or disappear completely [12]. An example of this is the road markings in the LWIR camera display in figure 1. Another example could be road signs. If a road sign is visible it appears as a grey shadow and it is not possible to tell what it displays. Infrared images based on long wave light are the most unfamiliar to the subject. Subjects may spend more time interpreting what they see.

Short wave infra red light has a shorter range than long wave infra red light (but longer than the cars headlights under normal circumstances). Its wavelength lies closer to the visible spectrum of light and is not only based on heat emitted from objects but also actively reflects natural existing infrared light on objects. Thus making the objects in the image more natural and familiar to the subject compared to the long wave image (Knoll, 2002). Apart from the distance covered and degree of familiarity for the subject, another important difference between short wave and long wave infra red light is that light from different sources is visible in the short wave image. For example, as seen in figure 1 the streetlight is visible in the SWIR camera display and light from vehicles further down the road. On the one hand this may increase the familiarity to the subject while on the other hand the glare from oncoming cars disturbs the picture and might in some cases make other high emissive objects harder to detect. It is possible that the fusion and the short wave recordings may have received more attention if the glare from oncoming cars and streetlight was limited. The data fusioned infra red recording has both advantages and disadvantages compared to both long wave and short wave. It has a long range and all infrared information is available in the data fusion display.

Several interpretations can be made on the background of these findings. It seems that the low amount of details shown in the long wave display may be an advantage in detecting wild life or humans. Apparently the unfamiliarity of the long wave infrared images does not affect the participants' preferences. With respect to support for maneuvering the car it seems that the long wave does not give enough details about the surroundings and the data fusion is used to aid this task in combination with long wave.

References

1. Andersen, H.H.K.; Hauland, G., (2004): Studying operator behaviour during a simple but safety critical task. In: Human error, safety and systems development. *IFIP 18. World computer congress* TC13 / WG13.5 7. Working conference on human error, safety and systems development, Toulouse (FR), 22-27 Aug 2004. Johnson, C.W.; Palanque, P. (eds.), (Kluwer Academic Publishers, Dordrecht, 2004) p. 209-221
2. Blanco, M. (2002): Relationship Between Driver Characteristics, Nighttime Driving Risk Perception, and Visual Performance under Adverse and Clear Weather Conditions and Different Vision Enhancement Systems. Available as http://scholar.lib.vt.edu/theses/available/etd-05202002-171805/unrestricted/Blanco-Dissertation.pdf
3. Buswell, G. T. (1935): How people look at pictures. A study of the psychology of perception in art. Chicago: The University of Chicago Press, 1935.
4. Foyle, D. C. (1991): Infrared Imagery in Flight. MS 262. Aerospace Human Factors Research Division, NASA Ames Research Center, Moffett Field, CA 94035-1000. http://human-factors.arc.nasa.gov/ihi/papers/publications/Foyle/visualissues91/visualissues91.html
5. Goldberg, J.H. and Kotval, X., P. (1998): Eye Movement-Based Evaluation of the Computer Interface, in Kumar, S.K. (Ed.), *Advances in Occupational Ergonomics and Safety*, Amsterdam: IOS Press, pp. 529-532.
6. Grönquist, H. (2002): Vision Enhancement Systems. The Importance of Field of View. Master Thesis. Cognitive Science Program, Linköping University, Sweden. September 2002. *LIU-KOGVET-D—02/09—SE.*
7. Hansen, J., P. and Itoh, K. (1995): Building a cognitive model of dynamic ship navigation on basis of verbal protocols and eye-movement data. In Lena Nooros (Ed.), *VTT Symposium 158, the 5th European Conference on Cognitive Science Approaches to Process Control*, Espoo, Finland, pp. 325-337.
8. Hauland, G. and Halbert, B. (1995): Relations between visual activity and verbalized problem solving: A preliminary study. In Lena Nooros (Ed.), *VTT Symposium 158, the 5th European Conference on Cognitive Science Approaches to Process Control*, Espoo, Finland, pp. 99-110.
9. Hollnagel, E. and Källhammer, J. (2003): Effects Of A Night Vision Enhancement System (Nves) On Driving: Results From A Simulator Study. *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment,* Training and Vehicle Design.
10. Just, M., A., and Carpenter, P., A (1980): "A Theory of Reading: From Eye Fixations to Comprehension," *Psychological Review*, vol. 87, no. 4, pp. 329-354.
11. Karlsson, J. (2002): Vision Enhancement Systems – Support for Night-time Driving? Master Thesis. Department of Computer and Information Science, Linköping University, Sweden, March 2002. *LIU-KOGVET-D-0063--SE*
12. Knoll, P. M. (2002): A NIR Based System for Night Vision Improvement. *SAE Convergence 2002*, Detroit.
13. Rumar, Kåre (2003): Infrared Night Vision Systems and Driver Needs. SAE Technical Paper Series. Lighting Technolgy, *SAE World Congress* Detroit, Michigan March 3-6.
14. Weber, Steen, T., Bove, C. (2000): Le Draoullec, F., Reuzeau, H., B., Andersen: Pilot's response and eye fixations during take off. *Risø National Laboratory Report Series.*
15. Wolf W., (1970) A study of eye movements in television viewing, *Final Reports (ERIC Reproductions Ser-vices No, ED 046 254* Columbus Ohio State University.
16. Yarbus, A. L (1967): *Eye movements and vision.* New York: Plenum press.

e&t attentiontracker®: measuring the visual conspicuity of elements on websites. An evaluation study

Raphael Indergand[1] , Christopher H. Müller[1], Daniel Felix[1]
[1] *ergonomie & technologie (e&t) GmbH, Zurich, Switzerland*

Abstract

As in advertising, online marketing works best if the users can immediately grasp the message. There is no easy to use and valid method for measuring what viewers really perceive and which elements are eye-catching. Therefore the e&t attentiontracker® was developed to measure the actually perceived elements on websites, advertising media, instruction manuals etc. The e&t attentiontracker® allows to identify the conspicuous elements and a relative ranking of the conspicuity. A website is briefly presented to subjects several times. After every presentation, they are asked to reiterate what they have seen on the website. With this method, it is possible to detect and rate the elements that were perceived consciously. In this study we evaluated the method of the e&t attentiontracker®. The results show that the e&t attentiontracker® is able to measure even slight differences in the conspicuity of elements on a website. This evaluation study proved that the e&t attentiontracker® effectively measures conspicuity of website elements.

Keywords

e&t attentiontracker, conspicuity, perception, usability, eye tracking

1 Introduction

Online communication pursues concrete goals: The E-shop must sell, a brand must be well perceived. The first couple of seconds determine the subjective impression users gain about the website and its offers. Countless guidelines of web design can support the designing process, but in the end it is the user alone who decides, whether a website really achieves its goals or not. It is a fact, that the first couple of seconds are crucial for the user's decision whether to stay on the website or to leave it.

Attention research can help answering that question. Several methods already exist to provide information about the path of attention on websites or print products. The most popular and commonly used technique is Eye Tracking. But lately, with the increasing importance of online communication, alternative methods have been developed, especially to examine commercial websites as a part of usability evaluation.

The only method with a cognitive, not a physiologic approach towards measuring attention is the e&t attentiontracker®. The e&t attentiontracker® focuses on what users perceived *consciously*. The outcome of an e&t attentiontracker® test is the order of the conspicuity of elements on a website. Disler, Müller & Felix [4] already showed that:

- there are significant differences between the various elements in terms of attention given to them by the viewers,
- the point of time when elements were perceived differs significantly between more and less obvious elements on the test objects.

In this study, both method and tools of the e&t attentiontracker® were evaluated.

2 Background/Related Methods

2.1 Visual attention

The attention of humans is strongly restricted. Only a small part of our environment is being processed, the rest is being filtered. The visual attention works with a very small area of our visual field, the fovea. The process of visual attention can be described with a spotlight metaphor. The eye moves over the visual field, while just a small part is perceived by the fovea. Normally, but not always, we fix the part of our visual field where our attention lies [1]. These circumstances have to be considered in designing interfaces. Which elements grasp our attention can be guided by design arrangements. But not all parts of our environment we look at is being percieved consciously. On the other hand, also parts of our environment that are not within our visual focus can be perceived consciously [8].

2.2 Related Methods

Eye Tracking is the most common used instrument to explore the users' path of visual attention on websites. The eye movement is recorded while subjects look at a picture. Recorded parameters are for instance fixation time, number of fixations and scanpaths [5]. These recordings must be interpreted and deliver information about the path of attention. But as mentioned above, these informations are interpretations. Participants of Eye Tracking experiments reported that they saw elements they never fixed, according to the interpretation of the collected data. Another disadvantage is the high costs of this method.

Another method of tracking the users' attention on websites works with mouse clicks. Participants have to click on the spots that attract their attention [9]. These clicks are recorded and provide a map of frequency of mouse clicks.

The Restricted-Focus-Viewer by Blackwell, Jansen & Marriott [3] works with the spotlight metaphor. Subjects use a mouse to move a spot over a blured picture on screen. The acuity of the picture increases from outside in of that spot while subjects have to explore the picture. Blackwell, Jansen & Marriott found similar results to an eye tracking experiment about reasoning of mechanical diagrams.

3 Method

3.1 Subjects

60 subjects participated in the experiment. They were selected out of the circle of friends of the experimenter. Most of them were students or academics with an average age of 27.4 (std. Deviation = 5.2). 31 of them were female, 26 were male. All of them have already surfed in the internet: The self rated average time in the internet per week was 296min (std. Deviation = 239.5) with an averaged 14.83 different visited websites (std. Deviation = 11.8).

3.2 The experimental design

The aim of the experiment was to show that with the e&t attentiontracker® even a slight change of the conspicuity of an object on a website can be measured. Therefore, three different prototype websites were designed in Adobe Photoshop 7 (A1, B1 and C1). On each site two changes were made seperatly, resulting in nine tested websites.

3.3 The tested Websites

The websites can be considered as average sites with a common general layout such as a main navigation on the left side or at the top, a content frame in the middle, login or search fields etc. [2]. The colors used are from the browser-safe web palette [7] and the fonts used are also commonly used in web design in order to provide a familiar look and feel. Texts and single words are in Latin from the website www.blindtexte.de [6] and not in German (the participants mother language), because the focus of this experiment was on graphical elements of websites and not on its contents.

The changes of objects concerned color, shape, placement, luminance contrast or size. All the variations from the prototype sites are intended to increase the conspicuity of the changed object. An overview of attributes that guide the deployment of visual attention can be found in Wolfe & Horowitz [11].

On the website A (Figure 1), the first variation was a different color of a button at the lower right of the site (A2). On the prototype, the color was in RGB-mode 0:0:255 (blue), in the changed version in 255:0:0 (red). As a second variation on the website A the position of a login field and a navigation frame was swapped (A3).

On website B (Figure 2), the changes concerned size and contrast. The first variation was applied to a search field at the bottom right of the site (B2). The size of the colored frame was increased. The second variation on this site was a grey layer behind a list of links to the right (B3).

On website C, variations were made in color and shape. First, an active navigation element on the left side of the page was made brighter (from RGB 102:204:51 to 102:255:0) (C2). The second variation was changing two boxes on the right to circles (C3). The size of the area was not varied.

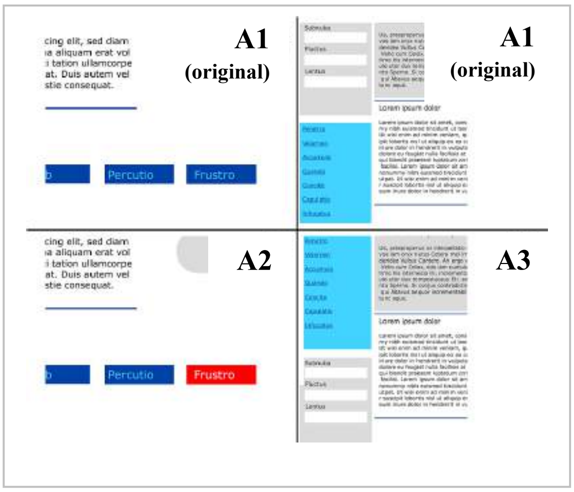


Figure 1. *Changed elements on website A: Left: Button from blue to red (A2). Right: Position of link list and login field swapped (A3).*

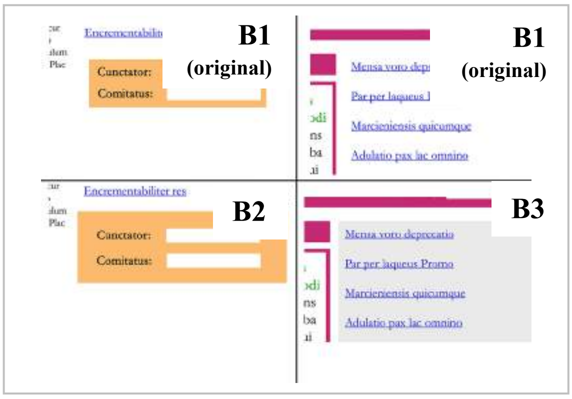


Figure 2. *Changed elements on website B: Left: Frame of search field is larger (B2). Right: Link-list with grey background (B3).*

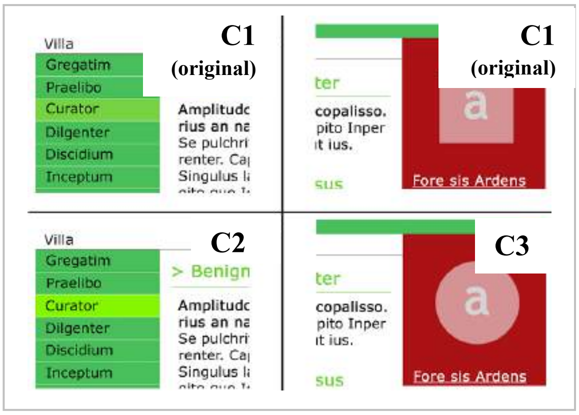


Figure 3. *Changed elements on website C: Left: Active navigation is brighter (C2). Right: Element is round (C3).*

3.4 Procedure

The experiment took 5 to 10 minutes per subject. The subjects were seated at a normal viewing distance (approximately 60cm) from the LCD monitor with a 1024x768 screen resolution. Testing procedure: A website is displayed five times in short succession. The duration of the first tachystoscopic presentation was 200ms. Each following presentation lasted twice the duration of the previous up to 3200ms for the fifth presentation. Participants started the presentations themeselves by pressing the space par. After a delay of 1.5 seconds, the screen shot appeared on the screen. After every presentation, the subjects were asked to report spontaneously what they had recognized. The

experimenter collected the data on print outs of the websites by marking the recognized elements together with the number of the presentation in which the element has been mentioned in. The order of recognized elements wasn't collected to avoid influences of primacy/recency-effects. Between every presentation, a mask of dark grey (RGB 128:128:128) appeared on the screen to avoid luminance transients [1]. In order not to influence the scan paths of the subjects, the experimenter asked only general questions if necessary at all to evoke answers from the subjects. Asking subjects about colors or shapes would have a significant impact on the perception of elements (expectations) in the next round, so the experimenter did not interfere by interviewing the subjects.
Each subject tested three websites, one of the A-Series, one the B-Series and one of the C-Series. So every Website was tested with 20 subjects. The order of the tested sites was randomized to avoid any influences of order.

4 Results

4.1 About the subjects

In a first step, participants were rated in their ability of percieving elements of the tested websites. Neither age, sex or internet experience showed any significant differences in the number of consacioulsy perceived elements. Therefore it can be assumed that any adult can participate in an e&t attentiontracker® test without distorting the overall results. Also the total number of recognized elements varied. Elements frequently recognized were also perceived early. Subjects with a low total number of recognized elements did not perceive elements consiously that were seen by subjects with a high number of recognized elements in a late presentation.

4.2 Conspicuity of the changed elements

The differences of conspicuitiy of the changed elements have been tested with the Kolmogorov-Smirnov test (with SPSS 11.0). Furthermore, we calculated a Conspicuity Index (CI), which includes both the number of subjects recognizing an element and the average trial the element was first recognized. For this purpose we calculated the integral of each element and projected it on a scale from 0 to 10. An element that has not been recognized by any participant has a CI of 0, an element that has been recognized by all the participants in the first presentation is rated as a 10. This index is useful to show customers of an e&t attentiontracker® test how the conspicuity of different objects within a website can be rated. The CI of an element should not be considered as a general but a relative parameter for elements on the same website.

Elements changed (from…to…)	**Kolmogorov -Smirnov-Z (n=20)**	**CI* before – after**
A1→A2: Button from blue to red	2.688***	0 → 3.19
A1→A3: Higher position of links	2.055***	.31 → 3.50
B1→B2: Login Field enlarged	1.581**	4.06 → 7.56
B1→B3: Links with grey background	.949	5.88 → 5.06
C1→C2: Square to Circle	2.214***	2.63 → 7.69
C1→C3: Active navigation field brighter	1.107	1.13 → 3.31

*CI=Conspicuity Index

*Table 1. Kolmogorov-Smirnov significance test of differences in conspicuitiy of changed elements. (Significance level: *<0.05, **<.0.01, ***<0.001)*

Results in Table 1 show that in five out of six hypotheses the e&t attentiontracker® measured the expected increase of conspicuity of the changed elements. On website A, both changes turned out to be highly significant. On website B, the enlarged search field was measured as significantly more conspicuous. The link list with grey background turned out to be less conspicuous than without background colour. It is likely that decreased contrast of text and background leads to this result. On Website C, both changes rendered the elements more conspicuous but only the change of the shape from square to circle resulted in a significant difference of conspicuity.

5 Conclusions

This experiment showed, together with the findings of [4], that the method of the e&t attentiontracker® is a valid tool to gather information about the conspicuity of elements of websites. Futhermore, we assume that any kind of product can be tested with this method in terms of the visual perception.
It is obvious that the CI can not be taken as an overall index of a websites' element. As the experiment showed, the change of one element on a website has an impact on other surrounding parts.
It is important that during the interview the experimenter does not ask specific questions as follow up to remarks by the subjects. This would lead to an intrusion into the natural scan path of the subject. Another important aspect is that a random scan on a display is quicker than a systematic search [11]. By asking specific questions the attention of participants could become led consciously and would lead to a slower performance.
To further improve the method it will be interesting to modulate certain parameters of the experiment such as the presentation time, the number of trials, the size of the presented website. Another interesting aspect could be the allegation of a search target combined with an alaysis of the number of recognized elements. As Wolfe et al. [11] found that a commanded search is slower than an anarchic search.
Finally to bridge to other methods as mentioned above, another experiment should be conducted, e.g. a combination of an Eye Tracking test with a e&t attentiontracker® test.

Supported by Prof. Dr. Hirsig of the Pschologisches Institut, Fachrichtung Psychologische Methodenlehre at the University of Zurich. Thanks to Andy Disler, Barbara Leu, Matthias Dreier, Markus Gubler and Samara Chatila who conducted first experiments with the e&t attentiontracker® which set the base for further developing the method.

References:

1. Anderson, J. R. (2001). Kognitive Psychologie. Heidelberg: Spektrum Akademischer Verlag.
2. Bernard, M. (2001). Developing Schemas for the Location of Common Web Objects. *Usability News 3.12001 http://psychology.wichita.edu/surl/usabilitynews/3W/web_object.htm*.
3. Blackwell, A. F., Jansen, A. R. & Marriott, K. (2000). Restricted Focus Viewer: A Tool for Tracking Visual Attention. In: M. Anderson, P. Chen, and V. Haarslev (Eds.): *Diagrams 2000*, LNAI 1889, pp. 162-177, 2000. Springer-Verlag Berlin Heidelberg 2000.
4. Disler, A., Mueller, C. H. & Felix, D. Attention Tracking – Measuring the focus of attention. (2003). *Human Computer Interaction INTERACT '03*. Zurich, Switzerland. 1st – 5th September 2003.
5. Duchowski, A. T. Eye Tracking Methodology: Theory and Practice. Springer, 2003.
6. http://blindtexte.de
7. http://lynda.com/hex.html
8. Luer, G., Lass, U. & Shallo-Hoffmann, J. (1988). Eye Movement Research, Physiological and Psychological Aspects. Göttingen: Hogrefe.
9. Scheier, C. & Koschel, K. (2002). Your Customer's Eyes. *Planung & Analyse*, 10, 2002.
10. Wolfe, J. M., Alvarez, G. A. & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature, Vol 406.*
11. Wolfe, J. M. & Horowitz, T. S. (2004) What attributes guide the deployment of visual attention and how do they to it? *Nature Reviews, Vol 5.*

Information seeking in the WWW: Detecting T-patterns in eye movements and navigational behavior

D. C. Unz [1], B. Werner [1], R. Mangold [2], M. Burmester [2]
[1] *Media and Organizational Psychology, Saarland University, Saarbruecken, Germany*
[2] *Information Design, Stuttgart Media University (HdM), Stuttgart, Germany*

Abstract
A lot of users are frustrated about using internet search engines. Therefore, the question how to facilitate the processing of search result lists is posed. The aim of our studies is to account for the temporal course of user's behavior and address the process in time. In particular, we ask, what kind of heuristics users pursue when inspecting result lists of search engines. We recorded subjects' eye movements and mouse clicks while subjects worked with search results pages. The observational records of eye movements and mouse clicks are analyzed for its temporal and sequential structure by means of the T-pattern detection method proposed by Magnusson [1,2,3] and operationalized in the software THEME. We asked if we can find different (observable) behavior patterns indicating that people pursue different (unobservable) heuristics. The results demonstrate that the study of the temporal structure of eye movements and mouse clicks can help to get a deeper insight into the structure of information seeking process.

Keywords
WWW-Search, User issues, Navigation, Eye-Tracking, THEME

1 Introduction
Information is a key resource in the 21st century. In order to get access to this resource, more and more people use the internet. Search engines are one of the most commonly used services of the internet. Half the users spend more than 70 % of their online-time seeking for information. But, on the other hand, nearly half the users are frustrated about using a search engine. Often, users act intuitively, furthermore, they evaluate their competence as very poor [4]. In order to facilitate the processing of search result lists it is important to analyze the behavior of search engine users. By describing behavior and by identifying strategies search engines may be improved and may be adapted to the user needs.

2 Research questions
A major challenge in the process of information seeking is the evaluation of the entries in a search result list. Users have to decide what document to open by means of only little information. A good deal of research has looked at the question of how to facilitate this. Typically, aspects of the user's behavior such as the opening of documents and overall search times have been recorded. But if one accepts the argument that information seeking is a complex process then such simple frequency data or overall search time data can only provide a relatively superficial view of the process. For the understanding of the process quantification alone is not enough, matters of structure must be considered. We have to consider the temporal course of user's behavior and address the process in time. One question that addresses this issue is: In what order do users look at the entries in a search result list? The usual design of such lists suggests a depth-first strategy: The user examines each entry in the list in turn, starting from the top, and decides immediately whether to open the document in question. In a first study (42 subjects were presented with a search results page that listed 25 web pages retrieved for a query about "assessment centre building blocks" [5]) we recorded subjects' eye movements and mouse clicks and looked at the order in which the search results were processed. Most of the subjects seem to apply a sort of strict depth-first strategy, a minority seems to apply a breadth-first strategy, and another minority seems to apply a partially breadth-first strategy. However, just looking at the data by appearance is not statistically firm and gives no objective measures, it may contain many mistakes and may overlook structures that are not visible at first glance.

In order to get a deeper insight in the process of information seeking, we ask, what kind of heuristics are pursued by users when inspecting result lists of search engines? Supposing that information seekers either consider few information and decide quickly or regard information at great length and in detail in order to come to an optimal decision (according to a differentiation of simple heuristics by Gigerenzer and Todd [6]), differences in behavior should emerge. The extent to which users follow such a strategy can be determined through analysis of eye tracking and log file resp. mouse clicks. So, we have to ask: Is there a (hidden) temporal and sequential structure in eye movements and mouse click behavior? What kinds of (hidden) structure exist in eye movements and mouse clicks? By what kind of (simple) mental heuristics may these patterns be produced?

3 Method
50 subjects were asked to perform three information seeking tasks. The task of the subjects was to answer questions about three different issues ((1) assessment centers, (2) backache and (3) investment in stock). Tasks varied in specificity [7]. Subjects were presented with a search result list (12 pages) for each task that listed web pages retrieved for a query about the three issues (Queries consisted of two or three search items and contained no operators or extended search functions. This corresponds to the general practice of users [4]). Subjects should perform the tasks by visiting at least some of the listed web pages. The subjects were assigned to different decision strategies or "simple heuristics" [6] according their statements in a questionnaire.

Our objective was to derive strategies of decision from observable data. Log files provide data about actions like mouse clicks, but it remains unclear what processes are in between these actions. These processes are exactly the essentials of decision strategies, for instance the order of looking at the entries in a search result list. Therefore eye movements are used as a further behavioral research method [8]. Eye movements are well measurable aspects of behavior that could directly be related to cognitive processes. They are indicators for interests and information processing. It could be taken for granted that information processing of stimuli only takes place during fixation periods.

In order to affect the process of information seeking as slightly as possible a remote eye tracker was used (Eyetracker ET-17 and Software ClearView from Tobii). The eye tracker is positioned below the screen and allows a contactless and non-reactive recording of eye-movements. The system additionally gathers event data similar to log files, so that different observable parameters of information seeking can be integrated. Behavior is continuously recorded in great detail, thus allowing conclusion about information processing of a person. The focal points recorded in an interval of 20 ms are aggregated in fixations. Additionally, areas of interest (AOI) are defined. Based on that data one could derive how long a person has fixated special areas of the web page. In our study each entry (containing title, summary, URL) of the search result list was defined as a rectangular AOI.

To study the temporal course of processing the search result list in detail, the observational records of eye movements and mouse clicks are analyzed for its temporal and sequential structure. The analysis is based on the T-pattern-model proposed by Magnusson (for a detailed description, see [1,2,3]) and operationalized in the software THEME. By means of the synchronized time series data we can look at the dynamic process of information seeking. The T-pattern-model views behavioral organization as a repetition of intra- and inter-individual behavior patterns. A process named T-pattern detection allows the detection of the temporal and sequential structure in real-time behavior records. The pattern detection method focuses on the relationships between the occurrences of behavioral event types in real time. T-pattern detection is based on an algorithm that tests the distribution of each pair of event types in time. The search for patterns is done by comparing the times for repeated occurrences of event types assuming a chance expectation in which behavior is independent, random, and uniformly distributed over time. If there is a significant relationship between event types they form a T-pattern. This means, a T-pattern is a combination of event types, the components of a T-pattern occur in the same order at similar time intervals in a behavior record.
The data of all subjects representing different observation periods were joined in a single data file to be analyzed on a significance level of p=.0001.

4 Results and Conclusion

We found no differences in behavior and performance depending on the assigned decision strategies. Possibly, the proposed strategies do not appear until subjects have a high level of previous knowledge in the domain. Regarding only subjects with a high previous knowledge there seems to be differences in behavior depending on different decision strategies.

Our results show a high number of nonrandom temporal patterns in the event time series of eye movements and mouse clicks as log file similar event data. The detected patterns indicate that information seeking process is temporally structured. The analysis results in different kinds of patterns. For example, we find longer, linear patterns and shorter, more complex patterns. The occurrence of the patterns depends on the search page a user inspects. On the first page of a result list short, non-linear patterns occur (see figure 1 as an example). Deciding what entry may be relevant and what link should be followed, users seem to move forward and backward between close-by entries. On subsequent pages of a result list longer, linear patterns occur (see figure 2 as an example). Typically, these patterns only consists of eye movements, no mouse click events are included. Users seem to look at on entry after another just in the order they are listed.

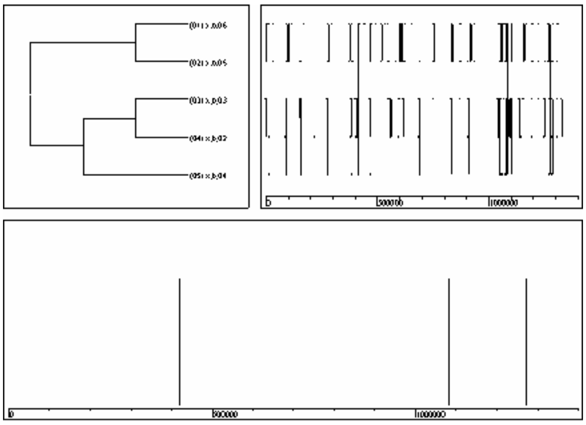


Figure 1. *Example of a temporal pattern of eye movements and mouse clicks. The Data output from Theme software shows temporal and hierarchical representation of a T-pattern. The top left box shows the hierarchical structure of the pattern. The top right box shows the times of occurrence of each of the event types and the way they are connected. The lower box shows the pattern trees as a function of time. The pattern includes five event types (fixation of entry no. 6 of the search result list is followed by fixation of entry no. 5, moving backward to entry no. 3 and no. 2 and then moving forward to no. 4).*

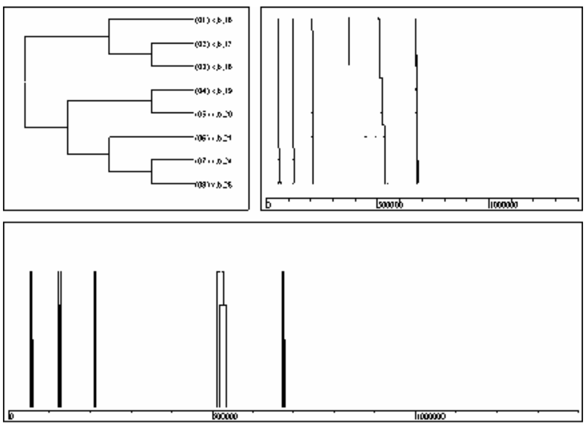


Figure 2. *Example of a temporal pattern of eye movements. The pattern is longer and more linear than the pattern presented in figure 1. The pattern includes eight event types (fixation of entry no. 16 of the search result list is followed by fixation of entry no. 17, moving forward to entry no. 18, no. 19, no. 20, no. 21, no. 24 and no. 26) and occurred 5 times in the behavior record.*

The analyses of data show that eye tracking in combination with recording event data similar to log file can yield valuable information. Whereas studies about information seeking in the WWW typically simply count certain aspects of user's behavior, as the opening of documents, or examine overall search times, our results indicate the necessity to address the ongoing process over time. The presented results demonstrate that by the means of T-pattern-analyses with THEME it is possible to identify structure in the eye movements and mouse clicks as observable parameters of the information seeking process. This highlights the potential of THEME to make a significant contribution to the examination of the information seeking process in the WWW.

References

1. Magnusson, M. S. (1996). Hidden Real-Time Patterns in Intra- and Inter-individual Behavior: Description and Detection. *European Journal of Psychological Assessment*, Vol. 12, Issue 2, 112-123.
2. Magnusson, M. S. (2000). Discovering Hidden Time Patterns in Behavior: T-Patterns and their Detection. *Behavior Research Methods, Instruments and Computers*, 32(1), 93-110.
3. Magnusson, M.S. (2005). Understanding Social Interaction: Discovering Hidden Structure with Model and Algorithms. In L. Anolli, S. Duncan Jr., M.S. Magnusson and G. Riva (Eds.) *The Hidden Structure of Interaction: From Neurons to Culture Patterns.* Amsterdam: IOS Press
4. Machill, M., Neuberger, C., Schweiger, W. & Wirth, W. (2003). Wegweiser im Netz. Qualität und Nutzung von Suchmaschinen. In M. Machill & C. Welp (eds.), *Wegweiser im Netz. Qualität und Nutzung von Suchmaschinen* (pp. 13-490). Gütersloh: Bertelsmann Stiftung.
5. Wirschum, N. (2003). *Erleichterung der Informations-suche im Internet.* Unpublished diploma thesis, Institut of Psychology, Saarland University.
6. Gigerenzer, G., Todd, M. P. & the ABC Research Group (1999). *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press.
7. Marchionini, G. (1995). *Information Seeking in Electronic Environments.* Cambridge: Cambridge University Press.
8. Körber, S. (2000). *Suchmuster erfahrener und unerfahrener Suchmaschinennutzer im World Wide Web. Ein Experiment.* Unpulished Master Thesis, Westfälische Wilhelms University, Muenster.

The importance of visual ability in behavioral testing of mice

R.E. Brown[1], A.A. Wong[1]
[1] *Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4J1*

Abstract
We evaluated visual discrimination, pattern discrimination and visual acuity in 13 strains of mice (129SI/Sv, A/J, AKR, BALB/cByJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, FVB/NJ, MOLF/Ei, SJL/J, SM/J and SPRET/Ei). As predicted by the physiology of their visual system, mice with no known visual defects (129SI/Sv, C57BL/6J and DBA/2J) performed well on all three visual tasks, mice with poor vision due to photoreceptor damage (AKR/J, BALB/CbyJ) showed moderate improvement and mice with retinal degeneration and other visual defects (A/J, C3H/HeJ, CAST/Ei, FVB/NJ, MOLF/Ei, SJL/J, SM/J and SPRET/Ei) performed at chance levels. These and other mice of the same strains were given tests of anxiety, motor performance and learning and memory. Strain differences in visual acuity accounted for a significant proportion of the variance in some measures of anxiety (center square frequency in the open field) and learning in the Morris water maze but motor learning performance in the Rotorod was not influenced by vision. Olfactory discrimination learning was enhanced in mice with visual defects. These results indicate that visual ability must be accounted for when testing strain differences in cognitive performance in mice.

Keywords
Vision, mouse, learning, memory, behavior.

1 Introduction
Inbred mutant and transgenic mice are now used for a wide range of neurobehavioral studies on the genetic basis of cognitive dysfunction [1, 4]. Poor performance on visuo-spatial tests such as the Morris water maze and the Radial arm maze is used to indicate cognitive impairment in many mouse models of neurodegenerative diseases [3, 7]. Many commonly used strains of mice, however, have visual defects, which may affect performance in visuo-spatial tasks. To examine visual ability as a possible confound in tests of cognitive function, we tested 13 strains of mice in three tests of visual discrimination using the visual water box [9] and correlated their performance in a battery of behavioral tasks with their visual ability.

2 Methods
Three strains of mice with reported normal vision [129SI/Sv (6M, 2F), C57BL/6J (10M, 6F), DBA/2J (6M, 10F], two albino strains with poor vision due to photoreceptor damage [AKR/J (6M), BALB/cByJ (6M, 4F)], four strains with retinal degeneration and other visual defects [A/J (9M, 6F), C3H/HeJ (8M, 7F), CAST/Ei (7M, 4F) FVB/NJ (6M, 6F), MOLF/Ei (8M, 7F), SJL/J (9M, 6F), SM/J (6M, 8F), SPRET/Ei (10M)] were obtained from JAX laboratories (Bar Harbour, ME) and subjected to a behavioral test battery that evaluated:

Anxiety-related behaviors
Open field maze (one 5-minute trial/day for 2 days) – measures line crossing, center square entries, rearing, defecation and freezing [4].
Light-dark transition test (one 5-minute trial) – measures the number of transitions from the light to dark compartments and time spent in the light compartment [4].

Motor learning
Rotorod (6 trials/day for 7 days) – provides a measure of neuromuscular co-ordination, fatigue and motor learning [6].

Spatial and olfactory learning and memory
Morris water maze (4 trials/day for 6 days and one probe trial on day 7) – measures swim latency, swim distance, average velocity, duration, frequency of thigmotaxic behaviour for 3 days of acquisition and 3 days of reversal training and percent time spent in the correct quadrant during the probe trial [8].
Olfactory discrimination learning task (4 training trials/day for 4 days, one test trial on day 5) – measures percent time digging in the CS+ [11].

Using the visual water box [9] (Figure 1), the visual abilities of mice were evaluated using a visual discrimination (VD) task (8 days, 8 trials/day), in which mice were trained to swim towards a vertical grating (0.17 c/deg) to locate a hidden platform versus a grey screen (no platform). They were then trained in a pattern discrimination (PD) task (8 days, 8 trials/day), in which mice were trained to swim towards a vertical grating (0.17 c/deg) versus a horizontal grating (0.17 c/deg) and a visual acuity (VA) task (8 days, 8 trials/day), in which mice were trained to swim towards a vertical grating of various increasing spatial frequencies versus a grey screen.

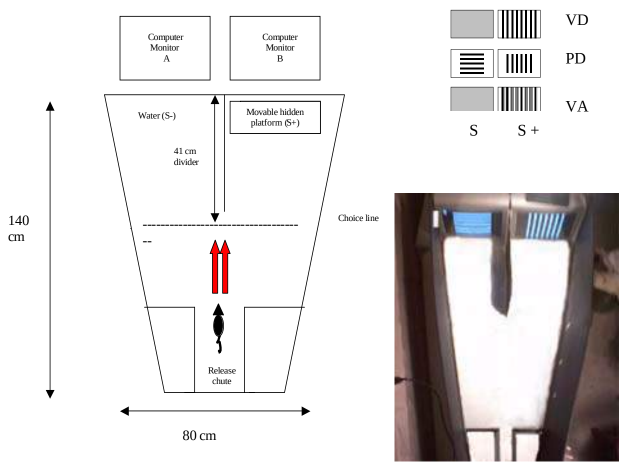


Figure 1. *Visual Water Box*
The visual water box, showing the location of the release chute, divider, hidden platform, two computer monitors and the stimuli used for the visual discrimination, pattern discrimination and visual acuity tasks.

3 Results

3.1 Visual Discrimination task

Of the thirteen strains of mice that were tested, all three strains with reported normal vision and the two albino strains were able to reach the criterion for learning the visual discrimination task (70% correct). Two strains (129S1/SvImJ and DBA/2J) reached criterion after only one day of testing, one strain (AKR/J) took three days and the other two strains (BALB/cByJ and C57BL/6J) required 6-8 days to reach criterion. No other strains reached criterion (Figure 2) [15].

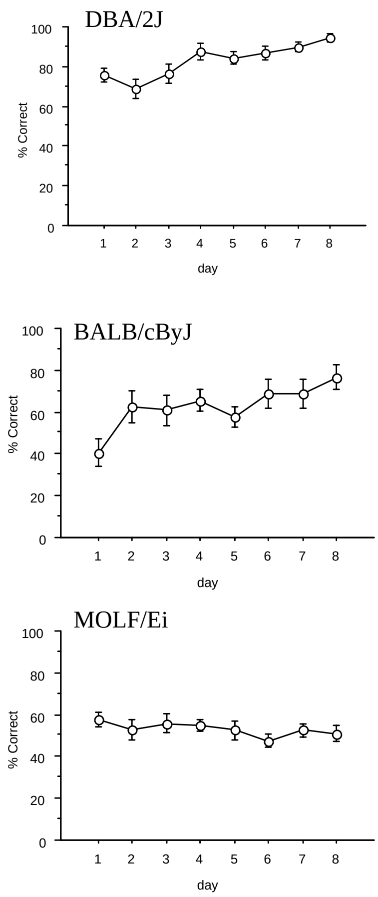
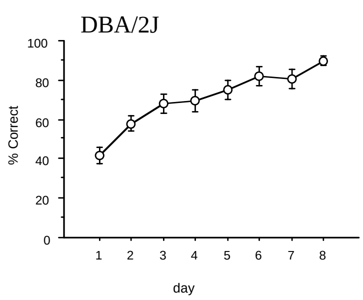


Figure 2. *Visual Discrimination task*
Mean (+ SEM) percent correct for a mouse strain with reported normal vision (DBA/2J), an albino strain (BALB/cByJ) and a strain with retinal degeneration (MOLF/Ei) on each of the eight days of testing in the visual discrimination task

3.2 Pattern discrimination task

All three strains of mice with reported normal vision and one albino strain (AKR/J) were able to reach the criterion of 70% correct within the testing period. 129S1/SvImJ and DBA/2J mice reached criterion on day five, while AKR/J and C57BL/6J required 6-8 days of testing. No other strains were able to reach criterion in the pattern discrimination task (Figure 3) [15].



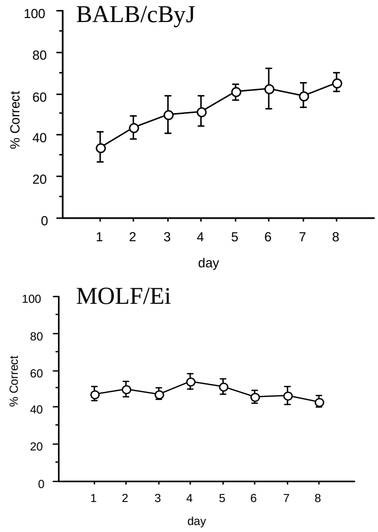
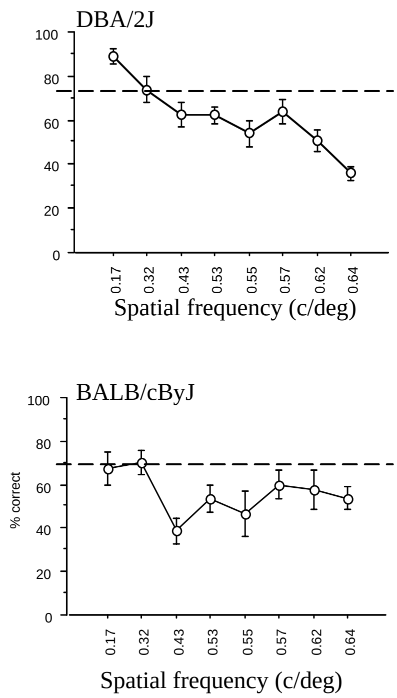


Figure 3. *Pattern Discrimination task*
Mean (+ SEM) percent correct for a mouse strain with reported normal vision (DBA/2J), an albino strain (BALB/cByJ) and a strain with retinal degeneration (MOLF/Ei) on each of the eight days of testing in the pattern discrimination task

3.3 Visual acuity task

A frequency-of-seeing curve was generated for each strain based on mean performance over eight days of testing and a visual acuity threshold was calculated when a strain dropped below 70% accuracy (Figure 4). All three strains of mice with normal vision and the two strains of albino mice (AKR/J and BALB/cByJ) performed above criterion with the low spatial frequency S+ (0.17 c/deg) that was used in the visual discrimination and pattern discrimination tasks. 129S1/SvImJ were unable to achieve 70% accuracy at the next spatial frequency (0.32), resulting in a visual acuity threshold of 0.245 c/deg. The albino strain (BALB/cByJ) was able to achieve criterion at 0.32 c/deg but not 0.43 c/deg. The other two normal sighted strains (C57BL/6J and DBA/2J) and the albino strain (AKR/J) had a visual acuity threshold of 0.365 c/deg, as they were able to achieve criterion at 0.32 c/deg but not 0.43 c/deg. No other strains were able to achieve criterion at any spatial frequency tested; therefore, their visual acuity threshold was set at 0 c/deg [15].



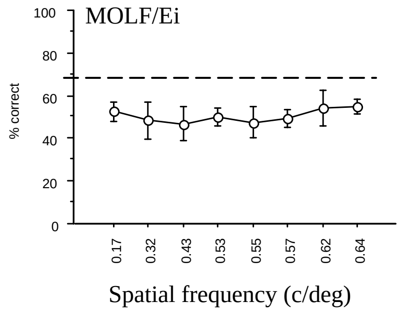


Figure 4. *Visual acuity task*
Calculation of visual acuity threshold from mean (± SEM) percent correct for an example of a strain with reported normal vision (DBA/2J), an albino strain (BALB/cByJ) and a strain with retinal degeneration (MOLF/Ei) on each of the eight spatial frequencies that were tested in the visual acuity task

3.4 Morris water maze

Mice with normal vision (129SI/Sv, C57BL/6J and DBA/2J) learned the task quickly, decreasing their latency (Figure 5) and distance to reach the platform over days (Figure 6) and spent the majority of the time in the probe trial in the correct quadrant (Figure 7). Albino mice (AKR/J, BALB/cByJ) showed a moderate decrease in latency (Figure 5) and distance to reach the platform over days (Figure 6) but did not show a preference for the correct quadrant in the probe trial (Figure 7). Mice with retinal degeneration (C3H/HeJ, MOLF/Ei and SJL/J) and other visual deficits (A/J, CAST/Ei, SM/J and SPRET/Ei) did not significantly decrease their latency (Figure 5) or distance to reach the platform over the six days of testing (Figure 6) and did not spend a significant amount of time in the correct quadrant during the probe trial (Figure 7) [14].

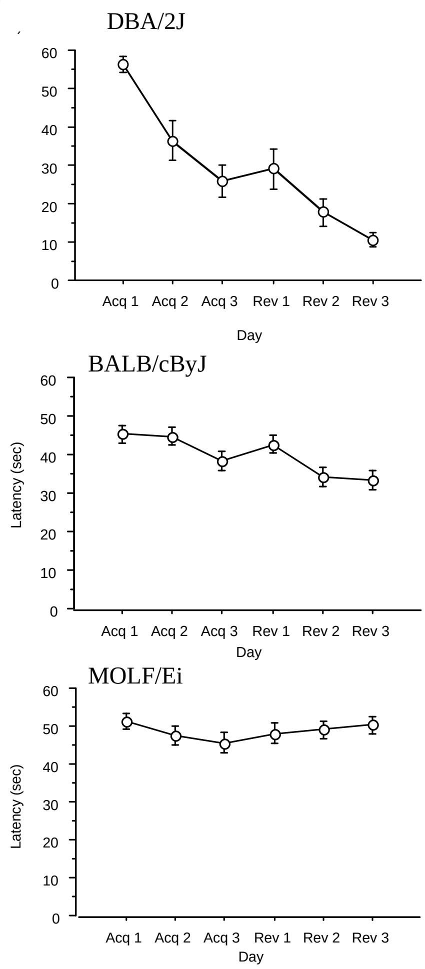


Figure 5. *Latency to find the platform in the Morris water maze*
Latency to find the escape platform for a strain with reported normal vision (DBA/2J), an albino strain (BALB/cByJ) and a strain with retinal degeneration (MOLF/Ei) over 6 days of training in the Morris water maze.

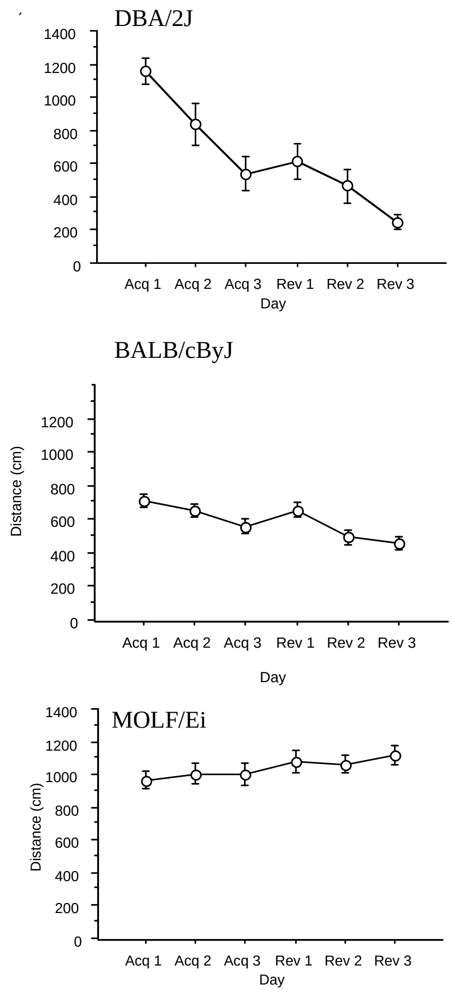


Figure 6. *Swim distance to the platform in the Morris water maze*
Swim distance to find the escape platform for a strain with reported normal vision (DBA/2J), an albino strain (BALB/cByJ) and a strain with retinal degeneration (MOLF/Ei) over 6 days of training in the Morris water escape task.

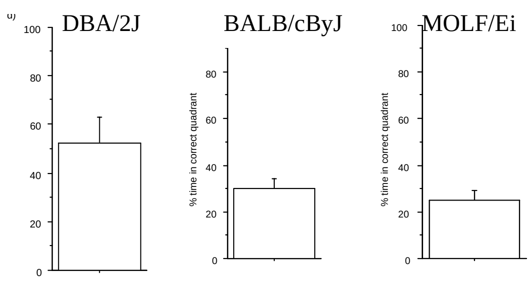


Figure 7. *Percent time spent in the correct quadrant in the Morris water maze probe trial*
Percent time spent in the correct quadrant for a strain with reported normal vision (DBA/2J), an albino strain (BALB/cByJ) and a strain with retinal degeneration (MOLF/Ei) during the probe trial on day seven of the Morris water maze.

3.5 Is performance in the Morris water maze correlated with visual ability?

There was a significant negative correlation between the % correct trials on day 8 of the visual discrimination task for each strain and the mean latency to find the platform in the Morris Water Maze during reversal days 4-6 (r = -.775, n = 13, p < .01) and the mean swim distance during reversal days 4-6 (r = -.718, n = 13, p < .01) but not the percent time spent in the correct quadrant during the probe trial (r = .347, n = 13, ns). There was also a significant negative correlation between the percent correct on day 8 of the pattern discrimination task of each strain and the mean latency to find the platform (r = -.737, n = 13, p = .01) and total swim distance to reach the platform during reversal days 4-6 (r = -.605, n = 13, p = .05) but not the percent time spent in the correct quadrant during the probe trial (r = .327, n = 13, ns). There was a significant negative correlation between visual acuity threshold and the mean latency to find the platform (r = -.754, n = 13, p = < .01) and total swim distance to reach the platform during reversal days 4-6 (r = -.705, n = 13, p = < .01) but not the percent time spent in the correct quadrant during the probe trial (r = .389, n = 13, ns) [14].

3.6 Is visual ability correlated with other measures that were evaluated in the behavioral test battery?

In tests for anxiety-related behavior in the open-field maze, frequency of center square entries was negatively correlated with percent correct trials on day 8 of the visual discrimination task (r = -.696, n = 13, p < .01) and visual acuity threshold (r = -.604, n = 13, p < .05) but not percent correct trials on day 8 of the pattern discrimination task (r = -.527, n = 13, ns). Proportion of time spent in the light zone in the light-dark transition test was not influenced by visual ability. Motor learning, as measured by latency to fall on day 7 of the Rotorod was not influenced by visual ability, as this task does not rely on visual cues. Interestingly, mice with poor visual ability performed better at an odour discrimination learning task, as percent correct trials on day 8 of the visual discrimination task (r = -.628, n = 13, p < .05) and percent correct trials on day 8 of the pattern discrimination task (r = -.758, n = 13, p < .02) were negatively correlated with percent digging in the CS+ [14].

4 Conclusions and recommendations

Strain differences in visual ability accounted for a significant proportion of the variance in behavioral tasks of anxiety and learning and memory tasks dependent on visual cues. The results reported here have important implications for studies that use these mouse strains for behavioral studies. Many behavioral tasks for rodents rely on visuo-spatial cues and a number of studies have suggested that performance in these tasks depends on visual ability [2, 10, 12, 13] but these warnings have gone unheeded, as many researchers have used visually impaired mice in visuo-spatial tasks. We recommend that the visual ability of all mice strains be tested in a vision task before they are used in behavioral tasks and performance data in these tasks be reported with respect to visual ability.

Supported by the JAX phenome project and NSERC of Canada (Grant A7441) to R.E.B.

References

1. Branchi, I.; Bichler, Z.; Berger-Sweeney, J.; Ricceri, L. (2003). Animal models of mental retardation: from gene to cognitive function. *Neuroscience & Biobehavioral Reviews*, **27**, 141-153.
2. Carman, H.M.; Mactutus, C.F. (2001). Proximal versus distal cue utilization in spatial navigation: the role of visual acuity? *Neurobiology of Learning and Memory*, **78**, 332-346.
3. Chen, G.H.; Wang, Y.J.; Wang, X.M.; Zhou, J.N. (2004). Accelerated senescence prone mouse-8 shows early onset of deficits in spatial learning and memory in the radial six-arm water maze. *Physiology & Behavior*, **82**, 883-890.
4. Crawley, J. N. (1985). Exploratory behavior models of anxiety in mice. *Neuroscience and Biobehavioral Reviews*, **9**, 37-44.
5. Crawley, J.N.; Paylor, R. (1997). A proposed test battery and constellations of specific behavioral paradigms to investigate the behavioral phenotypes of transgenic and knockout mice. *Hormones & Behavior*, **31**, 197-211.
6. LeMarec, N.; Lalonde, R. (1997). Sensorimotor learning and retention during equilibrium tests in Purkinje cell degeneration mutant mice. *Brain Research*, **768**, 310-316.
7. Lovasic, L.; Bauschke, H.; Janus, C. (2005). Working memory impairment in a transgenic amyloid precursor protein TgCRND8 mouse model of Alzheimer's disease. *Genes Brain & Behavior*, **4**, 197-208.
8. Morris, R. (1984). Developments of a water-maze procedure for studying spatial learning in the rat. *Journal of Neuroscience Methods*, **11**, 47-60.
9. Prusky, G.T.; West, P. W. R.; Douglas, R. M. (2000). Behavioral assessment of visual acuity in mice and rats. *Vision Research*, **40**, 2201-2209.
10. Robinson, L.; Bridge, H.; Riedel, G. (2001). Visual discrimination learning in the water maze: a novel test for visual acuity. *Behavioral Brain Research*, **119**, 77-84.
11. Schellinck, H.M.; Forestall, C.A.; LoLordo, V.M. (2001). A simple and reliable test of olfactory learning and memory in mice. *Chemical Senses*, **26**, 663-672
12. Thifault, S.; Lalonde, R.; Sanon, N.; Hamet, P. (2002). Comparisons between C57BL/6J and A/J mice in motor activity and coordination, hole-poking, and spatial learning. *Brain Research Bulletin*, **58**, 213-218.
13. Voikar, V.; Koks, S.; Vasar, E.; Rauvala, H. (2001). Strain and gender differences in the behavior of mouse lines commonly used in transgenic studies. *Physiology & Behavior*, **72**, 271-281.
14. Wong, A.A.; Brown, R.E. (2004) The influence of visual ability on the behaviour of 13 strains of mice in a battery of tests. *Society for Neuroscience, San Diego, CA*, Abstract # 83.7.
15. Wong, A.A.; Brown R.E. (2005). Visual discrimination, pattern discrimination and visual acuity in 14 strains of mice. Submitted to *Genes, Brain and Behavior*.

Principal component factor analysis of behavioral and neurochemical items in mice – a method to reveal relations between behavior and neurochemistry

M. Jähkel, L. Schiller, J. Oehler
Department of Psychiatry and Psychotherapy, Technical University, Dresden, Germany

Abstract

The principal component factor analysis [PCA] is a method to reveal interrelationships between measured items assembling parameters in factor groups by means of a multicorrelative procedure. The study reports mice data collected in different behavioral tests: magnetic impulse based running wheel activity, light beam based open field and plus maze activity, video recorded and manually scored social activity. In addition results of biochemical analysis are reported.

Parameters collected are subjected to test-related PCAs. Important steps and appropriate results as measure of sampling adequacy [MSA] criteria, communality, anti-image correlation values and factor loadings in a rotated factor matrix are described. PCA of the open field parameters demonstrates a clustering of behavior into activity, exploration and irritation. Results of the PCA including behavioral and biochemical data revealed linkages between behavior and neurochemical markers. Most evident were relationships between cortical dopamine, midbrain serotonin and aggressive behaviors on the one hand, but low midbrain serotonin and insecure behaviors on the other hand.

Keywords

running wheel, open field, plus maze, aggression, neurochemistry

Introduction

A variety of equipment allows more and more detailed observations and descriptions of animal's behavior. Commonly a lot of parameters is measured ranging from the frequency and duration of items to scoring the intensity of a behavior by a graded system. But the increasing number of parameters is difficult to handle especially when we need to choose the relevant parameters to be analyzed further. Using principle component factor analysis [PCA] as a multicorrelative procedure, variables were assembled into groups and possible functional relations may be revealed. Most relevant variables in one group [factor] were characterized by high loadings and may be used as parameters of further analyses.

Animal data of behavioral and subsequent neurochemical analyses should allow insight into the functional background of behavioral qualities as aggression, anxiety, drive and others. Revealing correlations between behavioral and neurochemical characteristics may be of impact to develop new therapeutic approaches in psychiatry where impulsivity and aggression on the one hand and depression and anxiety on the other hand were of special health care interest.

The presented investigations use the PCA of behavioral and neurochemical mice data to demonstrate correlative interrelationships. A large data pool was assembled to fulfil variety criteria. Male and female mice housed in groups and isolated were used for behavioral tests, and in addition trunk blood and different brain structures were collected from the same animals.

Considering the brief character of this paper only data management of the open field test and the results of the narrative PCA assembling the most important behavioral and neurochemical parameters are reported.

Methods

60 male and 60 female mice [NMRI, 4 weeks old, Charles River] were randomized for housing in standard group (5 per cage) housing cages [Ehret] or single (1 per cage) housing cages [Ehret] for 1 week, 6 or 12 weeks under controlled environments [Re. Prä. Dresden 24-9168.11-1-2002-5]. During the last week of housing behavioral tests were performed.

Immediately after the last test, animals were decapitated to obtain trunk blood for RIA corticosterone determination and brain structures for HPLC determination of catecholaminergic and indolaminergic transmitter parameters [cortex, hippocampus (hpc), striatum, midbrain / dopamine (DA), dopac (DO), homo-vanelic acid (HVA), serotonin (5HT), hydroxy-indol-acedic acid (HIA), norepinephrine (NE)]. For details see Rilke et al. 2001 (2.).

Behavioral analyses

Behavioral tests were carried out in a soundproofed chamber between 9 a.m. and 1 p.m. Between the tests two days of recovery were allowed. First the *running wheel* test was performed followed by the *open field* and *plus maze* test. The *social interaction* test was carried out on the last day immediately before decapitation.

Running wheel

The running wheel apparatus [TU Dresden] was comprised of 10 wheels placed side by side. The wheels were 13.5 cm in diameter and 4 cm wide. At the side of each wheel four magnets were arranged to record even slight movements of the wheel as 1 rotation impulse per 90° movement. Running wheel activity was scored every 15 minutes for 1 hour: The 30 minute record was revealed as most loading parameter by PCA.

Open field

The open field apparatus [TU Dresden] was made of brown plastic (56 cm x 28 cm x 19 cm) with 8 x 4 infrared light beams. Animals were placed in the center of the open field, facing towards the left small side. Light beam interruptions were recorded [n] and the movement pattern of the animal was determined. 18 different open-field parameters were computed for each animal within 10 min. For details see Jähkel et al. 2000 (1.). Parameters revealed as most loading by PCA are number of all squares crossed, long runnings in the middle, number of middle squares crossed, short runnings at the wall, time at the preferred square, long passive phases.

Plus maze

The plus maze [TSE] was made of gray plastic and was comprised of two open arms (32 cm long x 5 cm wide) and two enclosed arms (32 cm long x 5 cm wide) that extended from a central platform (6.5 cm x 6.5 cm). The test device was elevated 65 cm above the floor. Each mouse was placed on the central platform facing towards the left open arm. As activity parameters the number of entries into the open and enclosed arms as well time on these arms against time at the platform were recorded over a 6 min period by light beam interruptions. Two light beams, one at the side of the platform and the second 5 cm after each arm entry, must be disrupted before an arm entry was counted as valid. Time at platform, sum of all (closed + open arm) entries, number of open arm entries and % of time in closed arms were revealed as most loading parameters by PCA.

Social interaction

Mice were put into a clean glass container (Ø 16 cm, height 20.5 cm, without bedding) for 7 min to habituate as "resident". Thereafter, an unfamiliar group-housed mouse of the same sex ("intruder") was added and the session (6 min) was recorded on a videotape. Behaviors of the resident mice were observed blind with respect to the different strains. The number and time of locomotion, rearing, tail rattling, struggling, grasping, and any anal contacts were scored manually [TU Dresden].

Statistical analyses

Statistical management was performed using SPSS for Windows NT [SPSS Software GmbH, München, Germany)]. As dimension reduction procedure principle component factor analysis [PCA] with VARIMAX rotation was done.

Results and Discussion

PCA of open field parameters

At first those of the measured parameters have to be excluded which correlate ~1 with each other (active against passive time, or active time and all squares crossed) to receive a high score of measure of sampling adequacy [MSA] criteria, anti-image correlation values and communality values. Therefore the presented analysis includes only parameters, which were determined as "independent" by anti-image correlation values <0.950 received as a first step of the PCA (data not shown). Reducing "redundant" parameters the MSA criteria increased (data not shown).

Table 1. *Quality measures of open field parameters.*

parameter	anti-image correlation	communality
all squares crossed	0.784	0.973
central squares	0.727	0.941
time at center	0.620	0.840
frequency at first place	0.823	0.769
time at first place	0.841	0.873
frequency 15. / 1.square	0.528	0.542
mean of squares crossed per 40sec	0.825	0.910
passive time	0.906	0.927
frequency of passive phases >10sec	0.644	0.870
frequency of active phases at wall <4sec	0.883	0.804
frequency of active phases at centre >4sec	0.871	0.769
frequency of active phases at centre <4sec	0.766	0.711

The reported final PCA reached a MSA criteria (Kaiser-Meyer-Olkin criteria) of 0.805. This measure of sampling adequacy is *meritious* and hint on a representative data sampling and a high quality of the subsequent factor loadings in the rotated factor matrix. Table 1 presents the included variables, their anti-image correlation values and communality values. All parameters included - except the quotient frequency of crossings at the 15. place / frequency of crossings at the 1. place - reached anti-image correlation and communality values >0.600 indicating its adequacy and relevance as high loading parameters in the revealed factor groups of the rotated factor matrix.

Table 2 represents the factor loadings of each variable at the rotated factor matrix and the explained variance of the factors. Only factor loadings >0.300 are presented.

Table 2. *Rotated factor matrix of open field measures.*

parameter	F1 activity	F2 exploration	F3 irritation
all squares crossed	+0.943		
central squares		+0.932	
time at centre		+0.857	
Frequency at first place	+0.860		
time at first place	-0.387	-0.349	-0.800
Frequency 15. / 1.square			
mean of squares crossed per 40sec	+0.840		
passive time	-0.933		
frequency of passive phases >10sec	-0.660		+0.621
frequency of active phases at wall <4sec		+0.864	
frequency of active phases at centre >4sec	+0.877		
frequency of active phases at centre <4sec		+0.835	
Explained variance	48.133	25.860	9.561

The sum of explained variance (83.553 %) is very high and the loadings are clearly distributed between the three factor groups. Only frequency of passive phases longer than 10 seconds and time at the most often crossed place are relevant in more than one factor group. But considering its positive or negative correlation and its amount in the factor groups relations of most importance are obvious. Nearly 50% of the parameters measured are relevant for Factor 1, whereas parameters loading high in the second group reached a relevance of ~ 25% and parameters relevant for the third quality of locomotion registered during the open field test are relevant only up to 10% of the parameter sampling. Therefore our measurements reflect the different qualities of open field behavior in an non proportional manner and are mainly representative for the first factor group F1.

Parameters loading with highest values in factor group F1 are the commonly used activity parameter as squares crossed but also its negatively correlating counterpart passive time. Interesting is the loading of longer (>4sec.) activity movements at the centre in this F1 parameter group in contrast to short (<4sec) movements at the centre and at the wall, which loaded high in the second parameter group F2, where centre activity is characteristic. Obviously short movement patterns are qualitatively different from longer movement patterns. Considering the presentation of typical activity parameters in factor group F1 and the fact that centre activity should be rather less anxious with a high impact of curiosity, we suggest that variables loading high in F1 were mostly characteristic for the common activity state / drive, whereas parameters loading high in factor group F2 beside its activity dependence may be representative for explorative aspects of the behavior. In the third factor group only two parameters are of relevance correlating negatively with each other. This implies that longer (>10sec) phases of passivity are not registered at that square which is most often crossed during the test. Therefore we suggest that this component is representative for a rather anxious, uncertain behavioral state similar to freezing and named the factor group F3 irritation.

In general the open field PCA revealed three clearly distinct parameter groups which hint on distinct qualities of activity parameters and the PCA enabled the observer to find out the 6 most relevant parameters of 18 items measured and calculated by an automated and objective parameter registration.

In a similar manner parameters of all behavioral tests were evaluated by test related PCAs to find out the most representative parameters of each test performed.

A PCA was also performed with neurochemical

Table 3. *Quality measures of overall parameter sampling and loadings of F1 >0.300.*

parameter	anti image correlation	communality	F1 (activity)
all squares	0.544	0.879	+0.868
centre squares	0.579	0.861	+0.780
wall <4sec	0.540	0.856	+0.686
time at 1. place	0.397	0.837	-0.650
centre >4sec	0.611	0.635	+0.605
open arms	0.441	0.833	+0.598
all arms	0.559	0.755	+0.585
Passive >10sec	0.442	0.855	-0.057
midbrain DAin2	0.710	0.692	
cortex DAin	0.585	0.674	
striatum DA	0.521	0.802	
striatum HT	0.601	0.762	
midbrain NE	0.364	0.656	
cortex 5HTin	0.466	0.632	
hippoca. 5HTin	0.639	0.620	
running wheel	0.443	0.651	
t-struggling	0.411	0.834	
t-grasping	0.389	0.560	
t-locomotion	0.412	0.644	0.310
t-tail rattling	0.449	0.737	
cortex DAin2	0.414	0.641	
corticosterone	0.497	0.487	
t-anal sniffing	0.389	0.716	
midbrain HIA	0.414	0.826	
midbrain 5HTin	0.394	0.815	
striatum HVA	0.352	0.723	
t-selfgrooming	0.280	0.630	
Midbrain DAin	0.458	0.657	
t-rearing	0.499	0.771	
t at platform	0.595	0.733	0.306
t-freezing	0.406	0.660	
cortex NE	0.334	0.635	
t-closed arms	0.198	0.908	

and corticosterone data received after HPLC or RIA of materials obtained from the same animals.

The most relevant parameters of behavioral and biochemical tests are included in an narrative PCA.

Table 3 presents in the first two columns the qualitative values of the narrative PCA. This PCA reached a MSA criteria of 0.489 which is only *miserable*. Considering that some of anti-image correlation values are <0.500 not only for a lot of the biochemical data but also for behavioral parameters a improvement should be possible excluding those of the parameters which demonstrated also low communality values. For instance grasping could be excluded because of the fact that the other aggression related parameters struggling and rattling demonstrated higher measures of sampling adequacy. In the same manner all biochemical and behavioral data are to value. Only such parameters not to compensate by an similar item should be further included - for instance corticosterone can not be removed. Excluding the redundant parameters in further PCAs a better MSA criteria and a more clear picture of correlative interrelationships should be reached.

On the other hand the narrative PCA results are a necessary step to see correlations free from subjective data handling. Factor groups F1 – F5 reflect therefore this narrative result and are presented in Table 3 (F1) and Table 4 (F2-F5). Only factor loadings >0.300 are demonstrated.

In the factor group F1 only behavioral data are sampled. Obviously they have to be reduced to avoid dominance

Table 4. *Rotated factor matrix of the biochemical and behavioral parameters.*

parameter	F2	F3	F4	F5
all squares				
centre squares				
wall <4sec		+0.331		
time at 1. place				
centre >4sec		-0.407		
open arms				
all arms			-0.500	
Passive >10sec	+0.361	+0.327		
midbrain DAin2	+0.731			
cortex DAin	+0.715			
striatum DA	-0.713			
striatum HT	-0.626			
midbrain NE	+0.578			
cortex 5HTin	+0.559			
hippoca. 5HTin	-0.402			
running wheel	+0.308			
t-struggling		+0.675		
t-grasping		+0.612		
t-locomotion		-0.591		
t-tail rattling		+0.584		+0.462
cortex DAin2		+0.582		
corticosterone		-0.461	-0.316	
t-anal sniffing			-0.743	
midbrain HIA			-0.661	
midbrain 5HTin		+0.336	-0.598	
striatum HVA			-0.474	
t-selfgrooming			+0.448	
Midbrain DAin				
t-rearing	0.315		+0.320	-0.694
t at platform				-0.660
t-freezing				+0.586
cortex NE				+0.543
t-closed arms				-0.424

of correlations between activity parameters by itself. The same should be done with biochemical parameters considering its high and dominant presence especially in factor group F2. Only in parameter group F4 a balanced number of behavioral and biochemical data is sampled. Further steps of PCA consecutively excluding redundant parameters are necessary.

Nonetheless important interrelationships between behavioral parameters and biochemical items are found. In factor group F3 where aggression and activity parameters are sampled we see similar loadings of the cortical dopamine metabolism, midbrain serotonin metabolism and the negatively correlating trunk blood corticosterone content. This implies, that aggression will be observed, when midbrain serotoninergic neurons and cortical dopaminergic neurons are active but the HPA stress axis remains not activated.

On the other hand parameter loadings in factor group F4 suggest that low midbrain serotonergic activity is associated with “insecure” behaviors as self grooming and rearing. An interesting sampling pattern is also found in factor group F5 where high cortical norepinephrine and the aggression sign tail rattling are negatively correlated with orienting behaviors as rearing and platform observation. Considering that a balanced mode of activity - exploring before acting - in a insecure environment would improve fitness it may be suggested that a cortical norepinephrine overflow could force imbalanced impulsive behaviors such as aggression.

In general the reported procedures of PCA demonstrate its impact to choose out variables for subsequent data analyses on the one hand and on the other hand different qualities of behavioral characteristics my be revealed analyzing the pattern of parameter sampling in different factor groups. Considering that correlations are revealed also between behavioral and neurochemical items narrative PCA can be used to focus further neurochemical analyses.

References:

1. Jähkel, M.; Rilke, O.; Koch, R.; Oehler, J. (2000). Open field locomotion and neurotransmission in mice evaluated by principal component factor analysis – effects of housing condition, individual activity disposition and psychotropic drugs. *Progr. Neuro-Psychopharmacol. & Biol. Psychiat.* **24**, 61-84.
2. Rilke, O.; Will, K.; Jähkel, M.; Oehler, J. (2001). Behavioural and neurochemical effects of anpirtoline and citalopram in isolated and group housed mice. *Progr. Neuro- Psychopharmacol. & Biol. Psychiat.* **25**, 1125- 1144.

The starmaze: a new paradigm to characterize multiple spatial navigation strategies

L. Rondi-Reig[1*], G. Petit [1,2], A. Arleo[1&], E. Burguière[1]
[1] *Laboratoire de Physiologie de la Perception et de l'Action, UMR CNRS 7152, Collège de France, Paris, France*
[2] *Laboratoire Neurobiologie des Processus Adaptatifs, Equipe Développement et Vieillissement du Système Nerveux, UMR CNRS 7102, Université P&M Curie, Paris, France*
[&] *Present address: Neuroscience Group, Sony Computer Science Laboratory, Paris, France.*
** Correspondence should be sent to LRR*

Abstract

Spatial orientation disorders constitute a major problem in aged subjects. Among the different cognitive strategies used for spatial navigation, the most complex ones seem to be primarily deteriorated during ageing. Being able to dissociate the different cognitive strategies might help us in the early detection of age-related brain dysfunctions. In order to study multiple spatial navigation strategies, a new task was designed and named 'starmaze'. We first developed this paradigm for animal experiments. In this paradigm, up to three different strategies of navigation can be dissociated. The starmaze is composed of a central pentagonal ring from which five alleys radiate. A hidden goal has to be found. To locate the target, the subject can use either distal cues (allocentric strategy), or proximal cues located on the inner walls (guidance strategy), or a sequence of body movements leading to the goal (egocentric strategy). Four versions of the task can be used: *(i)* the multiple strategies, *(ii)* the allocentric, *(iii)* the egocentric, and *(iv)* the guidance version. In the multiple strategies version, the task has two components: the first assesses the learning capability of the subject; the second permits the identification of the strategy spontaneously used by the navigator to solve the task (multiple strategies can be employed: allocentric, egocentric and guidance). The allocentric version of the starmaze requires the subject to learn to reach the goal from different starting points and therefore necessitates a spatial representation of the environment. In the egocentric and guidance versions of the task there are no distal extra-maze cues (the apparatus is surrounded by black uncued curtains); subjects need to either perform a left-right-left movement sequence or to use intra-maze cues (e.g. to first approach a chessboard-like, then a black, and finally a white wall) to reach the target from different departure points. Data acquisition is performed by means of a video recording system and/or a tracking software. A set of parameters are measured to characterize the spatial behavior of the subject quantitatively (for example the distance traveled during each trial or the number of visited alleys). Data processing is automated via a MATLAB batch program developed in our laboratory.

Keywords

Spatial navigation – starmaze – allocentric strategy sequential guidance – sequential egocentric –

1 Introduction

Spatial navigation is a crucial function for many animal species including humans and is primarily deteriorated during aging. Spatial navigation toward an invisible goal involves the ability of a navigating animal or human to first acquire spatial knowledge (e.g. spatio-temporal relations between environmental cues) and to organize it properly. This requires the integration of multimodal sensory signals into a coherent representation. Second, the animal or the human needs to employ this spatial knowledge to adapt its motor behavior to the specific context in which the navigation takes place. Determining and maintaining a trajectory from one place to another calls upon multiple concurrent processes and demands the ability of the subject to adapt a goal-directed strategy to the complexity of the task. Multiple strategies can be employed, including route-based strategies such as learning a sequence of self-movements (egocentric strategy) or a sequence of visual stimuli (guidance) and map-based (allocentric) strategy (Arleo and Rondi-Reig, in press). The 'starmaze' task was designed in order to characterize the different navigation strategies used during a spatial behavior.

2 The starmaze apparatus

2.1 A five-branch maze

The starmaze consists of five alleys forming a central pentagonal ring and five alleys radiating from the vertices of this pentagonal ring (Figure 1). The entire maze is inscribed in a circle (diameter 204 cm) and all the alleys are filled with water made opaque with an inert non-toxic product (Accuscan OP 301). To solve the task, subjects have to swim to a platform hidden below the water surface.

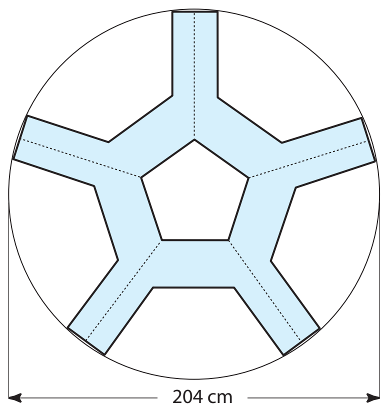


Figure 1. *Top view of the starmaze apparatus. The light blue regions are the alleys in which the subject has to navigate.*

2.2 Experimental protocols

We have developed four versions of the starmaze task.

***(i)* The 'multiple strategies' version**

This version was designed to permit the identification of the learning strategy spontaneously used by each animal. In order to achieve this characterization, this multiple strategies version of the starmaze task has two components. The first, called *training test*, lasts twenty days and assesses the animal learning abilities. The second component is called the *probe test* and is designed to identify the strategy used by an animal during the training part of the task. Every five days, one probe test is inserted between two training trials.

In order to learn and then perform the optimal trajectory to the goal (see Figure 2), animals can either use distal visual cues (represented in the figure by the white crosses, the black circle and the black and white stripes), or follow a sequence of intra-maze cues (chessboard-like, black, and white painted walls), or use a sequence of self-movements (turning left, right and then left). We called these three strategies allocentric, sequential guidance, and sequential egocentric, respectively.

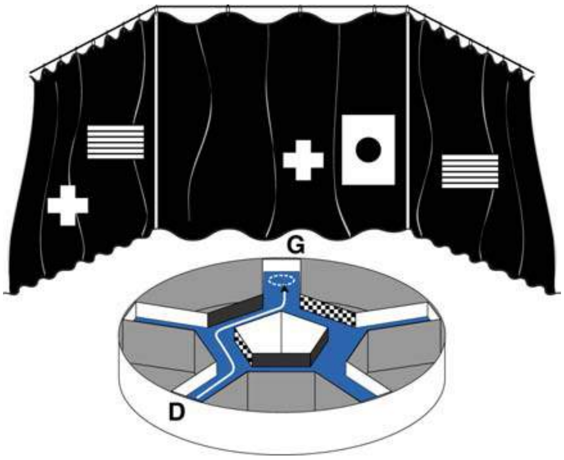


Figure 2. *The 'multiple strategies' protocol: the first component or training part. D is the departure location, G is the goal consisting of an immerged (and therefore invisible) platform (dotted circle). The arrow represents the optimal trajectory performed by the animal after learning. The alleys forming the central pentagonal ring have either black or chessboard-like walls, whereas the radial alleys have white walls. The maze is placed inside a black square curtain with distal visual cues attached on the curtains (crosses, circle, black and white stripes).*

The probe test has been designed to identify which strategy is used by each animal once it has learned to perform the optimal trajectory from the departure point D to the goal G. This probe test relies upon the assumption that an animal occasionally placed at a different departure point will continue to employ the same cues it used during training. This working hypothesis has been already successfully adopted in the cross-maze protocol developed by Barnes, Nadel and Konig (1980). The departure point of the probe test is chosen in order to dissociate the three kind of strategies described previously. This is possible when the D1 departure point is used (see Figure 3). If the animal used the distal visual cues when trained from the departure D, he will continue to use them when placed at the departure location D1 and therefore it will reach the same goal G. On the other hand, if the animal used the proximal visual cues, he will follow the sequence of chessboard-like, black and then white walls and it will arrive at the goal G1. Similarly, if the animal used the idiothetic cues during training, (i.e. it learned to turn left, then right, and then left), it will arrive to the goal G2 goal. During the probe test, three platforms are placed at locations G, G1, G2, corresponding to the three different navigation strategies. Most animals navigate without hesitation to the goals G, G1, or G2, allowing us to characterize their strategy during the learning component. In a few cases, the trajectory followed by the animal can be different from the three described above; we conclude that the animal demonstrates 'no clear strategy' during the probe test.

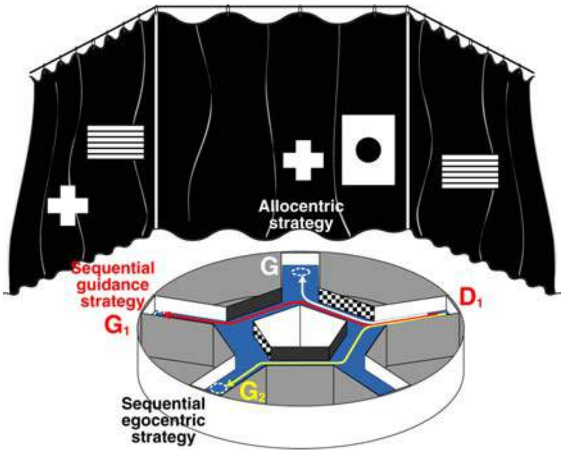


Figure 3. *The 'multiple strategies' protocol: the second component or probe test. As described in the text above, the new departure point is fixed (D1) and G, G1 and G2 goals correspond to allocentric, sequential guidance and sequential egocentric strategies, respectively.*

***(ii)* The allocentric version**

This version was designed in order to evaluate the ability of animals to learn a spatial navigation task using exclusively the distal visual cues. Proximal intra-maze cues are removed and each animal is placed in a randomly-selected alley (D1, D2, D3, D4) not containing the platform (Figure 4). In order to learn this task and use the optimal trajectory from each departure points, animals need to encode an allocentric representation of the environment. Similar to the Morris water maze (Morris et al., 1982), solving this version of the 'starmaze' task implies spatial learning capabilities to find a hidden platform from different starting points. However, here animals are constrained to swim in alleys that guide their movements, which permits a detailed analysis of the animals' trajectories.

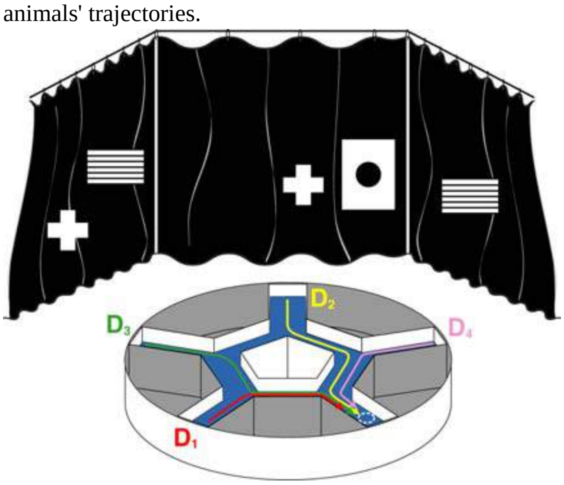


Figure 4. *The allocentric version of the starmaze. D1, D2, D3 and D4 represent the four possible departure locations randomly selected during the training period. White arrows represent the four optimal trajectories.*

(iii, iv) The guidance and egocentric versions

In these two versions extra-maze cues are removed and a circular black curtain is placed around the maze. During the guidance version (Figure 5) the same sequence of proximal cues is rewarded from different starting points. Each departure location D_i is associated to a specific goal location G_i requiring the animal to follow chessboard-like, then black, and finally white walls to solve the task. In the egocentric version of the task (Figure 6), intra-maze cues are also removed (note that all inner walls are white). During learning, the same sequence of movements is rewarded while varying the departure points. Each starting location D_i is associated to a specific goal G_i requiring the animal to turn left, right and then left. These strategies organized in sequence can be tested thanks to the existence of five alleys radiated from a pentagonal ring.

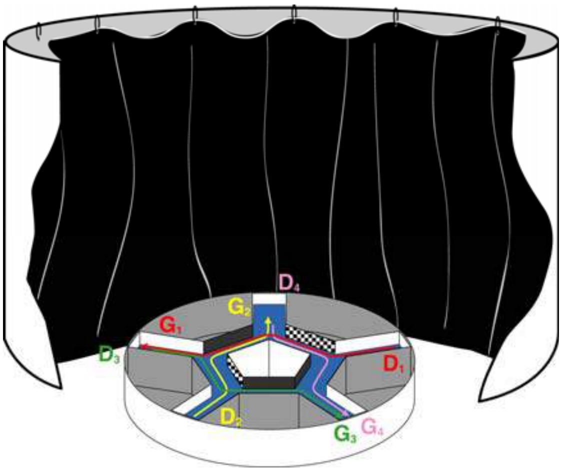


Figure 5. *The guidance version of the starmaze.*

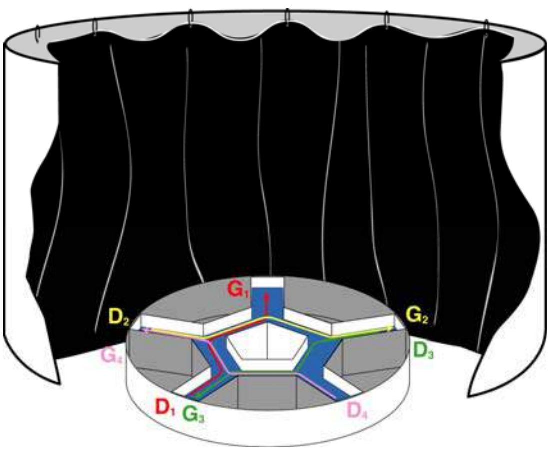


Figure 6. *The egocentric version of the starmaze.*

3 Data acquisition and analysis

We used a videotracking software (SMART@: Spontaneous Motor Activity Recording and Tracking, v2.0) designed for monitoring the behavioral activity of small animals in real time. This software provides a graphic editor suitable to define the zones in which we decide to measure the different behavioral parameters (Figure 7). Based on the defined zones, the tracking software generates a MS Excel file containing the position of the animal (the X-Y Cartesian coordinates of the center of mass of the image representing the animal, sampled every 200 ms) over time. The alley in which the animal swims (Z variable) and the mouse number (subject) are also reported in this file (see Figure 8).

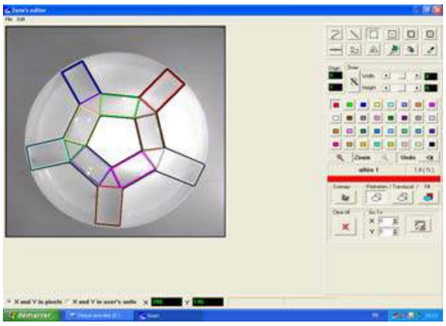


Figure 7. *Snapshot of the computer screen displaying the different zones we defined inside the starmaze (left part). The right part shows the different tools of the graphic editor available with this software.*

REPORT # 101						
From	00:00:00	to	00:00:08			
Sample	T (sec)	X	Y	Z. Nu	Z. Name	Subject
1	0	152.05	68.033	1	allée 1	89
2	0.2	149.18	73.361	1	allée 1	89
3	0.4	147.53	77.869	1	allée 1	89
4	0.6	145.07	81.967	1	allée 1	89
5	0.8	142.19	84.426	11	inter 1	89
6	1	137.67	87.705	11	inter 1	89
7	1.2	132.74	92.213	10	allée 10	89
8	1.4	128.22	95.902	10	allée 10	89
9	1.6	126.58	97.541	10	allée 10	89
10	1.8	127.4	98.77	10	allée 10	89
11	2	129.04	100	10	allée 10	89
12	2.2	131.51	102.46	10	allée 10	89
13	2.4	134.38	106.15	10	allée 10	89
14	2.6	136.85	110.66	10	allée 10	89
15	2.8	138.9	115.57	10	allée 10	89
16	3	139.73	119.67	10	allée 10	89
17	3.2	137.26	123.36	15	inter 9	89
18	3.4	134.38	125.41	15	inter 9	89
19	3.6	131.51	125.82	8	allée 8	89
20	3.8	130.27	127.05	8	allée 8	89
21	4	128.63	129.1	8	allée 8	89
22	4.2	128.22	133.61	8	allée 8	89
23	4.4	127.81	137.7	8	allée 8	89
24	4.6	126.16	141.39	8	allée 8	89
25	4.8	123.7	145.9	8	allée 8	89
26	5	121.23	150.41	8	allée 8	89
27	5.2	119.59	153.28	8	allée 8	89

Figure 8. *Example of MS Excel file provided by the SMART@ software. The alley in which the animal swims (Z. name) and the mouse number (subject) are reported in this file (see Figure 7).*

The data provided by the tracking system are employed for the further analysis of specific behavioral parameters to quantify the navigation performance of the subjects. A MATLAB batch program has been developed to process all the collected data automatically. The program goes through all trials (e.g. for example 25 animals, 4 trials per day for each animal, and 10 days result in 10^3 trials) and for each trial it computes: *(i)* the time needed to reach the target (termed escape latency); *(ii)* the average speed of the animal; *(iii)* the mean and cumulative distance between the animal and the target; *(iv)* the total distance traveled by the animal; *(v)* the mean and cumulative egocentric angle between the direction of motion of the animal and the direction to the platform; *(vi)* the amount of time spent in each region of the maze; *(vii)* the number of arms visited by the animal before reaching the goal (see Figure 9). Then, the program calculates the mean values of each behavioral parameter by averaging over all the day trials for each subject and/or for each group of subjects (e.g. control and mutant mice). The results are saved as MS Excel files, one for each behavioral parameter, and used to perform the statistical analysis (e.g. t- and Anova tests) by means of the Statview 5.0 commercial software.

Parameters measured \ Methods	Manually	Tracking software	MATLAB batch program
Latency to reach the goal	++	++	++
Number of arms visited	+	+	++
Sequence of arms visited	+	+	++
Mean animal speed	-	++	++
Total distance travelled	-	++	++
Mean distance between animal and goal	-	-	++
Mean angle between animal heading and goal	-	+	++
Time spent in different areas	-	+	++

Figure 9. *The principal behavioral parameters and their observation with different acquisition methods. Symbols: '-' not possible; '+' possible but not automatic; '++' very easy or automatically executed.*

The MATLAB program can also be employed to build three-dimensional (3d) representations of the trajectories followed by the animals at different learning stages (Figure 10). These plots provide a qualitative measure of the ability of a subject (or a group of subjects) to perform optimal goal-oriented behavior. To build these 3d plots, the MATLAB program samples spatial locations by means of a uniform grid (resolution 30x30, each grid cell corresponding to a 5x5 cm area). Then, each cell is given a value representing the time spent by the subject in that region (this values is normalized relative to the duration of the trial, then averaged over many trials and eventually over all the subjects of a same group).

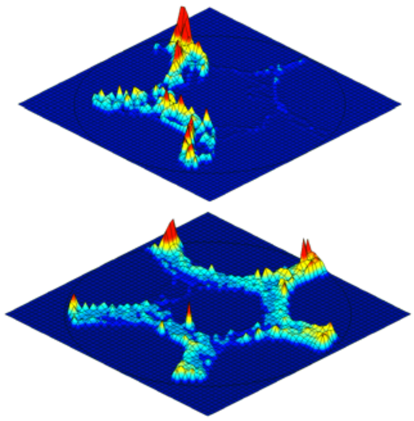


Figure 10. *Two samples of 3d plots providing a qualitative description of the animal behavior when solving the starmaze task. Example of an optimal (top) and a non optimal (bottom) trajectory during navigation.*

5 Conclusion

The design starmaze task took inspiration from the combination of the Morris water maze (Morris et al., 1982) and the cross-maze (Barnes et al., 1980; O'Keefe and Conway, 1980; Packard and Mc Gaugh, 1996; see White and McDonald for review). Similar to the Morris water maze, the complexity of the task requires a subject to develop a map-based (allocentric) strategy. As for the cross-maze, alternative strategies can be used and identified such the egocentric and the guidance strategies. Furthermore, this new task adds the possibility to test route-based strategies such as sequential guidance and sequential egocentric procedures combined with a possible map-based strategy. Therefore, the starmaze allows us to study the ability of a subject to use one of these complex strategies spontaneously. Note that sequential egocentric and sequential guidance as defined in the starmaze refer to route-based strategy as they require a sequential organization of the information. This is different from target approaching or stimulus-triggered response as defined in the cross-maze. Such a difference makes it possible to investigate the distinct neural and cellular bases mediating simple versus sequentially-organized strategies. Finally, the starmaze gives also the possibility to test the subject's navigation capability when it is forced to use a specific strategy.

Multiple applications of the starmaze can be imagined. We are currently employing it to study the neural bases of spatial navigation by using a behavioral genetic approach with conditional mutagenesis models (Rondi-Reig et al., 2004; Burguiere et al., 2004), and to develop a test for early detection of age-related dysfunctions (Petit et al., 2005). We have also adapted the starmaze to humans using virtual reality (Rondi-Reig et al., 2005).

This work was supported by grants n°SRI20001117030 from GIS "Longévité" (grant L0201) and ACI 'Neurosciences Intégratives et Computationnelles' (NIC 0083). We thank F. Maloumian for her help for the figures, S. Thierry for her help in developing the Matlab batch program and A. Lohof for reviewing the article.

References:

1. Arleo, A; Rondi-Reig, L. (in press). Multimodal sensory integration and concurrent navigation strategies for spatial cognition in real and artificial organisms. In F. Dolins and R. Mitchell (eds.), *Spatial Perception, Spatial Cognition*, Cambridge University Press.
1. Burguière, E.; Arleo, A.; De Zeeuw, C. I.; Berthoz, A.; Rondi-Reig, L. (2004). Deficit during spatial navigation in L7-PKCI mice lacking cerebellar long-term depression: a motor learning problem? *FENS Abstr.*, **2, A042.5**, Lisbon, Portugal.
2. Barnes, C.A; Nadel, L.; Honig, W.K. (1980). Spatial memory deficit in senescent rats. *Canadian Journal of Psychology*, **34**, 29-39.
3. O'Keefe, J. and Conway, D.H. (1980). On the trail of the hippocampal engram. *Physiological Psychology*, **8**, 229-238.
4. Packard, MG; McGaugh, JL (1996). Inactivation of hippocampus or caudate nucleus with lidocaïne differentially affects expression of place and response learning. *Neurobiol Learn Mem*, **65**, 65-72.
5. Petit, G.; Fouquet, C.; Tobin, C.; Berthoz, A.; Mariani, J.; Rondi-Reig, L. (2005). Early detection of spatial memory deficit during aging in mice. Abstract. Poster G-49 in *Proceedings of the 7th French Society for Neuroscience Meeting*, Lille, France.
6. Rondi-Reig, L.; Petit, G.; Tobin, C.; Tonegawa, S.; Mariani, J.; Berthoz, A. (2004). Impaired sequential-egocentric and allocentric memories in hippocampal CA1-NMDA receptor knockout mice during a new task of spatial navigation, *Society for Neuroscience*, San Diego, USA.
7. Rondi-Reig, L.; Berthoz, A.; Zaoui, M; Igloi, K. (2005). De la souris à l'homme: le labyrinthe en étoile, un paradigme pour caractériser les stratégies de navigation. *Séminaire Images virtuelles*, Paris, France.
8. White, NM; McDonald, RJ (2002). Multiple parallel memory systems in the brain of the rat. *Neurobiol Learn Mem*, **77**, 125-184.

The behavioral differentiation between the Carrier and the Non-Carrier profiles in groups subjected to the diving-for-food situation: a complex social model to study anxiety in rodents

H. Schroeder, D. Desor
Neurosciences Comportementales, URAPA, INPL-UHP, BP 239, 54500 Vandoeuvre-les-Nancy, France

Abstract
The diving-for-food situation is a complex social task that consists of rats swimming in apnoea to reach food through a fully immersed pool, bringing a pellet back to the cage and eating it. The progressive immersion of the only way of access to the feeder induces the emergence of a social differentiation in a group of 6 rats between the Carriers who dive and swim to get food, and the Non-Carriers who never dive and get their food only by stealing it from Carriers. Both ethological and pharmacological studies demonstrated a role for individual characteristics related to anxiety towards the water constraint and the social context in the acquisition of the Carrier or Non-Carrier profile This suggests that the diving-for-food situation may be relevant to investigate the neurobiological substrates of anxiety related to the social adaptation of a group faced to an environmental constraint, and to study the pharmacological properties of various drugs like anxiolytics or psychoactive substances.

Keywords
Social behavior, rat, anxiety

1 Introduction
Many studies dealing with social phenomena have emphasized the influence of environment on social systems [3,13]. Because of its importance for the survival of the individuals, access to alimentary resources is one of the most determining constraints for the social organization and structure of a group [14]. When animals forage in groups, individuals differ in their investment in searching for food and in their exploitation of the food found by the others. For example, individuals may specialize in stealing food from the others [1], or in joining the others that have already located a source of food [5]. The diving-for-food model is an experimental situation where the food accessibility is made difficult by progressively immersing the only way of access with water, leading to the emergence of a behavioral differentiation in groups of rats faced with this environmental constraint [2]. To be more precise, a single food source is located at the end of an aquarium, so that rats have to dive and swim under water to get the food and they have to go back to the cage where congeners are present, to eat it. In these conditions, a behavioral differentiation based on social interactions appears between Carrier rats which dive and bring the food back to the home cage, and the Non-Carrier animals which never dive and obtain their food by stealing it from the other members of the group. This situation reproduces a phenomenon observed on the Pô river bank in free-living populations of rats [4], and discloses several behaviors that are described in the wild rat such as diving, carrying or stealing [11,16]. The present paper will present the apparatus, the experimental procedure and the behavioral characteristics of this model. The perspectives about its use in the field of psychopharmacological researches will also be discussed.

2 Materials and methods

2.1 Animals
The experiments are performed on adult Wistar male rats (Iffa-Credo Breeding Laboratories, L'Arbresle, France) weighing 300-340 g. The animals are maintained under standard laboratory conditions on 12/12 h light/dark cycle (lights on at 06:00 a.m.) until the time of the study. Food and water are available *ad libitum*. The rats are randomly selected and mixed together to form groups of 6 rats each. The diving-for-food experiment is started one week after constituting the groups. All animals were treated according to the rules provided by the ASAB Ethical Committee.

2.2 Apparatus
The apparatus is depicted in Figure 1. It consists of a home cage (50 x 40 x 40 cm) and an aquarium measuring 150 x 25 x 40 cm (l x w x h) at the end of which a single feeder is placed, distributing one food pellet at a time. The cage is linked to the aquarium by a tunnel, and a sliding door controls the access to the tunnel. The maximum water level is fixed by a transparent cover over the aquarium, forcing the rats to swim under water. The apparatus is designed to oblige rats to dive and swim under water for a distance of one meter and to make it nearly impossible for them to eat the food at the feeder, and therefore to oblige the rats to carry the pellet of food back to the home cage. The floor of the cage consists in a grid in order to avoid the persistence of other sources of food in the cage such as faeces material or small particles of debris.

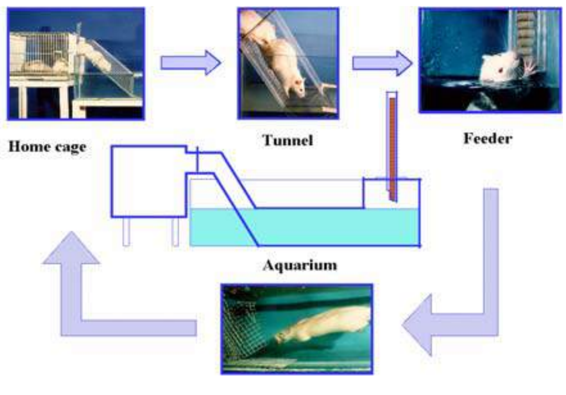


Figure 1. The diving-for-food situation: the experimental set-up

2.3 Experimental procedure
The diving-for-food experiment is conducted according to the following time schedule: On day 1, day 2 and day 3,

the aquarium is empty of water to allow the rats to learn the spatial characteristics of their environment and the food location; from day 4 to day 10, the aquarium is progressively immersed until the maximum water level is reached. At this latter stage, rats need to dive and swim under water to reach the feeder. Throughout the experiment, access to the feeder is limited to a daily 3-hour period in the experimental apparatus, and the rats do not have any other source of food during the rest of the day. Water is available *ad libitum*. Each day, rats are weighed 30 min prior to the experimental session (the weight of an animal on a given day reflect the food consumption of the day before). In these conditions, the body weight of rats faced with this experimental situation remains comparable to that of animals housed in standard cage [2,8]. Rats are observed with video recording on day 18. This day corresponds to the eighth day of complete immersion of the aquarium, and allows the individuals to be characterized for the behavioral profile they adopted in the group when confronted with the feeding problem. By that time, behavioral differentiation between Carrier and Non-Carrier rats is well established and remains stable [2]. The home cage is videotaped during the entire feeding session (3h). The session is recorded in daylight. The day before recording, each animal is individually marked on its coat with colored pens to allow individual identification on videotape.

2.4 Behavioral scoring

For each group and each experimental session, the observation lasts 3 h and is videotaped. Then, behavioral items related to feeding are analyzed. In this experimental situation, a food possession period, i.e., when a rat touches food, may start in one of the three following ways: The rat carries food, steals food, or picks up food from the floor. It may end in one of the four following ways: The rat eats the food pellet until its complete ingestion, the pellet is stolen by another rat, the pellet is lost (during an agonistic encounter), or the pellet is left on the ground [2,15]. The owner of the food may be attacked by other rats. Stealing attempts start by an approach to which the owner may respond by fleeing or kicking. Some approaches end by full attack, some of which evolve to simultaneous biting of the pellet for at least 1 s. The attacks may or may not lead to the appropriation of the food [15]. Rats are characterized for 3 groups of variables which are presented in Table 1. The first one was the social differentiation between Carrier and Non-Carrier rats (see the paragraph below). The second group of variables concerns the general behavioral characteristics and includes (a) the total time of food possession along the experimental session; (b) the total number of food possession periods along the experimental session; (c) the total number of dives; (d) the total number of carryings; (e) the diving efficiency, that is, the percentage of the number of food carryings in proportion to the number of dives; and (f) the total number of food possession periods ending by the complete ingestion of the pellet. The total number of pickings of food from the floor as the total number of losses during an agonistic encounter are also noticed but these events remain exceptional. The third group of variables is related to social interactions and food exchanges. It includes (a) the total number of thefts of food pellets they committed; (b) the feeding time they obtained by theft; (c) the total number of thefts of food pellets they experienced; (d) the feeding time they lost due to a theft after an agonistic encounter; (e) the food provisioning index, that is, the percentage of the quantity of food a rat lost in proportion to the total quantity of food the rat was able to obtain during the experimental session; (f) the total number of attacks committed on a rat eating food; and (g) the total number of attacks experienced by the owner of food.

2.5 Behavioral profiles

In the diving-for-food situation, three social profiles can be defined: one corresponds to rats which do not dive obtain food only by stealing it from the divers (Non-Carrier rats), whereas the two others correspond to animals which dive and bring food pellets back to the cage (Carrier rats). Depending on their abilities to protect food, the Carriers can be differentiated into two subgroups: (a) the Autonomous Carrier rats (AC rats) which are autonomous in the sense that they can hold the food they get by diving; and (b) the Supplier Carrier rats (SC rats) which have to feed various numbers of Non-Carrier congeners before eating.

2.6 Statistical analysis

Individuals are classified into groups of similar behavioral profiles by means of a hierarchical agglomerative cluster analysis of Euclidean distances using Ward’s method. Then, an inspection of the final data groups using distribution-free methods (Kruskal-Wallis and Mann-Whitney tests) is performed to identify traits that are associated with the three different behavioral profiles. The significance of all statistical tests is set at .05. Statistical analysis are performed in SPSS for Windows version 11.5 (SPSS Inc.).

3 Non-Carrier, Autonomous Carrier and Supplier Carrier behavior

As presented in Table 1, no significant variation of weight could be observed among rats of the same status, in any of the three behavioral profiles. According to these results, NC, AC and SC rats could not be distinguished in terms of food possession time and number of food possession ending by the ingestion of the pellet, suggesting that rats adapt to the diving-for-food situation with respect to food source utilization even if individual strategies differ. Concerning the behavioral characteristics of each profile, the Supplier Carriers touch food more often than the AC or the NC rats (Table 1),. They undergo many stealing attempts and food pellets losses, and then provide the food for the group, reflected by the highest value of the food provisioning index observed within the group. These animals are inefficient in stealing food from the others, and then have to adapt by increasing their number of carryings . The AC rats exhibit a different behavior: they experience less stealing attempts and lose smaller quantities of food compared to the SC animals. They also initiate a smaller number of attacks. So, the Autonomous Carriers possess some abilities which allow them to carry some food and to protect it against the congeners. The Non-Carriers are the rats which perform most of the stealing attempts and most of the thefts, and undergo few stealing attempts and food losses. They exhibit a high level of social interactions because the only way they have to get food is to steal it from conspecifics. Finally, a multivariate analysis of the behavioral data using a principal component solution with orthogonal rotation (Varimax) of the factor matrix suggests that, faced to the vital problem of feeding, the NC, AC and SC rats develop different strategies based on their own abilities in carrying

Variables	K-W	Non-Carrier rats	Autonomous Carrier rats	Supplier Carrier rats
Proportions	-	15/36 (42%)	9/36 (25%)	12 (33%)
Body weight (g)	n.s.	428 (403-457)	421 (396-463)	417 (396-455)
Total time of food possession (s)	n.s.	1108 (825-1608)	1394 (1172-1577)	1058 (688-1537)
Number of food possession periods	**	4 (3-6.5)	5 (3-5.5)	8 a, b (6-9)
Number of dives	-	-	2 (1-2.5)	6 c (5-8.5)
Number of carryings	-	-	2 (1-2.5)	6 c (5-8)
Diving efficiency (%)	-	-	100 (92-100)	100 (94-100)
Number of food possession periods ending by the ingestion of the pellet	n.s.	3 (2.5-5)	3 (2.5-4)	3 2-3)
Number of pickings of food from the floor	n.s.	0 (0-1)	0 (0-0)	0 (0-0)
Number of losses during an agonistic encounter	n.s.	0 (0-0)	0 (0-0)	0 (0-0)
Number of thefts committed	**	4 (3-5.5)	2 a (1-4.5)	1 a (0-2.5)
Feeding time obtained by theft (s)	**	1108 (825-1608)	947 (499-1154)	266 a, b (0-439)
Number of thefts experienced	**	1 (0-2)	1 (0-1.5)	5 a, b (3.5-6.5)
Feeding time lost by theft (s)	**	30 (0-374)	242 (0-599)	1395 a, b (1239-2261)
Food provisioning index (%)	**	1.5 (0-30.3)	11.2 (0-32.6)	59.3 a, b (54.9-77.4)
Number of attacks committed	**	48 (23-71)	6 a (3-10)	16 a, b (7-40)
Number of attacks experienced	**	6 (2-18)	4 (1-12)	20 a, b (11-37)

Table 1. Behavioral characteristics of NC, AC, and SC rats issued from a study including 6 groups of 6 Wistar male rats each. Results are expressed as median and quartiles (in brackets). K-W: the column indicates the results of the Kruskal-Wallis test. *a,b* $p < 0.05$, statistically significant differences from the Non-Carriers or the Autonomous Carrier rats, respectively (modified Mann-Whitney U-test for multiple comparisons). *c* $p < 0.05$, statistically significant difference between Autonomous and Supplier Carriers (Mann-Whitney *U*-test).

food or not, and to protect it or not against the congeners (Table 2). The diving-for-food situation seems to lead to an individual specialization: each rat being likely to adopt a Carrier or Non-Carrier profile, the NC, AC and SC statuses can be considered as supraindividual characteristics which can be assimilated to social roles, influencing the group which in turn influences the expression of these social roles [8].

4 The social differentiation

The emergence of behavioral differentiation in C and NC rats is a characteristic of all the groups tested in the diving-for-food situation until now. It regularly happens with respective proportions of Carriers and Non-Carriers that are very close to 50% [2,7,8,12,15]. Within the Carriers, about 40% of these animals become Autonomous. The differentiation remains stable over the time (6 groups were observed during 8 months without any changes in the social organization of the group [2]), and it has been observed in both mice [9] and rats, including Long-Evans, Sprague-Dawley and Wistar rat strains [2,7,12]. The emergence of the social differentiation is strongly related to the social context of this situation as demonstrated by the following results: (a) when they are individually trained in the device, about 100% of the rats are able to dive and to get food, whereas only 50% become Carriers in the 6 rat groups; (b) forming new groups included 6 previously differentiated Carrier or Non-Carrier rats leads to a new differentiation in both cases; (c) incentive behaviors from Non-Carriers towards Carriers are often observed, especially in the tunnel. The diving-for-food paradigm may then represent a suitable model to study social phenomena. At present, the neurobiological substrates involved in the social organization of the group remain to be investigated. Our hypothesis is that the differentiation between Carriers and Non-Carriers might be related to the anxiety level of rats faced with two anxiogenic constraints relative to the possession of food: a physical one (swimming in apnoea across the pool) and a social one (competition for food with conspecifics). Concerning the physical constraint, pharmacological data showed that the acquisition of the Carrier profile seems to be associated with a reduction of the level of anxiety in these animals (80% of rats which have been chronically treated with diazepam (DZ) at the anxiolytic dose of 1 mg/kg/day when the aquarium was progressively immersed became Carriers [6]). Concerning the social one, the chronic administration of DZ (0.5 mg/kg/day, i.p., 5 days) to all members of well-differentiated groups induced a significant reduction of the quantity of food lost by theft in Carriers whereas the number of thefts experienced and the number of attacks experienced were unchanged. These results suggest that DZ-treated animals are more efficient in protecting their food, possibly through the reduction of the level of anxiety of these animals faced to the social context [11]. Very recent investigations of the level of anxiety of the three behavioral profiles in the elevated plus maze test and the conditioned defensive burying paradigm have also confirmed this hypothesis, the AC rats being less anxious than the SC and the NC animals in both situations (unpublished data). All these results tend to confirm the role for anxiety in the emergence of the social differentiation.

ANXIETY →

ANXIETY ↓

	Carry some food	Don't carry some food
Protect the food pellet	AC rat	NC rat
Don't protect the food pellet	SC rat	unadapted animal

Table 2. Relationship between the acquisition of the NC, AC or SC social roles, the abilities of rats to carry some food and/or to protect it from the others, and the individual traits related to anxiety.

As suggested in Table 2, the adaptation of the NC, AC or SC rats to the situation depends on their own abilities to carry and protect the food from the others, and is probably the outcome of individual differences related to anxiety. Faced to the diving-for-food situation, some rats adopt the AC profile because they are the less anxious within the group, and then are able to overcome the two anxiogenic environmental constraints. Because they present a higher level of anxiety, the other rats develop different strategies to adapt to the situation through the acquisition of the NC or the SC statuses. The diving-for-food situation seems to be relevant in order to investigate the neurobiological substrates of anxiety related to the social adaptation of a group faced to an environmental constraint, and to study the pharmacological properties of various psychoactive substances.

References

1. Brockmann, H.J.; Barnard, C.J. (1979). Kleptoparasitism in birds. *Animal Behaviour*, **27**, 487-514.
2. Colin, C.; Desor, D. (1986). Behavioural differentiation in groups of rats subjected to a difficulty of access to the food. *Behavioural Processes*, **13**, 85-100.
3. Faulkes, C.G.; Bennett, N.C. (2001). Family values: group dynamics and social control of reproduction in African mole-rats. *Trends in Ecology and Evolution*, **16**, 184-190.
4. Gandolfi, G.; Parisi, V. (1973). Ethological aspects of predation by rats, *Rattus norvegicus* (Berkenhout), on bivalves *Unio pictorum* L. and *Cerastoderma lamarcki* (Reeve). *Bolletin di Zoologia*, **40**, 69-74.
5. Giraldeau, L.A.; Beauchamp, G. (1999). Food exploitation: searching for the optimal joining policy. *Trends in Ecology and Evolution*, **14**, 102-106.
6. Grasmuck, V.; Schroeder, H.; Toniolo, A.M.; Desor, D. (2000). Anxiety and social roles in the diving-for-food situation: are GABAergic and/or serotoninergic systems involved? *European Journal of Neuroscience*, **12** (suppl. 11), 91.
7. Helder, R.; Desor, D.; Toniolo, A.M. (1995). Potential stock differences in the social behavior of rats in a situation of restricted access to food. *Behavior Genetics*, **25**, 483-487.
8. Krafft, B.; Colin, C.; Peignot, P. (1994). Diving-for-food: a new model to assess social roles in a group of laboratory rats. *Ethology*, **96**, 11-23.
9. Nejdi, A.; Guastavino, J.M.; Lalonde, R.; Desor, D.; Krafft, B. (1996). Behavioral differentiation of mice exposed to a water tank social interaction test. *Behavioral Processes*, **36**, 11-18.
10. Nieder, L.; Cagnin, M.; Parisi, V. (1982). Burrowing and feeding behaviour in the rat. *Animal Behaviour*, **30**, 837- 844.
11. Schroeder, H.; Gelhaye, M.; Desor, D. (2003). Effects of a sub-chronic diazepam exposure on social behaviour related to food in rat groups exposed to the diving-for-food situation. *Acta Neurobiologiae Experimentalis*, **63** (suppl.), 49.
12. Schroeder, H.; Toniolo, A.M.; Nehlig, A.; Desor, D. (1998). Long-term effects of early diazepam exposure on social differentiation in adult male rats subjected to the diving-for-food situation. *Behavioral Neuroscience*, **112**, 1209-1217.
13. Sloman, K.A.; Taylor, A.C.; Metcalfe, N.B.; Gilmour, K.M. (2001). Effects of an environmental perturbation on the social behaviour and physiological function of brown trout. *Animal Behaviour*, **61**, 325-333.
14. Spinks, A.C.; Jarvis, J.U.M.; Bennett, N.C. (2000). Comparative patterns of philopatry and dispersal in two common mole-rat populations: implications for the evolution of mole-rat sociality. *Journal of Animal Ecology*, **69**, 224-234.
15. Thullier, F.; Desor, D.; Mos, J.; Krafft, B. (1992). Effect of group size on social organization in rats with restricted access to food. *Physiology & Behavior*, **52**, 17-20.
16. Whishaw, I.Q.; Whishaw, G.E. (1996). Conspecific aggression influences food carrying: studies on a wild population of *Rattus norvegicus*. *Aggressive Behavior*, **22**, 47-66.

Spatiotemporal sequences of neuronal activity encode odour information in the olfactory bulb

A.U. Nicol, A. Segonds-Pichon, K.M. Kendrick & *M.S. Magnusson
Babraham Institute, Cambridge CB2 4AT, UK
**Human Behaviour Laboratory, University of Iceland, IS-101 Reykjavik*

Abstract
Spontaneously recurring complex sequences of neuronal activity were recently demonstrated in action potential (spike) generation amongst simultaneously sampled multiple neurons and in postsynaptic potentials (Ikegaya et al, *Science*, 2004). Using software (*Theme*, Noldus) developed for detecting non-randomly recurring real-time sequential patterns (t-patterns) in behaviour and interactions (Magnusson, *Behav Res Meth*, 2000) we have inquired whether such patterns exist in the olfactory bulb (OB) and if so whether they have functional significance. In accordance with the Animals (Scientific Procedures) Act 1986 (UK), an electrode array (5x6 or 6x8 electrodes, 350µ or 250µ separation respectively, sampling area 2.2mm^2) was positioned in the OB (mitral layer) in 8 anaesthetised rats. Spikes were sampled in equivalent periods preceding and during 10s odour presentations (amyl acetate, N-butanol, DL-camphor, cineole) at various concentrations. Breathing was recorded as markers of peak inhale and peak exhale. Long t-patterns were identified. Often these patterns spanned the entire array. A considerable proportion incorporated breathing, and these patterns are the focus here. The incidence and length of breathing-related sequences varied with ventilatory phase and odour quality. Thus, in the OB, precise sequences of neuronal discharge encode olfactory information.

Introduction
Temporal sequencing is of great theoretical importance in cognitive neuroscience. Synchronised activity amongst neurons has been linked to perceptual cognition, namely “the binding problem” (Engel et al, 1997), whereby the combined features of a complex stimulus are associated by the synchronisation of the activities of neurons responsive to one or more of those features. The most widely accepted theory of the physiology of memory formation (Hebb, 1949) is also based upon the occurrence of such interactions in memory systems – “When an axon of cell A is near enough to excite B and repeatedly or persistently takes part in firing it… A’s efficiency, as one of the cells firing B, is increased”. Certain experimental models provide evidence that memory formation may adhere to hebbian principles. Long-term potentiation (LTP) is a physiological process that has been studied extensively since it was first described by Bliss & Lomo (1973). In LTP, pre- and postsynaptic elements in a neural pathway are simultaneously activated by repeated electrical stimulation of the presynaptic elements, thereby fulfilling the first of Hebb’s principles – neuron A repeatedly activates neuron B through the synaptic connection between the two elements. Subsequently, the efficiency of A in firing B is increased, and so is formed a simple hebbian assembly. In this paradigm, a large number of neuron A’s activate a large number of neuron B’s (i.e. there is little noise in the system), and the enhanced efficiency of transmission from one to the other manifests in the increased amplitude of the field potential generated when a single pulse is delivered to the presynaptic elements. However, evidence for such a process in a functioning memory system remains elusive.

Often comparisons between the experimental model (LTP) and memory formation in the functioning system assume that activation of a postsynaptic neuron may be reliably predicted by activation of an afferent presynaptic neuron. However, in a system where a postsynaptic neuron may receive many inputs from many afferent neurons, activation by the input from a single afferent neuron may have little impact on the activity of the postsynaptic neuron relative to ongoing activation by the large number of other presynaptic neurons. For this reason, t-pattern analysis of temporal sequencing in the activities of simultaneously sampled multiple neurons may be invaluable in understanding the mechanisms whereby neuronal networks encode sensory information.

Olfactory encoding is of specific interest as behavioural paradigms underpinning studies of the neurobiology of olfactory learning and memory are considered particularly robust (Bolhuis & MacPhail, 2001) and considerable progress has been made in establishing the neural substrates and pathways involved (Kendrick et al., 1992, 1997; Da Costa et al., 1997). Much of the encoding takes place at the level of the olfactory bulb (OB), the primary cortical projection area for olfactory input and an area that is entirely committed to processing this information. The area has been confirmed as playing an important role in olfactory memory formation. Thus, understanding the processes involved in encoding olfactory information is of great importance to understanding the neuronal mechanisms of learning and memory. Olfactory receptor neurons in the olfactory epithelium in the nasal cavity project mitral cells in glomeruli in the OB (Mori et al., 1999). Optical imaging studies demonstrate that different odorants elicit spatially defined patterns of glomerular activity in the olfactory bulb (Johnson et al., 2002; Spors & Grinwald, 2002). The quality of an olfactory stimulus is encoded by the specific combination of glomeruli activated by a given odorant. Gaining access to the olfactory bulb with a multiple electrode array (MEA) allows in vivo electrophysiological monitoring of neurons over a relatively large area of cortex. This makes possible the study of spatiotemporal patterns of activation across, and interactions between large numbers of simultaneously sampled and widely dispersed neurons in this area.

Methods
The present data have been collected from the olfactory bulb of anaesthetised rats (25% urethane, intraperitoneal, 1.5g per kg body weight). Using an MEA of either 30 or 48 electrodes advanced laterally into the OB, action potentials (spikes) were sampled from mitral layer OB neurons across an area of approximately 2.2mm^2 using a 100 channel laboratory interface (Cyberkinetics

Inc., USA). Typically, spikes were sampled from approximately 60 – 70% of electrodes. After completion of recordings, offline discrimination of spikes from individual neurons was performed using a PCA (principle components analysis) sorting algorithm developed specifically for these data allowing discrimination of activity from multiple neurons at each active electrode (Horton PM, http://www.sussex.ac.uk/Users/pmh20). Typically, spikes were sampled simultaneously from ≥100

Results

Temporal sequences often extending to 17 elements (neurons) were identified. These often spanned both dimensions of the array.

A considerable proportion of the patterns detected incorporated breathing. These patterns are the focus here. The number of breathing-related sequences detected in the data varied significantly with breathing phase, and also with odour presentation. The sequences most influenced

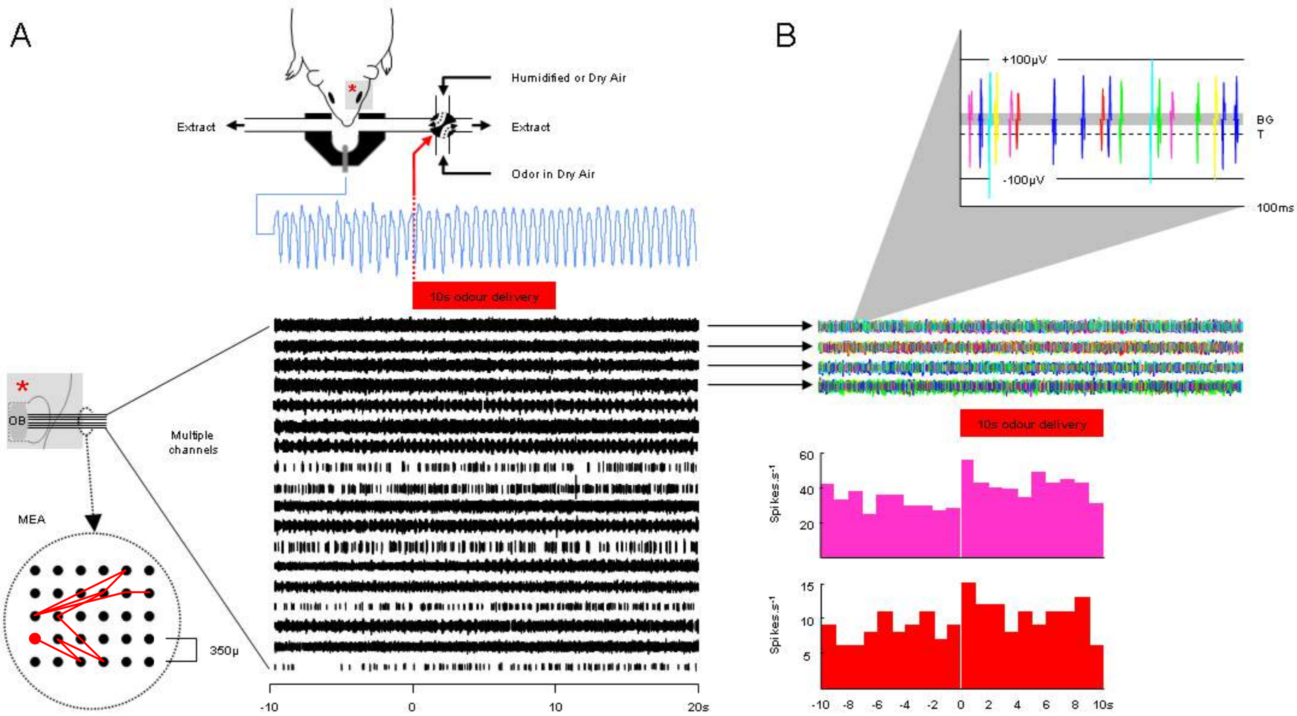


Figure 1. *Recordings from OB during odour presentations. (A) Neuronal activity (action potentials, shown, and local field potentials) was sampled from electrodes in a microelectrode array (MEA) positioned in the olfactory bulb (OB). Microelectrode arrays comprised either 30 electrodes (6×5, 350µm separation, illustrated) or 48 electrodes (6×8, 250µm separation). An example 2-dimensional plot of a sequence of neuronal discharges (red line) is superimposed on the diagram of a 30 electrode MEA, commencing at the electrode denoted by the red circle. Craniotomy and removal of the left eye permitted lateral access to the OB through the left orbit (inset*). During surgery and recordings, humidified air was supplied to the rat through a mask over the nose. A thermistor was used to record air temperature in the mask and so monitor breathing. Shortly (~30s) before delivery of an odour stimulus, the air to the mask was switched to dry air. Odours (amyl acetate, N-butanol, DL-camphor, cineole @ $5.42{\times}10^{-8}$ – $5.42{\times}10^{-6}$M) carried as saturated vapours in nitrogen gas (odourless), were mixed with dry air, and delivered for 10s to the rat via the mask. Onset of odour delivery was precisely timed to mid-expiration. Neuronal activity was recorded for a period spanning 10s before odour onset to 10s after odour offset. In the recordings shown, spikes were detected in 18 of 30 channels. (B) Spikes were detected when a triggering threshold (T) was crossed by the recorded signal from each electrode. This threshold was set at ≥2 × the background noise level (BG). A 100ms section of the upper spike train has been expanded to show times of occurrence of discriminated spikes. Two of the neurons, identified by the solid black and solid grey spikes, responded to the presentation of $5.4{\times}10^{-4}$ amyl acetate. The responses of these neurons to the presentation of this odour are represented by the purple and red histograms respectively, representing the firing rates of the two neurons over a period spanning 10s before to 10s after stimulus onset.*

neurons across the array. Times of occurrence of spikes generated by individual neurons were stored as events coded with the identity of the neuron and its location on the MEA. Also stored were events marking onset of expiration and onset of inspiration in the breathing cycle to allow this data to be related to patterns identified in the neural data. The data stored in this way are suitable for analysis by *Theme* (Noldus NL). Separate data files were generated for activity preceding and during presentation of odour stimuli to the rat over a number of trials using different odours at various concentrations. Event types entered into t-pattern analyses were times of occurrence of spikes from individual neurons, and times of onset of inhalation and exhalation in the breathing cycle.

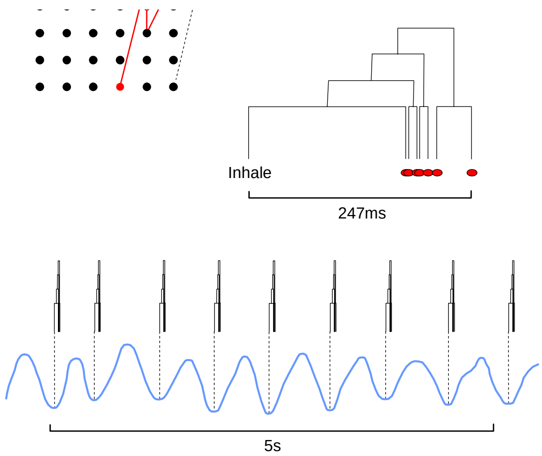


Figure 2*. Breathing-related sequence in neuronal firing. The neuronal sequence displayed on the enlarged grid (red line, commencing at the electrode indicated by the red circle) incorporated the marker for onset of inhalation. The mean duration of the spike sequence was 75ms, and the mean delay between the onset of inhalation and the onset of the spike sequence was 172ms.*

by odour presentation were those associated with the onset of inhalation or with both onset of inhalation and onset of exhalation (see Fig 3a). The incidence of these sequences increased on odour presentation. Sequences that were associated with the onset of exhalation were unaffected by odour presentation. The number of breathing related sequences also increased with odour concentration. Each of the experimental variables that were found to influence the number of breathing-related sequences also influenced the maximum length of patterns detected in each data set (see Fig 3b & 3c).

The findings were compared to those obtained when the activity of each neuron was rotated by random amounts relative to the activity of each other neuron so that the interval structure of each was maintained. In randomised data, only shorter sequences were found, and these were many times less frequent than in the real data.

Discussion

Precise sequences in spontaneous cortical neuronal activity were recently identified both in an isolated tissue preparation and *in vivo* (Ikegaya et al, 2004). These patterns were found both in intracellularly recorded postsynaptic potentials, reflecting release of discrete quantities of neurotransmitter, and in extracellularly recorded neuronal action potentials (spikes). Here we show that such sequences occur across large areas of the two-dimensional network of mitral neurons in the olfactory bulb, with functional connections spanning in some cases the entire area sampled by the MEA (>2mm^2). This perhaps is less remarkable given a recent account of anatomical connections ('short axons') spanning many mitral cells across the olfactory bulb (Aungst et al, 2003). Most notable in our findings is that these sequences, whilst present in spontaneous neuronal activity, have functional relevance. Presentation of an odour stimulus increases both the variety of sequences detected, and the number of patterns generated amongst the neurons. These increases are selective for sequences of neuronal firing that are associated with the onset of inhalation. Included amongst these sequences are those associated with both onset of inhalation and onset of exahalation, but not those associated exclusively with onset of exhalation. Thus stimulus quality is represented in the richness of neuronal sequencing in mitral cell activity.

These findings are consistent with other recent observations of neuronal activation in phase with breathing using *in vivo* optical imaging techniques (Spors & Grinvald, 2002) or electrophysiological techniques (Margrie & Schaefer, 2002), the latter study demonstrating that mitral cell membrane membrane potential fluctuations, and therefore likelihood of discharge, occur in phase with ventilatory rhythm. However, here we have shown that neuronal activity involves precise sequences of discharge that are related both to ventilatory activity and to odour information.

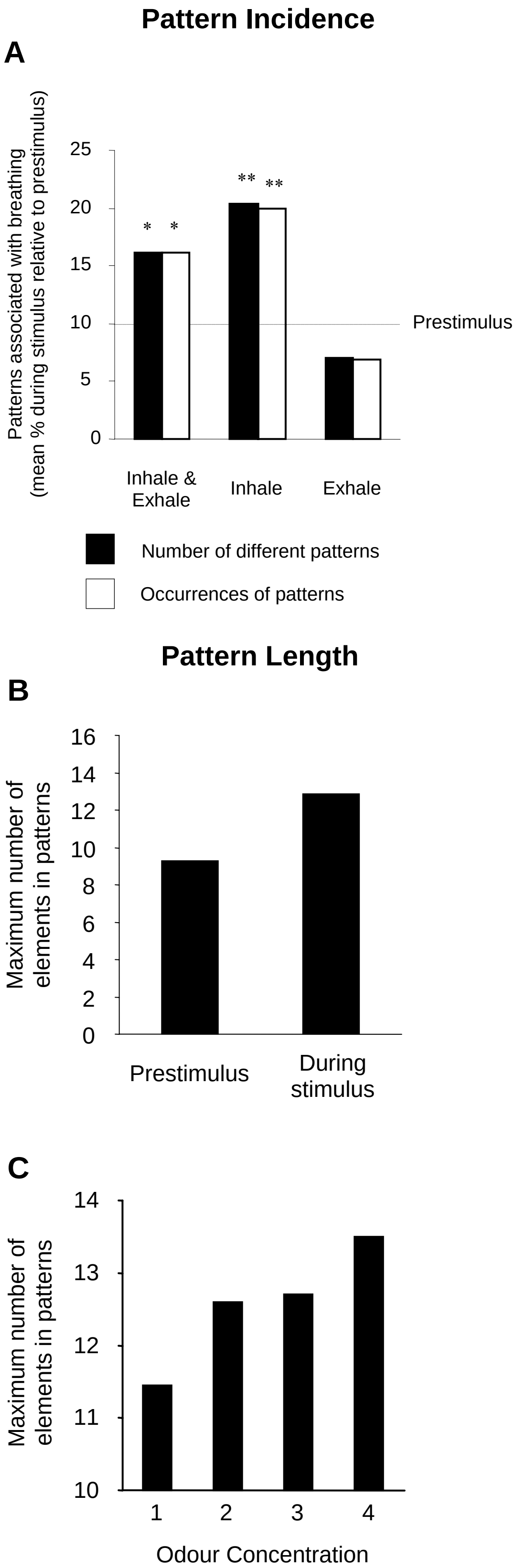


***Figure 3**. Experimental variables influencing the incidence and complexity of breathing-related sequences. (A) The number of sequences that incorporated the onset of inhalation increased on odour presentation. For the sequences that incorporated the onset of exhalation, but not the onset of inhalation, this was not so. (B) The complexity of breathing related sequences (i.e. the number of neurons participating in a sequence) also reflected odour onset. The maximum detected sequence length was greater during odour presentation than in prestimulus activity. (C) The maximum sequence length also increased as the concentration of the presented odour increased.*

References:

1. Aungst JL, Heyward PM, Puche AC, Karnup SV, Hayar A, Szabo G & Shipley MT (2003) Centre-surround inhibition among olfactory bulb glomeruli. *Nature*, **426**, 623-629.
2. Bolhuis JJ & MacPhail EM (2001) A critique of the neuroecology of learning & memory. *Trends Cog Neurosci*, **5**, 426-433
3. Da Costa APC, Broad KD & Kendrick KM (1997) Olfactory memory and maternal behaviour-induced changes in c-fos and zif/268 mRNA expression in the sheep brain. *Mol Brain Res*, **46**, 53-76.
4. Engel AK, Roelfsema PR, Fries P, Brecht M, Singer W (1997) Binding and response selection in the temporal domain - a new paradigm for neurobiological research. *Theory Biosci*, **116**, 241-266.
5. Griffith JS & Horn G (1963) Functional coupling between cells in the visual cortex of the unrestrained cat. *Nature*, **199**, 893-895.
6. Hebb DO (1949) *The Organisation of Behaviour*. Wiley, New York.
7. Ikegaya I, Aaron G, Cossart R, Aronov D, Lampl I, Ferster D & Yuste R (2004) Synfire chains and cortical songs: temporal modules of cortical activity. *Science*, **304**, 559-564.
8. Johnson BA, Ho SL, Yihan JS, Yip S, Hingco EE & Leon M (2002) Functional mapping of the rat olfactory bulb using diverse odorants reveals modular responses to functional groups and hydrocarbon structural features. *J Comp Neurol*, **449**, 180-194.
9. Kendrick KM, Guevara-Guzman R, Zorrilla J, Hinton MR, Broad KD, Mimmack M & Ohkura S (1997) Formation of olfactory memories mediated by nitric oxide. *Nature*, **388**, 670-674.
10. Kendrick KM, Levy F & Keverne EB (1992) Changes in the sensory processing of olfactory signals induced by birth in sheep. *Science*, **256**, 883-886.
11. Magnusson MS (2000) Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments and Computers,* **32**, 93-110.
12. Magnusson MS (2004) Repeated Patterns in Behavior and Other Biological Phenomena. In D. Kimbrough Oller & Ulrike Griebel (Eds) *Evolution of Communication Systems : A Comparative Approach (Vienna Series in Theoretical Biology)*. The MIT Press.
13. Magnusson MS, Burfield I, Loijens L, Grieco F, Jonsson GK & Spink A (2004) *Theme: Powerful Tool for Detection and Analysis of Hidden Patterns in Behavior*. *Reference Manual, Version 5.0*. Noldus Information Technology bv, Wageningen, The Netherlands.
14. Margrie TW & Schaefer AT (2002) Theta oscillation coupled spike latencies yield computational vigour in a mammalian sensory system. *J Physiol*, **546**, 363-374.
15. Mori K, Nagau H & Yoshihara Y (1999) The olfactory bulb: coding and processing of odor molecule information. *Science*, **286**, 711-715.
16. Spors H & Grinvald A (2002) Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. *Neuron*, **34**, 301-315.

Analyzing the flight behavior of malaria mosquitoes: a challenge in 3 dimensions

J. Spitzen, W. Takken

Laboratory of Entomology, Wageningen University, Wageningen, The Netherlands

Abstract

Single nocturnal malaria mosquitoes were exposed to human odors with or without a heat source in a wind tunnel. Their flight tracks were recorded with two solid state IR sensitive cameras, mounted under 15 degree angles above the wind tunnel. By subsequently combining the information from the two different camera views, the three-dimensional position of the mosquito was reconstructed and could be analyzed. Flight parameters for individual mosquitoes were compared for different zones in the wind tunnel viz. inside- or outside the odor plume. In this way we reconstructed the mosquito flight paths within and outside an odor plume, and studied the difference between the effect of odor alone and odor + heat. The effects of these combined host stimuli upon the behavior of these important mosquitoes will be discussed.

Keywords

3D image analysis, flight tracking, *Anopheles gambiae* sensu stricto, insect behavior, olfaction

1 Introduction

Malaria is still one of the scourges of the Third world, causing millions of infected persons and 1-3 million deaths each year [1]. The disease is transmitted by anopheline mosquitoes, which require blood for egg production. It has been proposed that interruption of mosquito-host contact would be the most effective method of malaria prevention as parasite transmission will then cease. Mosquitoes find their hosts by olfaction. The smell of the host can vary, and this determines the efficiency with which the mosquito can locate its host [2]. We are interested to identify the chemical cues that constitute the attractive odors and how the mosquito navigates upwind while stimulated by these cues. While flying upwind towards a host, mosquitoes use horizontal and vertical movements in odor plumes (pers. obs.). Therefore a wind tunnel setup was designed which could record flight in three dimensions. This paper presents a selection of data that demonstrate how 3D image recording of *An. gambiae* females is a useful tool to describe and understand a variety of flight parameters during host seeking.

2 Materials and Methods

2.1 Mosquitoes

The colony of *Anopheles gambiae* Giles sensu stricto originated from Suakoko, Liberia and was reared at Wageningen University, The Netherlands as described in [3]. One day before the experiments 5-8 day old female mosquitoes, which had not received a blood meal, were selected from their cages and placed in 2x3 cm plastic cups individually. Damp cotton wool was placed on top of the cup to prevent dehydration of the mosquitoes. Experiments took place around the 10th hour of the dark period, the time they are normally active in their natural environment.

2.2 Wind tunnel

A wind tunnel of 1.8x0.6x0.6m provided the flight arena for the mosquitoes (Fig. 1). The floor and walls were covered with infrared absorbing black coating, the ceiling was made of transparent polyacrylate. The upwind and downwind sides were closed off with messing gauze. A conditioned laminar air stream of 26°C and approx. 60%RH went through the tunnel at 20cm per second. Immediately in front of the upwind side of the metal gauze screen a glass funnel was placed horizontally, the opening (5cm diameter) facing the centre of the screen.

The flight images were taped synchronically by computer-controlled video recorders using two frame grabbers. After completion of a series of trials, the video recordings for each camera were played back. Within each image frame, the two-dimensional position of the mosquito was determined. By subsequently combining the information from the two different camera views, the three-dimensional position of the mosquito was reconstructed and stored. Special measurement and analysis software based on a 2-dimensional version of EthoVision, Noldus Information Technology, Wageningen, the Netherlands, was developed to automate the analysis process, reconstruct the flight path in three dimensions, as well as to extract behavioral parameters for characterization of the flight paths. Additionally, flight parameters for individual mosquitoes were compared for different zones in the wind tunnel viz. inside- or outside the odor plume. The estimated dimensions of the odor plume were assigned to the EthoVision program after a Ethylenediamine 1: acetic acid 4 smoke plume was created from the funnel (Cone shape plume with top angle of 19 degrees, diameter of 5cm at t = 0). A 5 centimeter wide boundary zone was assigned to the estimated ‘edges’ of the plume.

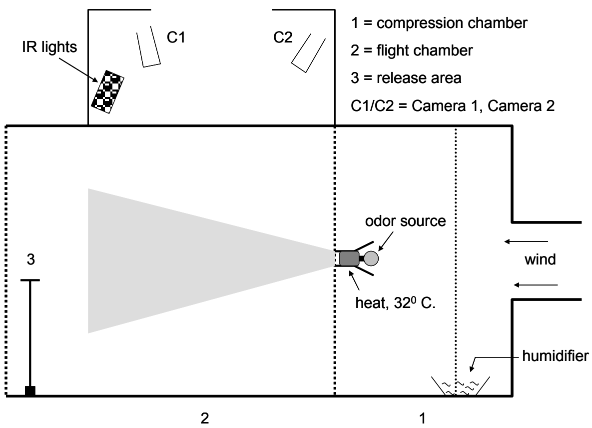


***Figure 1**. Side-view of wind tunnel setup. The estimated plume shape is indicated in light grey.*

2.3 Experimental procedure

Three different treatments were tested.

1. odor + heat
2. heat
3. odor

An odor sample was collected by a male volunteer wearing a nylon sock for 24 h. The sock was placed at the wide side of the funnel, immediately upwind of the heat source. The heat element was either set at 32^0C, which mimics human skin temperature, or turned off. Different treatments were tested on different days to avoid possible contamination of the setup. The following parameters were recorded and analyzed in order to test whether differences in host-seeking behavior could be determined by means of the 3D analysis techniques:

- Percentage landing on source (within 5 centimeter circle on upwind screen)
- Path straightness. Length flight chamber / ∑distance moved.
- Velocity when flying in different zones of interest
- Heading when flying in different zones of interest.

To reduce the large number of data points, only the last section of the recorded track was analyzed. This section is defined as the last time the mosquito flies downwind and/or subsequently upwind before landing at the upwind screen.

3 Results

A total of 144 individual mosquitoes were tested of which 105 (73%) took off for upwind flight. Figure 2 shows that the only treatment that repeatedly resulted in landing on the source was odor + heat (69%). Analysis of the first series of successfully recorded flights shows that when the mosquitoes were not attracted to the source, their path straightness increased (Fig. 3). In other words, they almost flew in a straight line to the screen. By contrast, mosquitoes that were exposed to odor + heat covered a greater distance before they landed on the screen. From the mosquitoes that landed on the odor + heat source flight parameters were analyzed for the three different zones of interest; edge, within and out of the plume. The mean velocity of female mosquitoes was reduced when flying in- or along the edge of the plume (Fig. 4). The heading was mostly parallel to the wind direction while flying in- or along the plume, but when the plume was lost mosquitoes tended to fly more cross wind (Fig. 5).

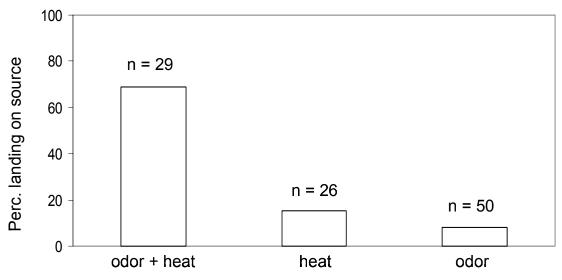


Figure 2. Percentage of female mosquitoes landing on the source when exposed to the treatments; odor + heat, heat and odor.

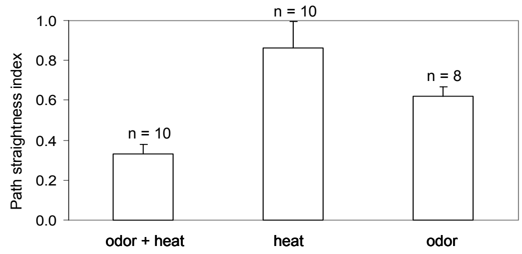


Figure 3. *Mean path straightness (Length flight chamber / ∑ distance moved) of female mosquitoes in flight while exposed to the treatments; odor + heat, heat and odor.*

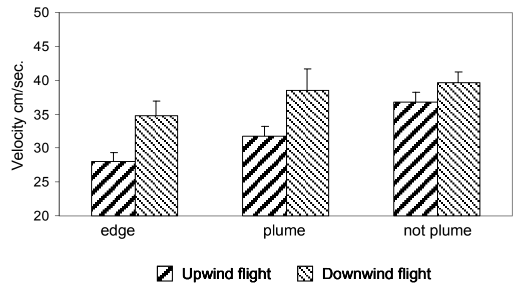


Figure 4. *Mean velocity per zone for female* An. gambiae *which were exposed to skin odor and heat. Means and SE are calculated for 10 individual tracks.*

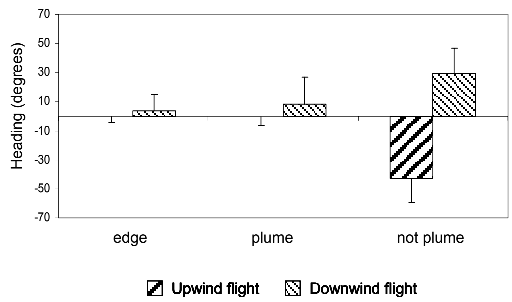


Figure 5. *Mean heading per zone for female An. gambiae which were exposed to skin odor and heat. A heading of 0 degrees represents a flight in line (either up-or downwind) with the wind direction. Means and SE are calculated for 10 individual tracks.*

4 Discussion

Female mosquitoes were attracted to a small artificial host stimulus in the present experimental setup. The combination of human body odor with a heat source proves to be essential to induce mosquito landing on the source [4] because heat or odor alone caused only weak landing responses. The study successfully produced a system that reconstructs the flight of a nocturnal insect in three dimensions and provides data for the analysis of flight behavior. The very fast movements of the mosquito, requiring a high temporal resolution of the observation, prevents at present an on-line analysis because of the specialized image detection algorithm used. The development of hardware to make real-time analysis possible is on its way. One of the best features of the program is the possibility to study the behavior of the flying insect in relation to the position of the odor plume. The program allows for the study of flights in the boundary layer of the plume. A good description of odor plume structure becomes essential in order to describe behavioral changes of mosquitoes that lose contact with the plume.

The results provide promise that behavioral responses of malaria mosquitoes to natural hosts and odor-baited entry traps can be better understood.

We thank Piet Huisman for assistance with the initial experimental design and Leo Koopman, Frans van Aggelen and André Gidding for rearing of the mosquitoes. The continuous support by Noldus Information Technology is much appreciated.

References:

1. Greenwood, B.M.; Bojang, K.; Whitty, C.J.; Targett, G.A. (2005). Malaria. *Lancet*, 365, 1487-1498.
2. Takken, W.; Knols, B.G.J. (1999). Odor-mediated behavior of afrotropical malaria mosquitoes. *Annual Review of Entomology*, **44**, 131-157.
3. Smallegange, R. C.; Qiu, Y. T.; van Loon, J. J.; Takken, W. (2005). Synergism between ammonia, lactic acid and carboxylic acids as kairomones in the host-seeking behaviour of the malaria mosquito *Anopheles gambiae* sensu stricto (Diptera: Culicidae). *Chemical Senses* ***30***, 145-152.
4. Healy, T. P.; Copland M. J. W.. (2000). Human sweat and 2-oxopentanoic acid elicit a landing response from *Anopheles gambiae*. *Medical and Veterinary Entomology* , **14**, 195-200.

The outflyer rate as a usable field method for verifying defense strategies in Western honeybees

Gerald Kastberger[1], Ronald Thenius[1] & Randall Hepburn[2]
[1] *Department of Zoology, University of Graz, A 8010 Graz, Austria*
[2] *Department of Zoology, Rhodes University, Grahamstown, South Africa*

Abstract

We investigated the responsiveness of 11 honeybee colonies of three African and European subspecies (*Apis mellifera capensis, scutellata, carnica*) to the exposition of normative mechanical shocks and alarm pheromones. We assessed foraging and defense status by tracking outflyer rates of disturbed colonies in relation to their undisturbed state. On mechanical stimulation the colonies resolved into two defense response types of considerable constancy; *releaser* colonies increased the rate of outflyers and *retreater* colonies reduced them. Exposure to alarm pheromones alone evoked only weak responses. Under combined mechanical and pheromonal stimulation the colonies further increased those behaviors exhibited under mechanical stimulation. This field method is usable to determine colony-specific traits distinguishing *aggressive* from *docile* defense strategies.

Keywords

Defense behavior / docility / aggressiveness / guard bees / *Apis mellifera*

1 Introduction

An elaborate defense system has evolved in honeybees in tandem with a need to safeguard their nests against the onslaught of predators and parasites. Even so, there is considerable variation in *defensiveness* of colonies among and within honeybee species, respective defense strategies scale from *docile* to *aggressive* [1,2,3,4,5]. *Docile* strategies avoid exposure of the bees to the predator, not stinging, retreating into the nest and discontinuing flight. *Aggressive* strategies of honeybees extend from guarding behavior to repelling, recruiting, mobilizing and releasing some or even a mass of counterattacking guard bees which sting or even kill intruders.

Here we report about a field method which we adapted in order to test two facets of the defensiveness in *Apis mellifera* colonies. We proved the significance of anecdotal observations on genetic dispositions of colonies (that some colonies are more *docile*, while others are more *aggressive*), and the further-going question whether one and the same colony may utilize the whole spectrum from *docile* to *aggressive* (from full retreat to the nest to the release of flying guards). For that, we mimicked natural disturbances using mechanical and pheromonal stimuli, and assessed *defensiveness* under variable foraging levels mainly by the rate of outflyers.

2 Methods

Test colonies. We conducted experiments with three subspecies of *A. mellifera; A.m. scutellata* and *A.m. capensis* during spring in South Africa and with *A.m.carnica* in summer in Austria. We used three colonies each of *A.m. scutellata* (82 experiments), *A.m. capensis* (52 experiments), and five of *A.m. carnica* (68 experiments). Each colony consisted of about 10.000 bees selected as reportedly *aggressive*, *intermediate* or *docile*.

Stimulation. We used three regimes: (1) the *m*-regime referred to mechanical shocks which were delivered to the bees in the hive. For that, 200 g weights were dropped through a 50 cm tube mounted on the top of the hive to transmit the impulse directly to the top-bars of the frames. These shocks were set for 3 min, at a rate of one per 2 s; (2) the *p*-regime exposed the colonies to alarm pheromones which were placed on cotton buds, doped with 10 stings, 10 cm in front of the hive, so that the bees had to fly from the hive entrance to reach it; (3) the regime *mp* combined mechanical shocks and alarm pheromones. Each experiment comprised three phases: a *pre-stimulation* (*preS*) phase of 3 min reflected the undisturbed conditions; a *stimulation (S)* phase continued for again 3 min; followed by a *post-stimulation* (*postS*) phase of 10 min. The first two minutes of the *S*- and *postS*-phases were subdivided into 4 intervals each of 30 s duration (*S1-4*, *postS1-4*). Colonies were tested for 10 days at different temperatures and times of day between 9:00 and 17:00, and after each experiment a colony remained untouched for at least one hour. Midday displays where young bees perform orientation flights were not considered.

Data assessment. We video-recorded the hive entrance throughout the experiments and observed three groups of bees: (a) *Crawlers* represented ground traffic at the hive entrance and were counted in 30s intervals. (b) *Scanners* facing the hive entrance were detected either in straight or in hovering flight. Their numbers were counted for every 2s (in order to minimize multiple scoring of the same bee) of the video film and were grouped into 30s intervals. (c) *Outflyers* from the hive were counted continuously for 30s intervals. The outflyer rate in the *preS* phase provided the baseline foraging level.

Data analysis. The outflyer rate was assessed in the 16 min of experiment and related to the mean outflyer rate of the *preS* phase. The resulting *net outflyer rate* confers the responsiveness to stimuli in the following details: positive values in the stimulation and post-stimulation phases signify the release of *flying defenders* (*releasing* response); negative rates define a *retreating* response, zero values refer to nil response. The correlation between net outflyer rates and foraging was assessed for the three experimental (*preS-, S- postS-*) phases and tested by conventional t-test.

3 Results and discussion

Behavior classes. The *outflyer* rate provides useful measures for two main behavioural features of the colonies: in undisturbed conditions it quantifies foraging (exclusively midday displays), and as a response to threatening stimuli it expresses the defense state of the colony. The latter can be described by the *net outflyer rate* defining colony defensiveness in two directions: negative rates denote down-regulation of foraging, increases in outflyer rate are more complex and do not necessarily indicate the opposite trait, the up-regulation of foraging. Detailed analysis of video-taped behaviors allowed us to distinguish here two cases: (a) up-regulation of foraging is proved if the *outflyer* and *scanner* rate curves had similar time courses but were phase-lagged by 3 to 5 min on the *scanners* side. This response was typical of retreater colonies in the *m*-regime, when they returned to normal foraging after their stimulus-induced retreat (Fig.1b). (b) we prove a release of guard bees if, as an immediate response to *m*-stimulation, the *outflyer* rate peaked synchronously with the *scanner* rate (Fig.1a,b,c),. Under this condition, bees attacked the experimenters regularly. This response was typical for *releaser* colonies in the *mp*-regime.

Responses to disturbance. Colonies exposed to alarm pheromones but not to mechanical disturbances did not mobilize flying guards, which suggests that they evaluate this *p*-stimulus as non-alarming signal. Remarkably, this reaction was essentially the same in *scutellata*, *capensis* and *carnica* colonies. However, colonies threatened by combined *mp*-stimulation increased their basic reactivity to *m*-stimulation by a factor of two. This result supports the hypothesis that the initial exposure to alarm pheromones alone must have informed, but did not arouse, the colony. Colony defense responses such as *releasing* guard bees or *retreating* to the nest obviously requires more than the pure exposition to alarm pheromones it needs also stimulation in mechanical or visual modality.

Constancy of defense pattern. The defense patterning changed throughout the experimental week in colony-specific ways, however, independently of subspecies. For the *m*-regime we found three schemes: (a) Some colonies remained in *releasing* flying guards and intensified this behavior in the consecutive days (Fig.1a,d); (b) other colonies always *retreated* to the nest (Fig.1b,e). Ten out of eleven test colonies maintained their specific defense patterns irrespective of changing environmental factors such as time of day or ambient temperature. (c) We found only one exception which *released* flying guards on the first experimental day, but switched to *retreater* type on the second day of testing (Fig.1c,f).

Foraging tunes defensiveness. Defensiveness is known to be influenced by the foraging activity of the colony. To investigate this we subjected the test colonies of a given foraging state (which was dependent, in particular, on time of day and ambient temperature), as described above, under three stimulation regimes (*m, p, mp*). This allowed us to classify them in respect of two sets of defense strategies.

Taking the *net outflyer rate* under the *m*-regime as the measure of *defensiveness*, the response patterns of the test colonies followed a rule which is probably universal for *Apis mellifera*. The eleven colonies which we had originally selected in three subspecies as *low*, *intermediate* and *high* according to an initial subjective assessment of their general *aggressiveness*, diverged into only two distinct categories, which we term (*m*-) *releasers* and (*m*-) *retreaters*. In categorizing them we identified as the main parameter the offset value between the stimulus-invoked portion of the *net outflyer rate* and the foraging level, taken in the first minute of *m*-stimulation. It appears highly probable that this value reflects genetic differences rather than environmental effects. The fact that *releaser* colonies mostly remain *aggressive* and *retreater* colonies *docile*, irrespective of ambient temperature, time of day or individual history, provides additional support for this interpretation.

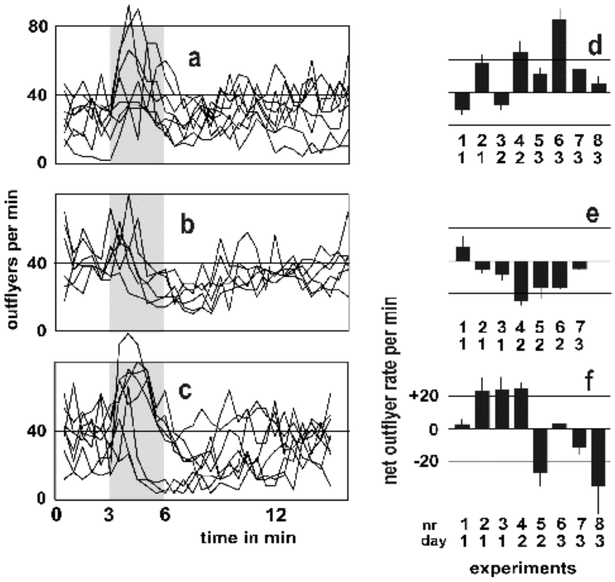


Figure 1. *The aspect of constancy of defense type in Apis mellifera, illustrated by three examples under m- stimulation: a,d colony capensis 3 as releaser type; b,e colony capensis 2 as retreater type; c,f colony carnica 3 as switching type from releaser to retreater. Graphs a-c: abscissa, time in min; ordinate, outflyer rate per min; the graphs of consecutive experiments have been superimposed; grey shaded zones mark the stimulation phase. Diagrams d-f: ordinate, net outflyer rates in the S3 phase, with arithmetical means (black columns) and mean errors (vertical bars). We took the S3 phase because here the contrast between releaser and retreater colonies is maximally established. Sequence coding: nr, number of experiment; day, relative day at which the experiment was performed; experiments were carried out at different times of day and different ambient temperatures within three consecutive days.*

Supported by the Austrian Scientific Funds (Grant N. P 13210)

References:

1. Breed M.D.; Guzman-Novoa E.; Hunt G. J. (2004) Defensive behavior of honey bees: organization, genetics, and comparisons with other bees. *Ann. Rev. Ent.*, **49**, 271-298.
2. Collins A.M.; Rinderer T.E.; Tucker K.W.; Pesante D. (1987) Response to alarm pheromone by European and Africanized honeybees. *J. Apic. Res.*, **24**, 217-223.
3. Moritz R.F.A., Southwick E.E., Harbo J.R. (1987) Genetic analysis of defensive behaviour of honeybee colonies *Apis mellifera* L. in a field test. *Apidologie*, **18**, 27-42.
4. Schneider S.S., McNally L.C. (1992) Colony defence in the African honey bee in Africa Hymenoptera: Apidae, *Env. Ent.*, **21**, 1362-1370.
5. Winston M.L. (1987) The biology of honey bees. *Havard University Press*, Cambridge.

Automatic segmentation of mouse behavior during video tracking in home cages

A.B. Cherepov , T.V. Mukhina, K.V. Anokhin
Department of Systemogenesis, P.K. Anokhin Institute of Normal Physiology, Moscow, Russia

Abstract

A common approach to analysis of behavior recorded by video tracking systems employs parameters (such as animal speed, moved distance, latencies to enter certain zones, times spent there, etc.) which only indirectly reflect the organization of complex behavioral patterns and their dynamics. To overcome this limitation we developed a novel approach to quantitative segmentation of animal behavior by automatic dissection of behavioral continuum into meaningful behavioral units. This method includes search for both "major" and "minor" breakpoints in the movement continuum. The former are complete stops of an animal; the latter can be found by analysis of animal's acceleration, angular speed, etc. The developed approach was successfully applied for segmentation of mouse behavior in a novel environment.

Keywords

Video tracking, elementary behavioral acts, functional systems.

1 Existing approaches to behavior analysis

Standard parameters (animal speed, distance moved, latencies to enter certain zones, times spent there, etc.) that are included in the up-to-date automated video tracking and analysis systems [1,2,3] supply important information about animal behavior in various tests. However, they cannot supply many relevant characteristics of motivated goal-directed behavior, learning and memory. Recent attempts to estimate different behavioral strategies of animals (for example, strategies of search for a hidden platform in the Morris water maze test) also do not give detailed information about discrete behavioral events and their dynamics [3,4].

Another approach to the analysis of behavior is implemented in the event recording systems (e.g. [5]). It allows more accurate registration and classification of the behavioral events. However, such systems are based mainly on expert estimations, do not employ automatic registration of behavioral dynamics and usually provide information without its precise relation to animal position in space.

Therefore in the present work we started to develop a new approach for an objective and quantitative recognition of biological units of animal behavior. Its main task is automatic demarcation of individual behavioral acts within a behavioral continuum, using animal path trajectories obtained from various video tracking systems.

2 Functional systems approach to segmentation of behavioral continuum

Our approach is based on the functional systems theory (FST) [6]. According to the FST behavioral continuum consists of individual behavioral acts performed by discrete self-organized functional systems. Each behavioral functional system starts from an afferent synthesis and decision making and finishes with evaluation of an achieved result. Achievement of the result of a functional system is a notable event in a goal-directed behavior. These moments are expressed as "breakpoints" between neighboring elementary acts within a behavioral continuum. Algorithmic recognition of such breakpoints will be independent from subjective estimates and can be used to explore relations between a natural structure of behavior and systems processes in the brain.

3 Results

Analysis of trajectories obtained in different behavioral tests shows that achievement of the result of a behavioral act is usually accompanied by a discontinuity of the animal movement. This is typically expressed in stops of the animals or slowing down of their movement. Therefore, in the first approximation of the algorithm of behavior segmentation we searched for "major" breaks of the movement continuum marked by complete stops of the animal motion. To find such zero speed segments we smoothed the trajectories by the running median algorithm as described in [7].

***Figure 1.** Three-dimensional view of the smoothed speed (dashed black line) and a trajectory (solid black line) for one of the segments between two stops found by processing of trajectory of C57Bl/6 mouse in a rectangular cage (34x20 cm) using running median algorithm. The smallest number of points within stop segment is set to 5 frames. Gray-colored line represents previous part of the trajectory. ○ , ● - Start and end points of trajectory, respectively.*

Figure 1 represents an example of a behavioral segment between two "major" breaks found by processing of trajectory of C57Bl/6 mouse. However, analysis of video records of the animals' behavior within such large segments revealed that they can often contain several "minor" acts (such as turning, rearing, etc.). Decrease of the minimum number of frames within stops can not solve this problem, since such segments do not contain points with zero speed. Use of speed threshold to classify parts of the trajectory with low speed as breaks between the "minor" acts is also inappropriate, since some acts with higher speed are missed by such a criterion.

Further search for additional signs of breakpoints between individual behavioral acts, has shown that projection of the acceleration on the speed direction in each point of the trajectory can be used for this purpose. Figure 3 shows that such "additional" breaks are defined as episodes of deceleration followed by acceleration, while the speed should be low, as determined by the speed threshold (dashed line at 5 cm/s in Figure 2B). To find these "additional" breakpoints another algorithms, such as analysis of changes of speed vector direction between each pair of trajectory points, can be employed as well.

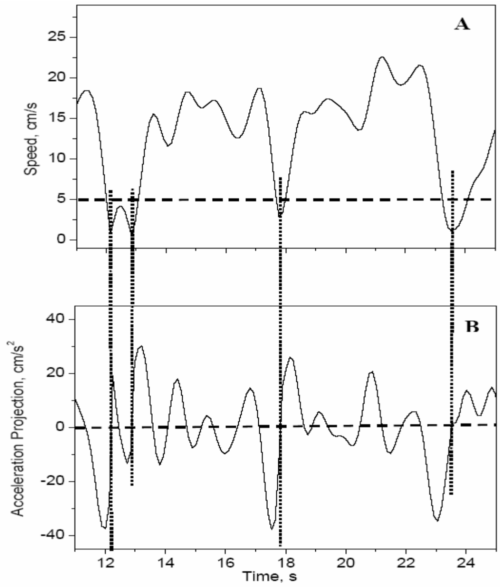


Figure 2. *Graphs of the instant speed (A) and projection of acceleration on speed direction (B) for a fragment of a segment between two stops. Dotted lines indicate breaks. Speed and acceleration values are obtained by 50 iteration of the second order polynomial smoothing (with 5-point sliding window) of the animal trajectory, what results in reduction of parasite noise and preservation of the curve features.*

In order to test the proposed algorithms we analyzed behavior of C57Bl/6 and BALB/C mice in a novel environment. Mouse activity was recorded in the home or novel cages using a FreeTrack video tracking software. Segmentation of behavior was performed with a SegmentAnalyser software that contains comprehensive options for the breakpoint detection and visualization [8].

Our tests revealed that accounting for the criterion of acceleration projection alone (without previous identification of stops) gives many spurious acts (even when option to merge neighboring breakpoints is turned on). The best segmentation was obtained by application of running median to find stops with subsequent analysis of acceleration projection for obtained segments to find points representing additional breaks and thus additional acts. Analysis of the video records of the segmented behavior demonstrated that the breakpoints found by this method show the best match to the segmentation performed by an expert observer. (about 85%). Mistakes were caused by inability of algorithms to account for low amplitude movements.

Supported by the Russian Foundation for Basic Research (Grant 03-04-49739) and International Science & Technology Center (Project 2368)

References:

1. Noldus, L.P.J.J.; Spink, A.J; Tegelenbosch, R.A.J. (2001). A versatile video tracking system for automation of behavior experiments. *Behavior Research Methods, Instruments & Computers*, **33**, 398-414.
2. Mukhina, T.V.; Bachurin, S.O.; Lermontova, N.N. (2001). Versatile computerized system for tracking and analysis of water maze test. *Behavior Research Methods, Instruments & Computers*, **33**, 371-380.
3. Wolfer, D.P. (2002). Exploiting video-tracking data to dissect water maze learning and strategy choice in mutant mice. *Int. Conf. on Methods and Techniques in Behavioral Research "Measuring Behavior 2002"*, Amsterdam, The Netherlands. 27-30 August, 2002.
4. Graziano, A.; Petrosini, L.; Bartoletti, A. (2003). Automatic recognition of explorative strategies in the Morris water maze. *J. Neurosci. Methods*, **130**, 33-44.
5. Noldus, L.P.J.J.; Trienes, R.J.H.; Hendriksen, A.H.M.; Jansen, H.; Jansen, R.G. (2000). The Observer Video-Pro: new software for the collection, management and presentation of time- structured data from videotapes and digital media files. *Behavior Research Methods, Instruments & Computers*, **32**, 197-206.
6. Anokhin P.K. (1968). Nodular mechanisms of functional system as self-regulatory apparatus. *Progress in Brain Research*. **22,** 230-251.
7. Drai, D.; Golani, I. (2001). SEE: a tool for the visualization and analysis of rodent exploratory behavior. *Neuroscience and Biobehavioral Reviews*. **25**, 409-426.
8. Mukhina, T.V.; Lukashov, A.O.; Anokhin, K.V.; Bachurin, S.O. (2002). Analysis of animal path trajectories in terms of individual behavioral acts. *4th Int. Conf. on Methods and Techniques in Behavioral Research "Measuring Behavior 2002"*, Amsterdam, The Netherlands. 27-30 August, 2002.

A new automatic video multitracking system able to follow a large school

J. Delcourt[1], Ch. Becco[2], N. Vandewalle[2] and P. Poncin[1]
[1] *Behavioral Biology Unit: Ethology and Animal Psychology, Department of Environmental Sciences & Management, Faculty of Sciences, University of Liège, Belgium*
[2] *Group for Research and Applications in Statistical Physics (G.R.A.S.P.), Department of Physics, Faculty of Sciences, University of Liège, Belgium*

The study of shoaling behaviours is tackled by scientists by three approaches: the field, laboratory and computer modelling approach [1]. The field studies are complicated and permit rarely an analysis of individual positions. The third approach, consisting of mathematical models, is artificial. It requires a comparison with field and laboratory results to validate and to determine the explications' degree of each model. Thanks to the improvements of digital imaging techniques, the laboratory studies on fish are able to measure the behavioural parameters with precision, impossible to realize with manual recording [2,3].
Multitracking systems are rare and of recent development. These are insufficient to study the real size of shoals, often composed of hundreds of fishes. The physicists of the G.R.A.S.P. has before realised some multitracking systems to follow hundreds physical objects [4]. Our two research units have developed a new system of multitracking applied to study the group of animals.
We have attempted to track juveniles of Nile tilapia (*Oreochromis niloticus* L., Cichlids) aged from two to three weeks (length: 11 to 15 mm) in an arena of 30 x 40 cm. The depth is limited by the height of the water, permitting to obtain a quasi 2D arena. We filmed the bottom by a monochrome 2/3 CCD camera.

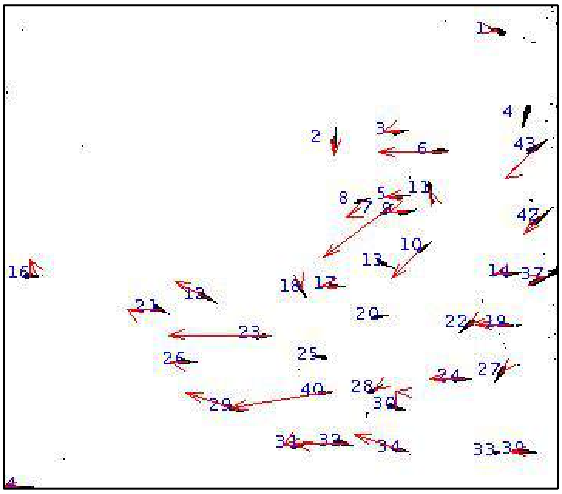


Figure 1. *Example of an analysed picture: Each black spot is a fish, with its speed vector.*

Our video multitracking system is able to detect and to track more than a hundred unmarked fishes by gray scaling technique during a few minutes [5,6]. It allows an analysis as well at the group level as at the individual level. The multitraking program is able to attribute a number at each fish and to follow each one during the whole duration of the track (up to 1 minutes). Our system allows the analysis of the movement of each individual, even if the trajectories of two fishes cross each other. It is possible thanks to the theoretical estimation of trajectory of each fish, compared with the real trajectory (analysis with feedback).

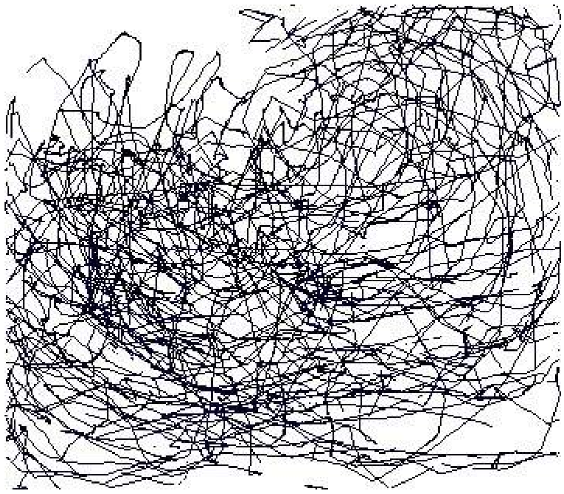


Figure 2. *Example of tracks (40 seconds) with more than 50 unmarked fishes.*

In videotracking system following only one unmarked animal, the track (displacement during a time period) is realized by the connection (in function of time) of the unique centre of gravity of each image (i.e. Ethovision developed by Noldus Information Technology, Videotrack^{TM} by View Point, …). However, the technique used to identify the numerous unmarked moving targets is different.

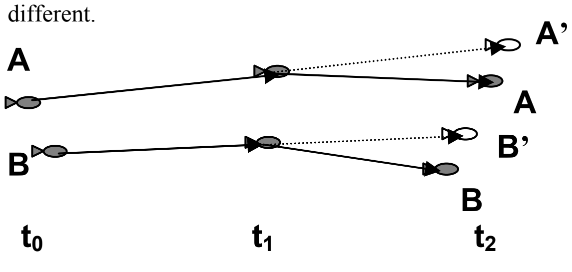


Figure 3. *Principle of the individual identification with feedback (see text for explications), A and B: two tracked fishes, A' and B': estimated position of the two fishes in time t_2 by extrapolation.*

In multitracking system, it is not possible because there is a large number of coordinate for each image. To obtain the track of each fish, we must keep account the close past of the fish. At the instant t_0, the fish is at the position X_0; at the instant t_1, the fish is at the position X_1. At the instant t_2, our system detects numerous targets: every one is potentially the tracked fish. Then the system estimates a theoretical position at the instant t_2 because we know the direction of movement and the speed of the fish between t_0 and t_1. By this extrapolation, the computer finds a theoretical point which is compared with the real detected positions. The nearest real position to the theoretical position is attributed to the tracked fish (see figure 3). Others tests are made by the tracking software before any number is attributed to a spot.

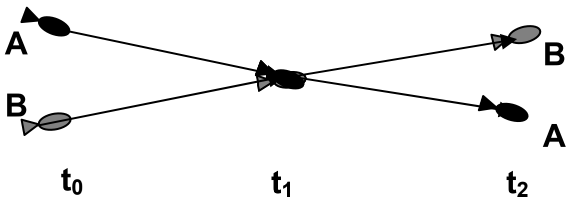


Figure 4*. The multitracking system is adapted to resolve the problem of individual identification thanks to the technique with feedback.*

This technique with feedback permits also to identify each individual after that their trajectories have crossed each other (see figure 4). Other parameters are automatically compared by the program, for example the size variation of the detected targets.

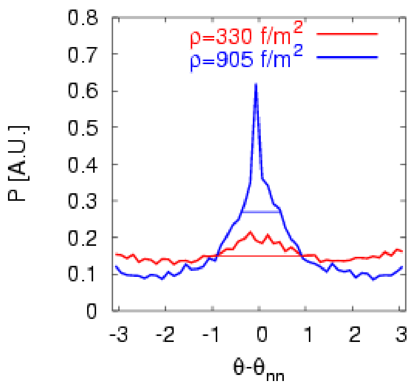


Figure 5. *Example of results obtained thanks to the multitracking system with identification of the individuals: distribution curve of the angular difference between the speed vector of two neighbouring in two density conditions (ρ=n fishes/m^2; A.U.: arbitrary unity; x-axis in radians).*

The human eye is always necessary to detect the errors in identification but reediting is possible. However, if the fishes don't move too rapidly, a rate of 25 images /s gives very good results, without too much visual expertise from the experimenter. If the fishes are too fast (i.e. escape response), it is better to use a high speed camera.
However the treatment of images is not possible in real time. The filmed sequences are digitized and treated after the experiment because of the important amount of data to analyse.
We collect a greater quantity of data permitting a better quality of the statistical analysis. Actually, thanks to this system, the sampling is practically continuous, depending of the rate of image of the camera.

Supported by a PhD student F.R.I.A grant and F.N.R.S. projects (n°2.4544.02 and 03-04-1.5.120-04) and the support of the contract ARC 02/07/293.

References:

1. Parrish, J.K.; Viscido, S.V. & Grünbaum, D. (2002). Self-organized fish schools: an examination of emergent properties. *Biological Bulletin* **202**: 296-305.
2. Noldus, L.P.P.J.; Spink, A.J.; Tegelenbosch, R.A.J. (2001). Ethovision : A versatile video tracking system for automation of behavioural experiments. *Behavior Research Methods, Instruments, & Computers* **33** (3): 398-414.
3. Ylieff, M.Y. & Poncin, P. (2003). Quantifying spontaneous swimming activity in fish with a computerized color video tracking system, a laboratory device using last imaging techniques. *Fish Physiology and Biochemistry* **28**: 281-282.
4. Vandewalle, N.; Trabelsi, S. & Caps, H. (2004). Block-to-granular-like transition in dense bubble flows, *Europhysics Letters* **65**: 316.
5. Delcourt J.; Becco Ch.; Ylieff M.; Caps H.; Vandewalle N. & Poncin P. (2004). Measuring the swimming behaviour in fish with two different computerised video tracking systems: comparison of the Ethovision system and a new multi-tracking prototype system. *XI European Congress of Ichthyology*, September 6 - 10, 2004, Tallinn, Estonia, in abstract.
6. Delcourt, J. ; Becco, Ch. ; Vandewalle, N. & Poncin, P. (2005). Influence de la densité d'individus sur les comportements de bancs chez les tilapias du Nil. *XX^{ème} conférence de la Société Française d'Etude du Comportement Animal*, 2-4 may 2005 Université Rennes 1, France, in abstract.

Interpretation of behavioral data from a computer controlled milk feeder

P. P. Nielsen[1], M. B. Jensen[2] and L. Lidfors[1]
[1] *Department of Animal Environment and Health, Swedish University of Agricultural Sciences, Skara, Sweden*
[2] *Department of Animal Health, Welfare and Nutrition, Danish Institute of Agricultural Sciences, Foulum, Denmark*

Abstract
When calves are fed milk through a computer-controlled automatic milk feeding station various factors affect their feeding behavior. Some of the factors are related to the programming of the milk feeder, e.g. number of meals per day, meal size and milk allowance, and others are related to management, e.g. number of calves per milk feeder, age difference between calves and breed. The computer-controlled milk feeder records all visits to the feeding station and these data can be used to evaluate the different feeding and management schedules. Several experiments have been performed to evaluate the various schedules influences on the calves' behavior, and it has been shown that the collected data from the milk feeder can be used to interpret the calves' behavioral reaction to the different feeding schedules and thereby improve the adjustment of the feeder to better suit the calves' preferred feeding behavior.

Keywords
Feeding behavior, calves, milk feeding, diurnal rhythm

Introduction
The trend in dairy production is a move toward more automatic systems both in milk production, e.g. automatic milking systems, and feeding management, e.g. automatic feed and milk dispensers for calves. These automatic milk feeders are often programmed in a way that forces the calves to disperse their meals over a 24 hrs period, even though it has been shown that calves prefer to eat/drink in the early morning and late afternoon [1, 2].
Valuable behavior data may be automatically generated from these automatic milk feeders. This data can be used to evaluate the effect of various feeding and management schedules on milk feeding behavior. Furthermore, we can investigate the calves' preferred diurnal rhythm of milk intake, how much, how often and also how fast the calves prefer to ingest milk. This type of information can be used in optimizing the programming and the use of these feeders to better suit the calves' social and feeding behavior.

Methods
several studies have been performed examining calves' use of milk feeders. The calves' feeding behaviors were continuously recorded by the computer-controlled automatic feeding station (HL100; Calvex, Hørslev, Denmark) (figure 1). Each calf was fitted with a collar holding a transponder that, when placed close to the feeder, communicates to the computer which calf was occupying the feeder (figure 1).
The computer recorded number and duration of visits and how much the calf was drinking or eating during each visit. The visits were recorded as five different types: visits to concentrate feeder without eating; visits to concentrate feeder and eating concentrate; visits to milk feeder without access to milk; visits to milk feeder with access to milk, but did not drink, and finally visits to milk feeder with access to milk and drinking. However, the collected data have to be manipulated before any statistical analysis can be performed.

Figure 1. Calf drinking milk from a computer controlled milk feeder (Photo: Nielsen, P. P.).

In an experiment performed during 2004 – 2005 at DIAS behavioral recordings were collected during the milk feeding period and the observer noticed that some of the visits might be misinterpreted by the computer as the wrong type of visit due to either the calves' behaviors while feeding or the programming of the computer. E.g. after a meal which was recorded as 'a visit to milk feeder with access to milk and drinking' the calves often stayed in the feeder sucking on the dry teat, the time after the meal was then recorded as 'a visit to milk feeder without access to milk' even though it was a part of the whole meal visit.

Application
Data from the computer-controlled milk feeder have been used to measure total duration of all visits and duration of unrewarded visits in relation to milk allowance.

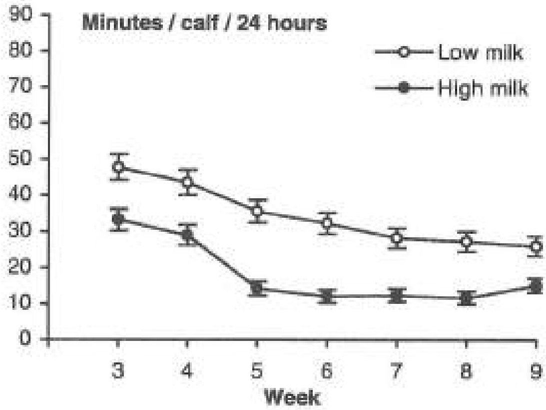


Figure 2. Mean duration of unrewarded visits for calves on low and high milk allowance (S.E. indicated) [1].

The results showed that calves of heavy breed (Danish Holstein Friesian and Danish Red) on high milk allowance had a lower total occupation of milk feeder compared to calves on low milk allowance. Additionally calves on high milk allowance had a lower duration of unrewarded visits than calves on low milk allowance (figure 2) [1].

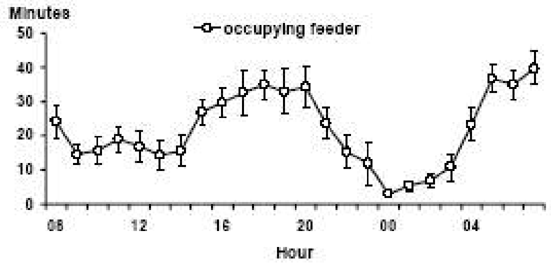


Figure 3. Mean duration of feeder occupancy, minutes per h during a 24 h period (S.E. indicated) [mod. from 1].

Even though the milk feeder producers by various programming encourage the calves to distribute their milk meals evenly over a 24 h period, the calves still show a distinctive diurnal rhythm (figure 3) [1, 2]. This diurnal distribution of feeder occupancy have been shown not to differ between groups of 12 calves per feeder and groups of 24 calves per feeder, which support the theory that calves are reluctant to feed outside their activity periods. Rather than changing their diurnal rhythm, the calves ingested the milk faster when kept in groups of 24 calves per feeder compared to when kept in groups of 12 calves per feeder [2].

4 Conclusion

Automatically collected data from an automatic milk feeder for calves can be used to evaluate various feeding and management schedules on milk feeding behavior. This information enables us to optimize the programming to better suit the calves' behavioral preferences.

References

1. Jensen, M.B.; Holm, L. (2003). The effect of milk flow rate and milk allowance on feeding behaviour in dairy calves fed by computer controlled milk feeders. *Appl. Anim. Beh. Sci,*. 82(2), 87-100.
2. Jensen, M.B. (2004). Computer-controlled milk feeding of dairy calves: the effects of number of calves per feeder and number of milk portions on use of feeder and social behaviour. *J. Dairy Sci.*, 87, 3428-3438.

flyTracker: real-time analysis of insect courtship

J.A. Heward, D.A. Baker, T.C. Lukins and J.D. Armstrong
School of Informatics, University of Edinburgh, Edinburgh, Scotland.

Abstract
We have developed a software package designed to automate the analysis of the courtship ritual of male Drosophila flies. Our system comprises three linked packages that (a) track objects in live video feeds or from files; (b) replays video files with and without animation/simulations of the tracked log files and allows expert annotation to be added and; (c) which learns classification rules from expert training data and applies it to new datasets. In this paper we focus on the tracking package.

Keywords
Tracking, bounding box, occlusion, Drosophila.

1 Introduction
The small fly *Drosophila melanogaster,* often referred to as the fruit fly, and still used in high-school biology genetics classes has been studied by biologists for close to 100 years. It is one of the key model organisms for studying biological processes including behavior and neuronal function. The courtship ritual is one of the commonly used behavioral assays for the organisms yet despite intense interest it is still quantitatively measured by visual observation by trained staff. An overview of the biological motivation and summary of the technical advances is presented in the accompanying manuscript by the same authors [1]. In this short paper we focus on the technical challenges and solutions we considered in building the tracking system.

Our requirements analysis identified a number of essential and desirable features. The key ones were that the system should: (1) be modular and extendable. (2) be platform independent. (3) run from live video feed or video files. (4) be able to visualize all processes.

We chose JAVA 1.4 as a development language since it would enable all of our key objectives. We had available, and tested all components on, a wide range of platforms from multi-cpu linux clusters to single processor laptops and workstations.

2 Materials and methods
The tracking module is the most important of the modules. To simplify the task we chose a single camera platform, the AXIS 205 net camera (www.axis.com). These small cameras have built in web servers that control the camera settings and return single snapshot images, live video feeds or movie files in response to a http style URL. We initially recorded several courtship encounters on the system under different pixel resolutions and frame rates to see what the minimum specifications would be for a human observer. All settings were for grayscale video footage taken at 320 by 240 pixels and 10 frames per second.

We then collected a series of test examples that were 5 minutes each and where a varying amount of courtship activity was noted during casual observation.

3 Tracking

3.1 Initial Processing
The task for the tracker component is to convert each frame of video into a log file that describes the identity, the number, position and angle of orientation of each fly sized blobs located in the video. It does require a basic parameter – either the expected number of flies or the expected size range of flies (given one the other can be estimated from the video).

Although all footage we used was in grayscale, we identified at an early stage that color is often used to identify flies (e.g. through genetic determinants of eye color). Therefore all processing was performed in RGB color space.

3.2 Filtering
To make the blob detection simpler, some initial image smoothing was required. We investigated the use of Gaussian Convolution, Sobel Filters and Average Convolution. In terms of both speed and image quality we found the average convolution methods to be the most appropriate.

3.3 Thresholding
Next we considered a number of thresholding methods to identify pixels in each frame that we should consider for segmentation. We evaluated mean thresholding, adaptive thresholding, histogram threshold detection, background removal and pseudo adaptive thresholding (see below). Adaptive thresholding and histogram threshold detection gave the best results but were expensive computationally. Background removal was implemented as an option but was not required for the task of removing ill defined shadows etc. Mean and pseudo adaptive thresholding worked well and were computationally cheap. We chose the later since it was better and handling uneven illumination. We also included a simple erosion operator to split barely touching objects at this stage.

In pseudo adaptive thresholding, the frame is split into a small number of non overlapping sub frames and each of these is mean thresholded. This results in a system that is resilient to small changes in light intensity over the image but is much faster to compute (around 2ms on our systems compared to 20ms for adaptive thresholding). Best results are when the image is split into between 9 and 25 sub frames.

3.4 Blob Detection
We considered a model based approach but decided that it would be too slow and inflexible. Moreover, one of our longer-term applications involves the analysis of behavior in a microgravity environment where new emergent behaviors are possible. Thus a predefined model would be inappropriate. We adopted a labeled feature approach which worked well with the small blob size processing an entire frame in approximately 5 ms. We then fit a bounding box to the blobs (with some simple constraints for maximum and minimum size of a fly sized blob).

3.5 Blob Tracking

To track the flies we implemented the Bounding Box method of [2]. We added an extra rule that is invoked when there are unmatched blobs in consecutive frames (i.e. the fly moves fast enough to move beyond its previous bounding box in a single frame). In this situation we resort to a best-fit algorithm on the positional data of the unmatched blobs. To obtain size, positional, and orientation data the area, centre of mass and the vector between the centre and point furthest from the centre are calculated.

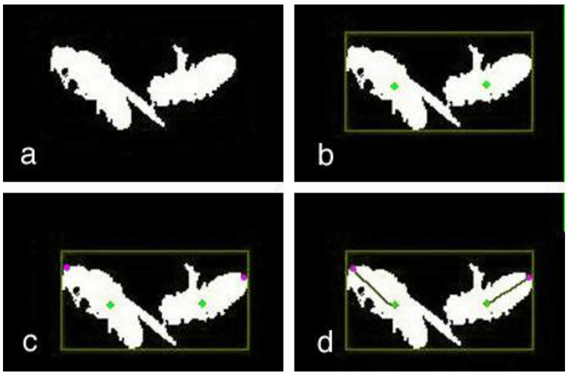


Figure 1 *Resolution of occluding flies: (a) Two touching flies within a bounding box. (b) K-means clustering (K=2) finds the centers of the two flies (green blobs). (c) Locate the point furthest away from the centers (magenta). (d) The angle between these points is taken as the orientation.*

When flies occlude we attempt to maintain the size, position, orientation and identity information. First, K-means clustering is performed within the bounding box (K=2). Next, the means obtained are matched on a best-fit basis with the positions of the corresponding blobs from the previous frame to maintain the identity and position. The size is retained from the previous frame. Calculating the furthest point from each mean within the bounding box is used to generate the orientation (figure 1). Combined, these methods handle the merging and splitting of more than two objects.

4 Evaluation

The tracking package generates a text log file. The log contains a list of blobs in each frame, their identity, area, X and Y coordinates and their angles of orientation. The change in position between each frame is usually less than 20 pixels which is feasible for the resolution, frame rate and fly size used. Viewing the footage confirms that the flies move about half a body length per frame on average. Occasionally the movement is higher, around 40 to 60 pixels per frame. This is likely due to periods of footage where the male is chasing the female around the container as the movement here is very rapid or jumps in the video footage. Using the methods described here we observed just a single label error in 3000 frames of footage and further testing is on-going.

5 Conclusions

Instead of trying to deal with occlusion by using image based separation techniques we demonstrate that it is possible to deal with occlusion at a higher level. The methods described here reduce the computational time required sufficiently for the tracking of multiple fast moving and often occluding objects to be possible on common PC hardware. With respect to our primary aims, we have used the system to track two interacting flies in real time from a live video feed.

This work was supported by grant no. S18944 from the BBSRC (JDA).

References:

1. Armstrong, J.D., Baker, D.A, Heward, J.A. and Lukins, T.C. (2005) Sex, Flies and no Videotape. *(this volume).*
2. Fuentes, L.M. and Velastin, S.A. (2001) People tracking in surveillance applications in 2nd IEEE International Workshop on Performance Evaluation on Tracking and Surveillance, PETS 2001 (Kauai, Hawaii-USA, 2001).

Use of image analysis to measure the impact of shimmering behavior in Giant honeybees on predatory wasps

E. Schmelzer, G. Kastberger
Department of Zoology, Karl-Franzens University of Graz, Austria

Abstract

The Southeast Asian giant honeybee *Apis dorsata* nests in the open. Therefore, a series of defense strategies have been evolved to cope with the predatory pressure of birds and wasps. Here, *shimmering behavior*, evoked by approaching wasps, includes hundreds of the roofing bees of the nest which flip their abdomens synchronously and cascadically. Observations on natural wasps hovering in front of the nest allowed us to trace the spatial and temporal behavioral patterns of wasp and surface bees. For that, image analysis of video sequences enabled both a semi-automated assessment of the flight trajectories of the wasp, and the quantification of the waving as colony response of the giant honeybees. Our data address the multipleness in threatening cues and reactions in the interplay between hunting predator and self-defending prey. It provided evidence for the repellence effect of defense waving against predatory wasps.

Keywords

Images analysis, Giant Honeybee, *Apis dorsata*, defense behavior, shimmering behavior

1 Introduction

The hornet-sized giant honeybees live in single-comb nests, which are attached to branches, rocks or buildings. A nest measures up to two meters in the horizontal span and may comprise over 100.000 individuals. The bees themselves cover the comb in multiple layers, forming the *bee curtain*. On the surface, the bees are locomotorily active only in the *mouth zone*, where they arrive, depart, dance and show trophallaxis. In the *quiescent region* they hang around mostly uniformly oriented, with the head up and the abdomen down. As these bees nest predominantly in the open, they are more exposed to predators than cavity-dwelling species. It is this predatory pressure that has evolved a series of defense strategies [1]. One of these is the *shimmering behavior* [3], which is particularly evoked by approaching wasps, moths or birds. This colony response can be termed more interpretatively as *defense waving* [4].

What we only know so far is that this behavioral trait displays a remarkably fast communication within the nest, unique in the animal kingdom. During waving thousands of colony members are aligned to flip their abdomens in a highly coordinated reaction within a fraction of a second. It is evoked by visual stimuli of mainly predatory impact and is adaptive for internal and external defense purpose: internally, it may spread information about predatory stimuli in order to arouse and alarm colony members; externally, it may provide visual patterns in order to confuse, misguide and repel nest marauders.

Observations of natural wasps hovering in front of the nest allowed us to watch the interplay of the potential predator and the potential prey. In this paper we ask whether defense waving has a repelling impact on wasps. For that, we used image analysis techniques to quantify the spatial and temporal behavioural patterning of wasp and bees as a sequence of recurrent threatening and reactive behaviour in the interplay of hunting predator and self-defending prey.

2 Methods

2.1 Video recording

Wasp attacks on giant honeybee nests were video recorded (using the DV camcorder Sony VX 2000) in Chitwan area, Nepal. The *Apis dorsata* nest, which was selected for the recording, was located 20 m above ground on the interior ceiling of a water tower (Fig.1. arrow). This location had two advantages; first, hovering wasps regularly occurred in front of the nest, and second, the ceiling of the water tower provided contrastful detection of individual bees and wasps.

Figure 1. *The experimental nest of Apis dorsata on a water tower in Chitwan, Nepal. Left-side picture, the water tower with the red arrow pointing to the location of the nest which was attached at the interior ceiling; right-side picture, nest attachment site viewing upwards. At the center (encircled by green line) the experimental nest can be seen.*

2.2 Data analysis

We transferred the video film into tif-images at a rate of 25 frames per s, using Avid X-press pro as full-featured editing software for real-time DV and analog video. Image analysis was mainly done with the software Optimas, Flir. In total, we analyzed a video sequence of 7727 images.

Measuring defense waving. Macros in Optimas established a movement detector by which changes of grey pixel values in nest zones were spotted imagewise within specific noise thresholds. Such divisions are used as markers for bees which flip their abdomen during waving, or which change their position by locomotion. For the *quiescent* nest area, the scoring of movement detection is a usable measure of waving activity itself, for the *mouth* area it would be a marker for locomotion.

Measurement of flight traffic. The hovering wasp was located in single frames interactively by mouse clicks on its thorax. We assessed its shortest distance to the nest, its momentary direction, ground velocity and ground acceleration; additionally, the trajectories of outflying and incoming bees., Again, interactive detection was preferred, because automation was uncertain and isolation of flying wasps or bees not possible, when they were too fast or too slow.

Higher-level information. Waving is a colony response to approaching wasps. In this paper, however, we were interested to measure the reactivity of the wasp to the waving. For that, as trigger event for pooling we chose the lowest intensity value of waving during recurrent activity. We expected to get a turning response of the wasp, away from the nest, after the waving reached a certain intensity level. Lastly, sequences of predation were defined, and the hunting success of the wasps determined.

3 Results and Discussion

Approaching wasps released defense wavings at the giant honeybee colony. After the onset of the wave the wasp, when hovering nearby, responded within 320ms by 3,23 ±1,20 cm (mean and mean error; n = 31); it turned off the nest and continued hovering at a greater distance from the nest (53,7 ± 0,3 cm; mean and mean error; n = 4407 40ms intervals). This data support the *repellence hypothesis.* Wasp touching the nest of giant honeybees would act at high risk to be grabbed, dragged into the bee curtain and heat-balled to death [2]. Obviously out of this reason the wasps choose a less hazardous strategy to exploit giant honeybee nests as food resources; the wasps wait at secure positions to hunt arriving or departing worker bees. The experimental wasps tried in the average of 6s to catch a bee; one in 30 trials were successful.

The bees have evolved a series of defense strategies against wasps, they repel them away from the nest, not allowing direct contact, otherwise they would grab and kill them. Despite these antipredator capacities they are not prepared to chase off wasps if they are only 50 cm apart the nest. Mass counterattacks, which are regularly triggered by approaching birds, have not been observed against wasps. We speculate that the hunting success of wasps when hovering in front of the nest does not essentially lower the fitness of the honeybees. In this respect, co-evolutionarily seen, the predator is one step ahead its prey.

Supported by i-swarm project (IST FET-open Project 507006) and Grant from University Graz November 2004

References:

1. Kastberger G, Winder O (2001) Defence strategies of Giant honeybees - Documentary Film, Wissenschaftlicher Lehrfilm für den Unterrichtsgebrauch an Universitäten. 23 min. Produktion epo-film Wien-Graz 2001
2. Ono M, Igarashi T, Ohno E, Sasaki M (1995). Unusual thermal defense by a honeybee against mass attack by hornets. *Nature* **377**, 334-36
3. Sakagami SF (1960) Preliminary report on the specific difference of behaviour and other ecological characters between European and Japanese honeybees. *Acta Hymenoptera* **1**, 171-198
4. Schmelzer E, Hartbauer M, Kastberger G (2004) Defence waving in the Giant honeybee (*Apis dorsata*). *Proceedings of the Deutsche Zoologische Gesellschaft, 97. Annual Meeting at Rostock.*

Two real space tests to study spatial navigation impairment in different stages of Alzheimer's Disease

K. Vlček[1], E. Kalová[2], J. Laczo[3], E. Jarolímová[4], J. Hort[3], J. Bureš[1]
[1] *Department of Neurophysiology of Memory, Academy of Sciences, Czech Republic*
[2] *Neurology, Na Homolce Hospital, Prague, Czech Republic*
[3] *Neurology, Motol Hospital , Prague, Czech Republic*
[4] *Gerontology centrum, Prague, Czech Republic*

Abstract

We developed two tests to specify impairment in allocentric and egocentric navigation and in spatial memory for several locations. Both real space version in a 2.8 m diameter circular arena and map-view computer version was used. Control group was compared with subjects in several stages of AD. In the Hidden Goal Task (HGT), the subject should find a hidden goal inside of the arena. The start position and/or cues on the arena wall could be used for orientation. In the Spatial memory test (SMT), the subject was shown 2, 4 or 6 locations inside of the arena and then was asked to recall their position in the order they were presented. The errors in the remembered order and in locations of the points were evaluated separately. Our results suggest general spatial navigation impairment in AD and spatial working memory impairment in the early stages of AD.

Keywords

Spatial memory, Allocentric Navigation, Episodic memory, Real space test, Virtual reality, Alzheimer's disease

1 Introduction

Enhancing the early detection of Alzheimer's disease (AD) has been the aim of much effort. The neuropsychological approach has focused on developing tests aimed at the most early impaired cognitive functions in AD, the episodic and prospective memory and spatial navigation.

The structure among the most early and most heavily impaired in AD is the hippocampus. The role of the hippocampus in spatial orientation was suggested first in animals and was based on the discovery of spatial specific firing neurons and on the effects of hippocampal lesions in navigational tasks. The task used in these original studies, and used widely until now, is the Morris water maze. The rat has to swim for a platform in a pool filled with opaque water. Two types of navigation can be tested in this task, egocentric relative to the subject's current position and allocentric, relative to distant cues. Only the allocentric navigation was shown to be dependent on hippocampal function.

The most often used definitions of episodic memory rely on the conscious experience of recollection. Concept of 'episodic-like' memory trace, consisting of information 'where', 'when' and 'what' concerning a unique event was developed. The merit of this definition is that it may be demonstrated behaviorally in animals.

The navigational tasks and table-top tasks of spatial memory for human differ in terms of the perspective from which the observer is required to operate and also in their frames of reference. In order to gain more complete control over experimental stimuli while allowing simulated exploration of space from a first-person perspective, computer-simulated or virtual reality environments have been widely used to study navigation. A serious limitation of these experiments is that some aspects of the environment are not easily understandable to subjects who are not sufficiently familiar with computers and that the virtual world is not immediately comparable with the corresponding real world.

We developed two real space navigation tests for human together with their computer analogues. In our previous study [1], we documented the impairment of AD subjects in both of these tests. The present study was aimed at characterization of spatial navigation impairment in several stages of AD and Mild Cognitive Impairment.

2 Methods

2.1 Subjects

The total of 77 subjects was diagnosed into four groups: early and middle stage of Alzheimer's' disease groups (E-AD and M-AD), Mild cognitive impairment group (MCI) and control group. The diagnoses were based on NINCDS–ADRDA, NINCDS–AIREN and DSM-IV criteria and supported by MRI volumetry, CSF biomarkers assessment and standard neuropsychological examination.

2.2 Apparatus and testing environment

The real navigation version of the tests was performed in a fully enclosed arena 2.9 m in diameter surrounded by a 2.8 m high cylindrical dark blue velvet curtain, the Blue Velvet Arena (BVA). A TV camera above the center of the arena was connected to a computerized tracking system recording position of an infrared LED on the top of a standing pole used by the subject to indicate places on the floor. The TV camera was surrounded by eight laser pointers aimed at eight circular areas of the floor spaced at 45° intervals. Eight computer controlled light patterns could be projected at 45° intervals on the cylindrical wall 1.5 m above the floor. They were only visible when one or two of them were turned on to serve as landmarks for allocentric navigation. Similarly controlled was the projection of a single red point on the wall indicating the start location. The computer version was performed on a PC computer in the same room where the BVA was placed. Both real space and computer versions of the test were controlled by self-made programs written in Quick BASIC for DOS.

2.3 Hidden Goal Task

This task was designed as a real space analogue of Morris water maze for human. It consisted of 3 tests, in each of which a computer version was followed by a real space version. Individual tests examined specific modes of navigation: while in test 1 both the start and the cues could be used to find the goal, only start was displayed in test 2, so that the subject had to rely on egocentric strategy when searching the goal. Only the two cues could be used to find the goal in test 3, implying the use of allocentric navigation.

In the computer test the experimenter showed the subject the arena displayed on the computer screen as a large circle with a starting point marked on the arena circumference by a small circle and the orientation cues represented by short red and green arcs at the arena circumference. The subject was asked to use the mouse to locate the goal position. The correct position of the goal was then displayed for 5 s, the screen was cleared for 5 s and new trials was started with the positions of the goal, cues and/or start rotated around the arena center by a multiple of 45°. After eight trials, each virtual test was completed and followed after a short pause by real space version in the BVA.

The experimenter took the subject into the BVA and showed him/her the positions of the cues and of the start, which were in the same mutual position as on the computer screen. After the subject located the goal by use of the LED carrying pole, its exact position was projected on the floor by a laser pointer aimed at that particular point. After a 5-s pause the start and cues were rotated in a pseudo-random sequence to new positions and the subject was instructed to go to the new start position and proceed with the test. After the eighth trial of the first real test was completed, the examination continued with the computer and real versions of test 2 followed similarly by test 3. The performance was evaluated as the distance between the position recalled by the subject and the position of the goal.

2.4 Spatial Memory Task

The objective of the Spatial Memory Task was to remember and then recall positions of 2, 4 and 6 locations. The locations were presented in sequence and the subject was instructed to recall the locations in the same order as they were presented. The test thus includes two aspects of the episodic-like memory, the 'where' and 'when' information. It consisted of a computer version on a computer screen and a real space version in the BVA. The computer version always preceded the real space version.

The subject was instructed that s/he will be presented one by one 2, 4 and then 6 locations inside of the arena and will be asked to recall their position in the presentation order. The subject was presented each point for 3 seconds, wirh 10 second pause with black screen between. After marking the assumed position of the locations, the subject was shown for 2 seconds their correct position with the size of error s/he made. After this test with 2 locations, then subject was similarly tested recall of 4 and 6 locations.

In the computer version, the subject was shown on a computer screen the locations as red discs inside of a circle representing the BVA. In the real space version, the subject was shown the locations on paper cards. On each of them was drawn a circle representing the arena and inside the circle there were blue dots marking the target positions. To mark the locations inside of the BVA, the subject navigated from the start towards the presumed position of the location, marked its positions with the LED carrying pole and returned back to the marked starting point.

The performance in each test was evaluated as the distance between the position recalled by the subject and the correct position of the locations. The evaluation was performed both considering and not considering the order, in which the subjects recalled the locations.To compare the success in finding the locations with some baseline random level, 20 pseudo-random persons were used generating hits randomly distributed all over the arena surface by a self-made program.

3 Results and discussion

We found spatial navigation impairment in all experimental groups and similar results in the real space and virtual versions of the tests. Both AD groups were impaired in all three subtests of the HGT, suggesting general navigational impairment. In agreement with the developing hippocampal damage in the progress of AD and of the role of hippocampus in allocentric navigation, the middle stage AD group was worse than early stage AD group in the third subtest of HGT testing allocentric navigation. The MCI group was also impaired relative to controls in this third subtest, supporting the reports of hippocampal damage in MCI.

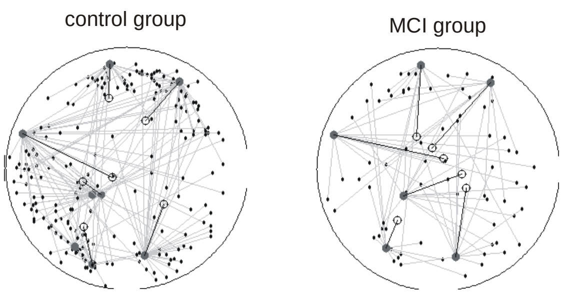


Figure 1. Example of results of the Spatial Memory Task with 6 points to remember. The small gray discs are the target locations, the black dots, connected by gray lines with the gray discs, are individual attempts to hit the target by the subjects. The gray lines connect the individual attempts with the target locations, which the subject tried to hit. The empty circles connected with the gray disc represent the gravity center of all the attempts to hit the target point specified by the line.

In the SMT, the order of the locations turned out to be more difficult to remember and more differentiating between the groups than their position itself. While the MCI group was in all cases similar to controls in remembering the position, it was impaired in remembering the order of 4 and 6 locations. The middle AD group was in both real space and virtual tests impaired relative to controls and in the real space tests similar to the pseudo-random group. Our results suggest episodic-like memory impairment in the MCI group and general episodic and spatial memory impairment in the AD.

Supported by GACR grant 309/05/0693, MSMT project 1M0002375201 and AV0Z50110509.

References:

1. Kalova E.; Vlcek, K.; Jarolimova, E; Bures, J. (2005). Allothetic orientation and sequential ordering of places is impaired in early restages of Alzheimer's disease: corresponding results in real space tests and computer tests. *Behav. Brain Res.*, **159** (2):175-186

Measuring the behavior of mink with permanent access to running wheels: the design of the wheels and a description of the digital video system used to register the behavior

E.L. Decker
Dept. of Animal Health, Welfare and Nutrition, Danish Institute of Agricultural Sciences, Tjele, Denmark

Abstract
In order to quantify the activity level of mink in different farming systems 32 running wheels equipped with magnetic switches, were installed on alternate cages in a set of 64 cages. The switches were connected to a digital IO card in a PC. All signals from the switches were timestamped and registered in a file. The PC was also equipped with AD video cards for 8 video cameras. The digital video was stored in a proprietary video database, allowing datetime stamped keypress behavior scoring of the recordings.

Keywords
Digital video, keypress scoring, running wheels, mink.

Introduction
In order to quantify the activity level of mink in different management systems 32 running wheels equipped with magnetic switches, were installed on alternate cages in a set of 64 cages (Fig. 1). The mink had permanent access to the wheels.

Figure 1. *General and detailed views of the wheels.*

Equipment
The wheels were designed to be installed on standard mink cages. Entry to the wheels was through a hole in the cage, identical with and opposite the entrance of the nest boxes of the mink (Fig. 1). After any activity the wheels always returned to a position with the magnet on the opposite side of the switch. Mink entering the wheel had to make at least half a turn for movement to be registered, making certain that only full revolutions were measured.
The switches were connected to a digital IO card [2] in a PC. All signals from the switches were timestamped and registered in a file.
The PC was also equipped with AD video cards for 8 video cameras. Moving the cameras every day resulted in 24h video for each cage, every 8th day.

The programme collecting the video is a “standard” security system from MSH-video [1] in Latvia. It is a Client-Server system with everything necessary for surveillance like, network access to the server, reduced frame rate for “time lapse” video, software motion detectors (motion intensity and area can be defined) for each camera. What is not standard about the system is saving the recordings in a video database instead of in a video file (.mpg), allowing huge video files and full control of the recorded video with features such as a timeline for all cameras showing detected motion (Fig. 2), instant access to all records, the possibility to scan the video with a predefined timespan, conversion to standard .mpg for automatic analysis and, most important for analysing video, showing, registering and timestamping (video time of course!) all keypresses while viewing the video (Fig. 3). The resulting log file can be handled in any data treatment system.
For analysing the behavior in predefined areas like the feeding area or the entrance to the wheel, the system can automatically search through the recorded video and find all video sequences with movement in just this area (Fig. 3).

Conclusion
Installing the digital video recorder in the same PC as the cards controlling the equipment, allows faultless synchronisation of all data. The digital video system used, besides being cheap, was exeptionally suited to behavior studies.
The results of the project will be published by S.W. Hansen from DIAS. Just one note for the technically minded: make sure to put good bearings in the wheels, some mink ran up to 17000 revolutions a day (about 25 km).

References:

1. www.guard.lv
2. www.measurementcomputing.com

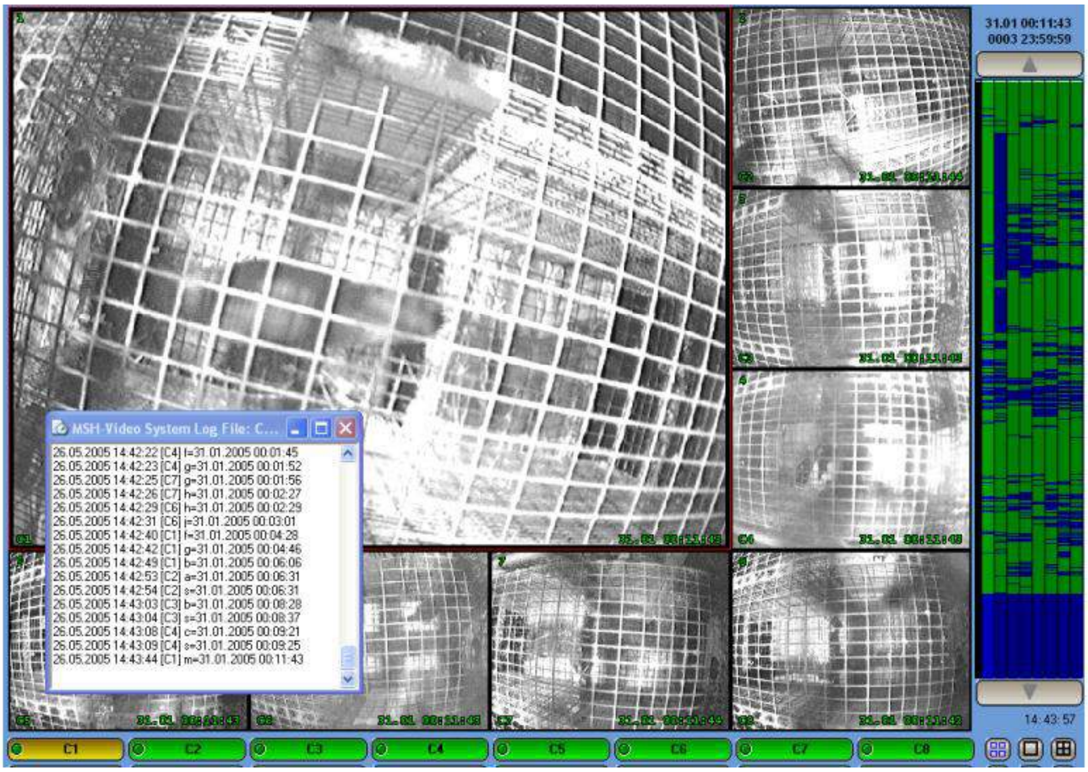


Figure 2. *General view of the recorded video. Note the window with the recorded keypresses, showing the date and time of viewing, the camera in focus, the key pressed and the date and time of the recording.*

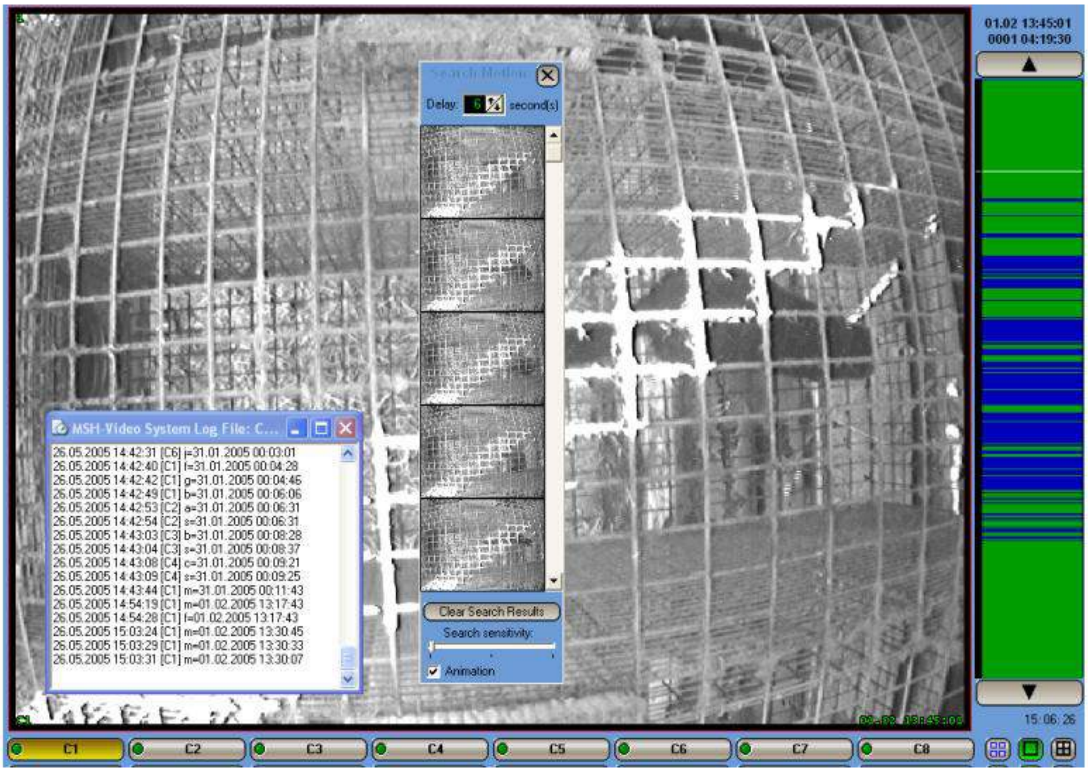


Figure 3. *Zooming in on one camera. Note the window with the automatically generated video with movement in f.x. the feeding area. When pressing a picture in this window the main screen will also switch to this time.*

Automated recognition of rat and mouse body landmarks using 2D kinematical model

T.V. Mukhina, K.V. Anokhin
Department of Systemogenesis, P.K. Anokhin Institute of Normal Physiology, Moscow, Russia

Abstract
This paper describes an approach for video tracking of an animal (mouse or rat) to recognize its body landmarks. The recognition is performed using two-dimensional kinematical model of the animal developed by the authors. The model accounts for anatomical features of the mouse or rat body and describes relative position and rotations of its parts, such as tail base, nose, waist and neck. It can be used to find the animal's head and tail, recognize turns of the body and the head and discriminate certain forms of animal behavior. One of the features of this method is that it does not require artificial body markers. This significantly expands its possible applications in animal behavior experiments.

Keywords
Video tracking, model-based motion capture, computer vision.

1 Introduction
Video tracking, namely automated determination of animal movements by video camera, is an important tool for automation and standardization of behavioral experiments. However, representation of an animal by the gravity center of its body, that is employed in most of the video tracking programs (for example, [1,2]) and calculation of both static (trajectory length, times) and dynamical parameters (for example, distribution of linear and angular speeds over the trajectory [3]) does not allow to describe animal behavior in terms of postures and different types of behavior. Unfortunately so far great majority of suggested posture recognition algorithms are aimed on people rather than animals [4].

This paper describes an approach to video tracking of a rodent (e.g. mouse or rat) to recognize its postures. Recognition of rodent postures is much more difficult than that of humans' or even of horses' postures, for example. A human body is build of fairly rigid parts which can be relatively easy simulated using geometrical figures. On the contrary, mouse body is "soft", and its visible surface usually cannot be reduced to its skeleton. So, it is relatively difficult to approximate such a shape-changing body by a set of stiff geometrical figures as it is usually done with the humans. In the current work recognition of mouse body postures is performed using two-dimensional kinematical model of the animal body. One of the features of this method is that it does not require the use of markers that are often employed in the optical motion capture techniques [4].

2 Model description
The model was developed for the case of video tracking in the horizontal projection (video camera is set above the experimental arena), since such configuration provides the most complete information about animal movements required in most of behavioral experiments. The model describing animal movements should account for anatomical features of its body and be relatively simple. Snout, neck, waist, tail base as well as relative shifts and rotations of these body parts are among most significant for understanding of movements of the mouse or rat body.

A kinematical geometrical 2D model of a rodent body was developed (Figure 1). The model has 2 axis of rotation in the horizontal surface: 1 – rotation of the frontal part of the body (shoulders + neck + head as a whole) relative to the center of the middle circle simulating the abdomen (angle α_1); 2 – rotation of the neck and head (as a whole) relative to the shoulders (angle α_2). Points 1, 2, 3 and 4 correspond to the center of rodent's spinal column, neck basis, pelvis and skull base, respectively. Parameters D_1 and D_3 characterize changes of the body length at its flexions in vertical plane relative axis A-A′; D_2 is responsible for stretches of the neck. Orientation of the model in space and size of its parts were determined at the initialization stage of the model-based tracking.

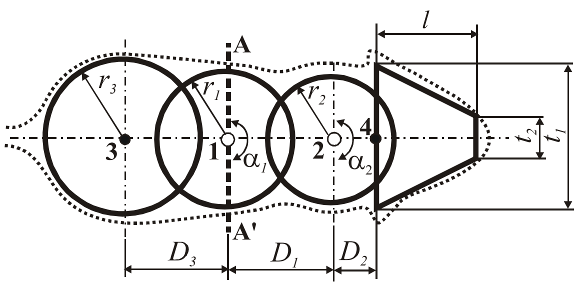


Figure 1. *Model of a mouse body. Dots indicate contour of the body.*

Each frame was treated in the following way: segmentation of the outline of an animal body, isolation of tail from the body and then fitting of the model to the body. Initialization of the model parameters was performed at the first frame of the video flow. The center of gravity of the tail was used to determine anterior-posterior orientation of the model as a whole. Center of the model (point 1) was initially determined as a center of an ellipse approximating contour of the body. Then both constant (sizes of the elements) and variable (angles and shifts) parameters of the model were computed. For all other frames only distances D_1, D_3 (as the result of flexion of the body relative axis A-A′), D_2 and angles α_1 and α_2 were allowed to change.

3 Results
The model was tested by video-tracking mouse behavior in a rectangular cage of 36x24 cm. The model satisfactory describes relative positions of different parts of the

moving mouse body. It can be used to find the animal's head and tail, recognize turns of the body and head and distinguish certain behavioral acts (for example stretching). Examples of the model application are shown in Figures 2 and 3.

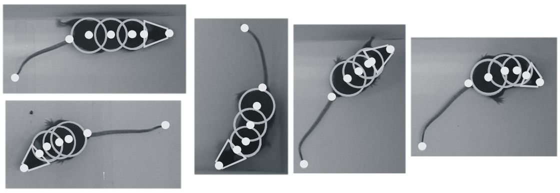


Figure 2. *Configurations of the model elements at different postures of the mouse.*

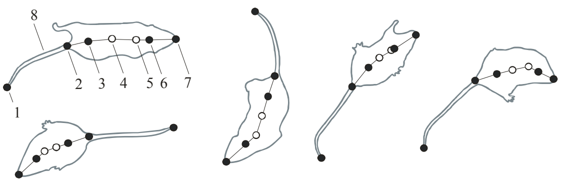


Figure 3. *Fragments of tracking of mouse movements using 2D* model. *1 – End of the tail, 2 – tail base, 3 – pelvis, 4 – point of flexion in the waist area of mouse spine, 5 – neck base, 6 – skull base, 7 – nose, 8 – mouse contour.*

4 Conclusions

We have shown that a difficult problem of "soft" body landmarks recognition can be solved within a framework of relatively simple geometrical model. Further improvement of the model may include establishment of animal-dependent anatomically-based restrictions on relative positions of the elements of the model, introduction of additional elements accounting for extremities of an animal, *etc*. However, creation of a realistic 3D body model will be the most constructive approach for the detailed recognition of animal's movement.

Supported by Russian Foundation for Basic Research (Grant 03-04-49739).

References:

1. Noldus, L.P.J.J.; Spink, A.J; Tegelenbosch, R.A.J. (2001). A versatile video tracking system for automation of behavior experiments. *Behavior Research Methods, Instruments & Computers*, **33**, 398-414.
2. Mukhina, T.V.; Bachurin, S.O.; Lermontova, N.N. (2001). Versatile computerized system for tracking and analysis of water maze test. *Behavior Research Methods, Instruments & Computers*, **33**, 371-380.
3. Mukhina, T.V.; Lukashov, A.O.; Anokhin, K.V.; Bachurin, S.O. (2002). Analysis of animal path trajectories in terms of individual behavioral acts. *4th Int. Conf. on Methods and Techniques in Behavioral Research "Measuring Behavior 2002"*, Amsterdam, The Netherlands. 27-30 August, 2002.
4. Moeslund, T.B.; Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, **81**, 231-268.

Determining minimal space requirements of group-housed rabbits based on their spatial distribution

E. Van Poucke[1], L. Maertens[2], J. Van den Abeele[3], L. Lens[3], F. Tuyttens[1]
[1] *Department of Mechanisation, Labour, Buildings, Animal Welfare and Environmental Protection, Agricultural Research Centre, Merelbeke, Belgium*
[2] *Department of Animal Nutrition, Agricultural Research Centre, Melle, Belgium*
[3] *Terrestrial Ecology Unit, Department of Biology, University of Ghent, Ghent, Belgium*

Abstract
Current commercial housing of fattening rabbits in small barren cages highly constrains the performance of natural behavior. Housing these animals in larger pens may be a step in the right direction, but little is known about the optimal group and pen size. To assess the spatial requirements of group-housed animals a new method, based on the 'critical inter-individual distance', was tested. Additionally, the effect of environmental enrichment was evaluated. Fourteen groups of 22 fattening rabbits were housed in pens of seven different sizes in either a barren or an enriched interior. Top-view digital photographs were taken every week to determine the spatial distribution of the individual rabbits inside each pen, and to calculate a heterogeneity index (homogeneity indicating a restricting environment). This index increased with pen size and the presence of enrichment. Rabbit age negatively affected the index but this effect decreased with pen size and the absence of enrichment. These results suggest that when housing rabbits in groups, the larger the pen the better, but smaller pen size may be compensated by enrichment. The developed distribution method seems efficient for determining minimal space requirements of group-housed animals but ought to be validated and refined.

Keywords
Spatial distribution, spatial requirements, group housing, enrichment, fattening rabbits

1 Introduction
Fattening rabbits are generally housed in small groups in small barren wire net cages, which constrains the performance of natural behavior [3, 5] and causes severe animal welfare problems [1, 4, 7, 8]. Pens for the housing of larger groups of rabbits have therefore been developed as an alternative housing system for fattening rabbits [3]. The increased group size has the advantage of increasing the absolute space available per animal, but the disadvantage of an augmented risk of aggression and mortality and a slightly lower production [3, 6]. However, little research has been conducted towards optimizing pen-housing systems with regards to stocking density and environmental enrichment. Although such an optimization has the potential of reducing the abovementioned disadvantages of pen-housing systems, it is hampered by the lack of an objective method for determining the spatial requirements of group-housed animals. The aim of this study was to investigate the effect of pen size and environmental enrichment on the distribution of fattening rabbits in the pen.

2 Material and Methods
To assess the spatial requirements of group housed animals, a new method was tested. It hypothesizes animals to distribute themselves in a homogeneous way if the allocated space is restrictive, while heterogeneously if there is no space restriction. This hypothesis is based on the 'critical inter-individual distance', i.e. the minimum distance an animal will allow a conspecific to approach before responding by either aggression or avoidance [2].

Fourteen groups of 22 (randomly distributed) weaned rabbits were housed in pens of seven different sizes (1 m x respectively 1.32 m, 1.44 m, 1.60 m, 1.81 m, 2.10 m, 2.54 m, 3.27 m). To evaluate the effect of environmental enrichment, each pen size was provided in a barren and an enriched form. The enrichment, which consisted of two parallel wooden boards offering an opportunity to hide, gnaw and scratch, was placed in the centre of the pen in order not to affect the distribution pattern of the rabbits. For this reason all pens were also equipped with two feeders (placed at the short sides) and two drinking-lines (placed at the longitudinal sides) (Figure 1).

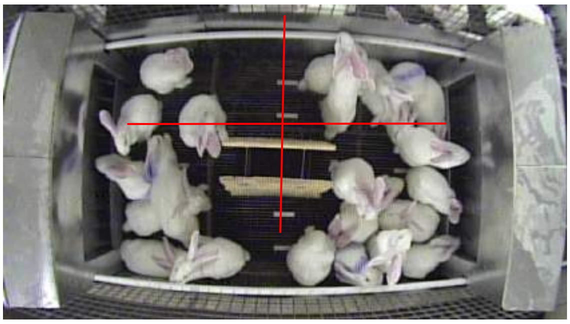


Figure 1. *Enriched pen (1 x 1.81 m). In order not to affect the distribution the enrichment was placed in the centre of the pen and (also for the barren pens) two feeders and two drinking-lines were placed at the sides. A longitudinal and transverse line was drawn on each photograph to determine the distribution pattern of the rabbits.*

During the fattening period (age four – ten weeks) photographs were made every 30 min, for the duration of 48 hours per pen per week, using a digital camera installed on an automatically controlled rail system (Figure 2).
The spatial distribution of the rabbits in the pens was determined by subdividing each pen in four imaginary identical parts, i.e. by drawing a longitudinal and transverse line on each photograph (Figure 1).

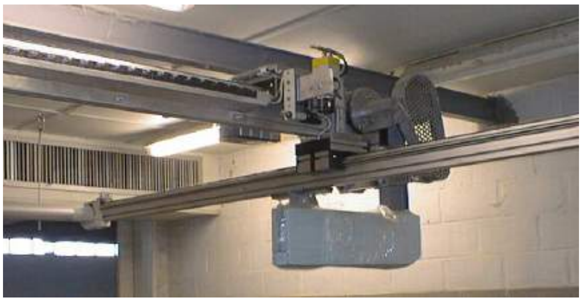


Figure 2. Digital camera installed on an automatically controlled rail system.

Subsequently, for each photograph the number of animals per subdivision was recorded and the standard deviation was calculated. To obtain the heterogeneity index the standard deviation was transformed to a value between 0 and 1 so it became independent of the number of animals in the pen (group size may differ slightly between pens because of mortality).

3 Results and discussion

The heterogeneity index increased with pen size ($F = 8.82$, $P < 0.01$) and the presence of enrichment ($F = 43.87$, $P < 0.001$). Rabbit age negatively affected the index ($F = 910.62$, $P < 0.001$) but this effect decreased with pen size ($F = 83.69$, $P < 0.001$) and the absence of enrichment ($F = 8.48$, $P < 0.01$). As the frustration of the rabbits, caused by the environment, is expected to increase with body size (and hence age) and with stocking density (both factors reinforcing one another), our heterogeneity index seems a useful tool for assessing the spatial requirements of group-housed animals. If so, our results indicate that the spatial requirements of rabbits depend on the quality of their environment: a smaller pen size (and accompanying higher stocking density) may be compensated to some extent by providing environmental enrichment.

This developed distribution method seems efficient for determining the minimal space requirements of group-housed animals but ought to be validated and refined. For example, the dominance hierarchy and behavior was not taken into account whereas both aspects might have a marked influence on the distribution of group-housed animals [2].

References

1. Keeling, L.; Jensen, P. (2002). Behavioural disturbances, stress and welfare. *The Ethology of Domestic Animals. An Introductory Text*, 79-98.
2. Keeling, L. (1995). Spacing behaviour and an ethological approach to assessing optimum space allocations for groups of laying hens. *Applied Animal Behaviour Science*, **44**, 171-186.
3. Maertens, L.; Van Oeckel, M.J. (2001). The fattening of rabbits in pens: effects of housing and gnawing material on performance level and carcass quality. *12th Symposium on Housing and Diseases of Rabbits, Furbearing Animals and Pet animals*, Celle, Gemany, 9-10 may 2001.
4. Morton, D.B. (2002). Behaviour of rabbits and rodents. *The Ethology of Domestic Animals. An Introductory Text*, 193-209.
5. Podberscek, A.L.; Blackshaw, J.K.; Beattie, A.W. (1991). The behaviour of group penned and individually caged laboratory rabbits. *Applied Animal Behaviour Science*, **28**, 353-363.
6. Rommers, J.; Meijerhof, R. (1998). Effect of group size on performance, bone strength and skin lesions of meat rabbits housed under commercial conditions. *World Rabbit Science*, **6**, 299-302.
7. Stauffacher, M. (1992). Group housing and enrichment cages for breeding, fattening and laboratory rabbits. *Animal Welfare*, **1**, 105-125.
8. Verga, M. (2000). Intensive rabbit breeding and welfare: development of research, trends and applications. *Proceedings 7th World Rabbit Science Congress. In: World Rabbit* Science, **8 (S1)**, 491-509.

A real-time video analysis system for automatic analysis of rat behavior

Zhang M, Zhang H.Y, X.X Zheng
Key laboratory of Biomedical Engineering of Ministry of Education
College of Biomedical Engineering & Instrumental Science, Zhejiang University, Hangzhou, Zhejiang,China

Abstract

Behavior analysis has become an essential part of neural science. Methods and technologies of behavioral analysis have progressed from direct observation, transducer-detection to the most versatile method - video tracking technology. However, current video tracking systems are based only on locomotory analysis, with the posture information ignored. As a result, detailed ethogram can't be acquired. In this paper, a novel method for automatic posture recognition was presented. Algorithms of posture classification and locomotory analysis were combined to perform behavior recognition. An 8-pattern ethogram was used to describe rat behavior, i.e., groom, stretched attend, stationary rotation, sit, rear, walk, rotation, fast walk. Based on the algorithms, we developed an automatic behavior analysis system for real-time description of ethogram. The advantages of the system are that it can do more in-depth analysis and even novel parameter analysis using postures and posture changes of animals, and also it can give a real-time description of detailed ethogram.

Keywords

Ethogram, Posture, Behavior, Real time, Automatic Analysis System

1 Introduction

Assessment of animal behavioral effects is widely used in researches of central nervous functions, such as motorial function, learning and memory. It has become an essential part of neural science, especially neural physiology and ethopharmacology etc. Traditionally, experimenters observe animal directly and record the behavior manually or using an event-recording program [4]. For some behaviors, manual recording of behavior may be the only way to detect and record their occurrence. However, directly observation is time-consuming, subjective, arbitrary, and also with low investment. As a result, Methods and techniques of behavioral analysis have evolved dramatically in the past decade. Automatic behavior analysis methods have progressed from hard-ware detection to video tracking technology [6]. Recently, digital image processing technique has been applied to tracking the position and path of an animal, and also further locomotory parameters such as distance moved, average speed. Though this technique has more advantages than former methods, it is based only on analysis of locomotory patterns with the animal represented as spatial coordinates of its barycenter [1-3]. However, behavior consists of postures and changes of postures (including changes of positions). As an essential part of behavior, posture contains more psychological information. Consequently, recognition of animal postures and detailed description of animal behavior have become an interesting research area in ethology, especially in behavior researches of rodents.

In this paper, we presented a robust automatic posture analysis method. Based on the algorithms, we developed a novel behavior analysis system named RTEtho, which could give a real time and more detailed description and online analysis of rat behavior.

2 Algorithms of posture recognition

2.1 Feature selection

Rat postures were categorized into five classes for empirical reason, i.e., groom, stretched attend, stationary rotation, sit, and rear (in Table 1). During experiment, rats can move freely in one area, which may bring its orientation and position changes of images constantly. As a result, rat posture classes have multiformity. i.e., in one posture class, there may be many sub-classes according to our observation. Also, in real application, rat size variety and camera focus changes may introduce resolution (DPI: Dot per Inch) changes of images. Relying on this, feature of the shape of rat image boundaries should be rotation, translation and scale invariant (RTS). In order to be robust in these changes, four "cursory" features of rat contour were extracted, which were also proved to be RTS invariant [7]. The features were "cursory", which result in high immunity to intra-class multiformity.

Among all the descriptive features of a rat contour, we selected four effective ones.

The four features are:

(1) $X_1 = \frac{1}{N}\sqrt{\frac{(d_i - \bar{d})^2}{N-1}}$

Where, d_i is the distance from the i[th] pixel of a contour to the contour's barycenter; $\bar{d}$ is the average distance of d_i ; N is the number of pixels along the contour.

(2) $X_2 = \frac{d_{min}}{d_{max}}$

Where, d_{max}, d_{min} are the maximal and the minimal d_i respectively.

(3) $X_3 = \arccos\frac{<\vec{d}_{max}, \vec{d}_{min}>}{\|\vec{d}_{max}\| \cdot \|\vec{d}_{min}\|}$

Where, $\vec{d}_{max}, \vec{d}_{min}$ are vectors corresponding to the maximal and minimal d_i respectively.

(4) $X_4 = \frac{M}{N}$

Where, M is the number of the pixels along the concave segment of contour.

2.2 Posture Classification

We need categorize the posture of a rat in current frame into the five defined classes with the above vector. Here we set up a Naive Bayes Classifier.

A Naive Bayes classifier is a probabilistic classifier in which the features are assumed conditionally independent of each other given the class. Naive-Bayes classifiers have a surprisingly very good record in many classification problems, such as text classification [3]. Although the Naive-Bayes model has a much simpler structure than other models, it is still observed to be successful as a classifier in practice.

Consider a classification problem with $y \in \{0,1,2,\ldots,m\}$ (class label) and $X \in R^n$ (feature vector) the observed data, the classification problem under the maximum likelihood framework (ML) can be formulated as:

$$\hat{y} = \arg\max_{y} P(X \mid y) \qquad (1)$$

If the features in X are assumed to be independent of each other conditioned upon the class label (the Naive Bayes framework), equation (1) reduces to:

$$\hat{y} = \arg\max_{y} \prod_{i=1}^{n} P(X_i \mid y) \qquad (2)$$

Learning of parameters of the model is a probability density estimation problem. Because of the assumption of the conditional independence, we need to learn the following $n+1$ prior probabilities from the data: $P(y), P(X_1 \mid y), \ldots P(X_n \mid y)$. For a continuous density estimation problem, $P(X_i \mid y)$ is a value at certain point of a probability density function. In practice, the common assumption is that we have a Gaussian distribution. Here we use the 1-D kernel density estimation. A Gaussian kernel smoothing function was used to estimate the density. Figure. 1 shows the probability density estimation functions of X_1 in the first four defined posture classes.

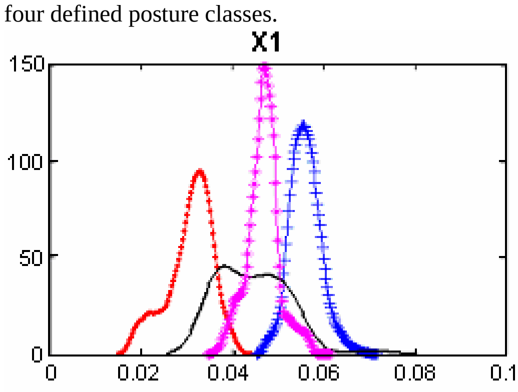


Figure 1*. Probability Density Estimation of X_1 in the Four Defined Posture Classes*

2.3 Behavior Classification

We use an 8-pattern ethogram to describe rat behavior, and each pattern was labeled with a number: 1-Groom, 2-Stretched attend, 3-Stationary rotation, 4-Sit, 5-Rear, 6-Rotate, 7-Walk, 8-Fast Walk. For each image, the ethogram pattern was recognized according to the animal's postures in current image and in a few prior images.

Postures	Description
Groom	The rat licks and rubs its paws or its body.
Stretched attend	The body of the rat is stretched and immobile.
Stationary Rotation	The rat rotates the front part of the body without moving the hind legs.
Sit	The rat is immobile.
Rear	The rat stands on its hind legs, and the back is straight.

Table 1. Typical postures.

With the ethogram, a lot of parameters can be calculated, such as total distance, mean speed, percent time in mid-section of the open-field, percent time of still, walk, fast walk, groom, stretched attend, stationary rotation, sit, rearing, and groom times, rearing times, rotation times etc.

3 Real-time behavior analysis system

3.1 System structure and advantages

Implementing the behavior recognition algorithms mentioned above, a new behavior analysis system was developed. The whole system consists of a subject container (Open-field box, etc.), digital video-capture device, a PC computer and the behavior analysis software we developed.

The software consists of six main modules: info-input module, parameter-setting module, data-sampling-analysis module, info-output module, data-view module, and off-line analysis module.

3.2 Info-input module

This module is for management of all input information, including experiment information and rat information, such as experiment directory, experiment name and group, rat weight, sex, species, age, and comments etc.

3.3 Parameter-setting module

This module is for users to set experimental parameters before experiment starts. Parameters can be FPS (frame per second), experiment time, arenas & zones etc.

3.4 Data-sampling-analysis module

When the experiment starts, the video of rat behavior is displayed in an active video window directly. Every frame of the video is processed and showed in another window. Synchronously, rat behavior is recognized using the algorithms mentioned above. With the experiment going on, tracking path, ethogram can be showed in the corresponding window in real time dynamically. (Figure 2)

With the system, we can analysis ethogram or ethogram sub-sections online. Statistical analysis method and sequence analysis were introduced into the system.

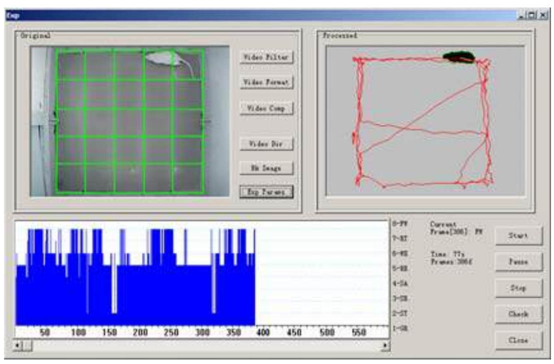


Figure 2. *Real-time Ethogram showing*

3.5 Info-output module

When an experiment ends, the system can output experiment logs and analysis results in self-defined format, such as *.ini.

3.6 Data-view module

This module allows users to view locomotory analysis results, ethogram analysis results, and experiment logs etc. Also, results from different experiments can be compared.

3.7 Off-line analysis module

This module is a special part of the system, which allows users to do offline analysis on the recorded video or image sequence, then output experiment results, logs and etc. This function gave experimenters an opportunity to observe the experiment process and calculate the result once again.

4 Conclusion

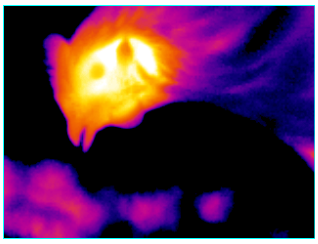
Compared with some existing systems based on locomotion analysis, this system can do more in-depth analysis and even novel parameter analysis using postures and posture changes of animals. And also, it can give a real-time description of ethogram. Since this software is compatible with USB video camera, it sharply reduces the hardware cost while still keeps the performance high.

Supported by National 863 Program of China (Grant N. 2003AA2Z3537), Key Laboratory of Biomedical Engineering of Ministry Education of China (Zhejiang University), Zhejiang Provincial Key Laboratory of Chinese Medicine Screening, Exploitation & Medicinal Effectiveness Appraise for Cardio-cerebral Vascular & Nervous System.

References:

1. Bonatz, A. E., Steiner, H., & Huston, J. P. (1995). Video image analysis of behavior by microcomputer: Categorization of turning and locomotion after 6-OHDA injection into the substantia nigra. *Journal of Neuroscience Methods*, **22**, 13-26.
2. Dai, H., & Carey, R. J. (1994). A new method to quantify behavioral attention to a stimulus object in a modified open-field. *Journal of Neuroscience Methods*, 53, 29-34.
3. Dumais S.T., Platt J., et.al. (1998) Inductive learning algorithms and representations for text categorization. *Proceedings of the 7th International Conference on Information and Knowledge Management*, pp. 148-155.
4. Lucas P.J.J. Noldus, Andrew J. Spink et.al. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers*. 33 (3), 398-414
5. Kruk, M.R., (1997) Measuring behavior into the twenty-first century, *Trends in NeuroSciences*, vol.5, pp. 187-189, 1997.
6. Noldus, L.P.J.J., (1991). The Observer: a software system for collection and analysis of observational data. *Behavior Research Methods, Instruments, and Computers* 23, 415–429.
7. Zhang M., Zhang H, Zheng X, "Automatic recognition of rat's postures based on Naive Bayes classifier," *Space Medicine & Medical Engineering*, in press

The use of thermal imaging to score plumage condition in laying hens (*Gallus gallus domesticus*)



S.J. Pope, P.H. Zimmerman, E. Glen, S. Lambton, C.J. Nicol
Department of Clinical Veterinary Science, University of Bristol, Langford, United Kingdom

Abstract
The feather coverage of laying hens can decrease due to birds pecking the feathers of their companions. This can become a major welfare issue. Also, various studies have shown that the loss of feathers can lead to substantial increases in feed intake and compromise the thermo-regulatory ability of birds. At present, methods used for scoring the plumage condition of hens are primarily subjective, and are scored on a subjective ordinal scale for the feather condition of different parts of the body. In order to quantify this we have developed a method to determine how the thermal output of individual birds varies with feather cover. Using a FLIR E4 camera we have shown that areas that have been de-feathered have an increased temperature output compared with feathered areas, and can quantify feather damage. There are also indications that the method could be used to carry out large-scale flock analysis to identify 'hot spots' of feather pecking.

Keywords
Feathers, Thermography, Poultry

Introduction
At present the most common method used for scoring the plumage conditions of laying hens is a subjective technique. For example, Tauson et al. [1] classified birds into 4 main categories.

Birds were scored for feather condition of five individual parts of the body, i.e. neck, breast, back, wings and tail according to a scale: Score 4 = a very well feathered body part with no or few worn or otherwise deformed feathers. Score 3 = a body part where feathers had deteriorated, but where the body was still completely covered. Score 2 = a body part where feathers had clearly deteriorated and/or with larger naked areas. Score 1 = a body part with heavily damaged plumage with no or only a very small areas of the body covered with feathers.

This method of categorizing provides the best technique at the moment for quantifying plumage cover. It is however a non-quantitative method of analysis and so depends upon a subjective categorization of the perceived feather cover. At present there is no known truly quantitative method for assessing the plumage cover of laying hens.

Various studies have shown that the loss of feathers can lead to substantial increases in heat loss [2] and feed intake [3]. Poor feather cover could also lead to poor feed efficiency [4]. There is also a marked effect on thermoregulatory capacity in relation to the plumage cover of a bird. Nichelmann et al. [5] showed that the maximum thermoregulatory increase in heat production (HP) of defeathered laying hens exceeds that of the feathered ones at thermo neutral temperatures by 238%. All of the above studies have been carried out in metabolic chambers and have provided valuable information about the physiological effects of feather loss in laying hens. They do not however indicate the thermal loss experienced by laying hens in commercial conditions. The new technology of infrared thermography could provide an indicator to the thermal output of laying hens in these conditions and could also be developed as a tool to quantify the effects of different levels of feather cover and its affect on thermal regulation.

Materials and Methods
All of the birds used in this study laying hens (*Gallus gallus domesticus*) from a free-range system. The birds were sampled between 41 and 72 weeks of age. The birds were sampled at random from 6 different areas within the house and were visited on 4 times during the laying cycle. Each bird was held while a second experimenter took thermal images and video footage. In order to further validate the readings, the birds were placed in a container for a period of 5 minutes and second thermal images were taken of the same bird before it was released.

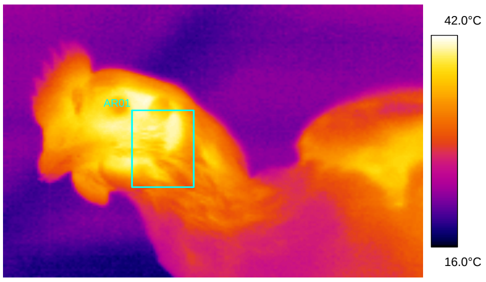
Each of the birds were feather scored according to Tauson et al.'s (1984) method by two people independently on-site. The same birds were then scored by another 5 people using video footage taken on farm. Pictures were also taken of the four body parts using an FLIR E2 thermal imaging camera, which measures and images the emitted infrared radiation from a target object. In order to further validate the procedure the measures were repeated. After each bird was feather scored and a thermal image was taken the birds were placed in a holding pen for 5 minutes, and then another thermal image was taken, the two images could then be tested for their repeatability.

The fact that the radiation is a function of the target surface temperature makes it possible to calculate and display this temperature. Various object parameters must be corrected in order to get a true reading. The following object parameters were supplied for this correction to work; Emissivity, Ambient temperature, Atmospheric temperature (Often equal to the ambient temperature.), Distance, and the Relative Humidity of the air.

The emissivity is the amount of radiation coming from an object compared to that of a blackbody and was on a setting of 0.95 [6]. The atmospheric temperature was recorded using in-house thermometers. The distance was kept at a constant 1.5m using a fixed measure. Because of the short measuring distance that relative humidity was left at the default setting of 50%.

The camera took radiometric JPEG images, which could then be analyzed using ThermaCAM^{tm} Reporter 2000 Professional software. This software enabled the images to be analyzed in various ways. For the purposes of this study each of the images was measured with a fixed area, which gave the minimum, maximum, and average temperature readings.

The range at which the temperatures were analyzed was fixed between 16^{o} – 42^{o}.



IR information	***Value***
File name	***Ir_0346.jpg***
Label Value
AR01 : max	***38.0°C***
AR01 : min	***27.9°C***
AR01 : avg	***33.7°C***

Figure 1 *ThermaCAM[tm] Reporter 2000 of head and neck using fixed area measure.*

Results

The thermographic temperature readings and the feather score measures were compared for each of the five different regions of the birds' bodies. Figures 2 shows that the higher the feather scores the lower the temperature recorded. This relationship was repeated for all five regions of the birds' bodies examined.

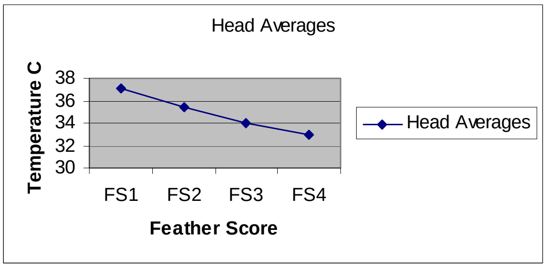


Figure 2. *Total head average temperature against feather score.*

A similar relationship was seen during the 4 different visits. As expected within normal commercial conditions, the feather scores per bird worsened through the laying cycle. Figure 3 shows that correspondingly, the thermographic temperature increased through the laying cycle.

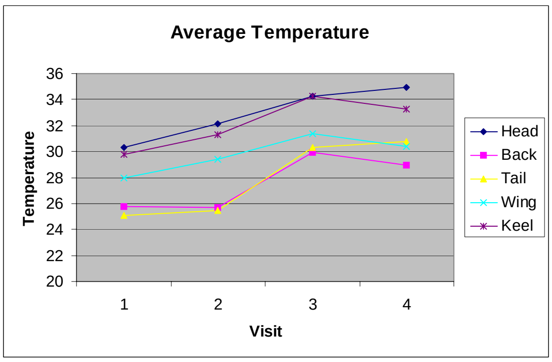


Figure 3. *Average temperature against over the 4 visits.*

Further statistical analysis was carried out to compare the repeatability of the thermographic measures (Table 1). The relationship between the feather scores and the temperatures recorded on each of the five different body regions was then compared (Table 2).

Table 1. *Repeatability measures using Pearson correlation.*

	R-Sq	P
Head Temperature	0.75	<0.001
Back Temperature	0.94	<0.001
Tail Temperature	0.90	<0.001
Keel Temperature	0.54	0.006
Wing Temperature	0.40*	0.047

Table 2. *Polynomial regression analysis of body temperature against feather score.*

	R-Sq	P
Head Temperature	0.41	<0.001
Back Temperature	0.35	<0.001
Tail Temperature	0.43	<0.001
Keel Temperature	0.35	<0.001
Wing Temperature	0.06*	<0.001

** We only had data for feather scores 3 and 4*

Conclusion

The development of thermal imaging technology and the cost of the equipment now mean that it is possible to carry out assessments of animals in on-farm situations and obtain, with the use of strict protocols, quantifiable measures of plumage damage. The results show that the loss of feathers has a highly significant effect on the thermoregulation of individual birds. The increase in heat loss associated with the loss of feathers was, in some cases, likely to impose severe metabolic difficulties in sustaining core body temperature. Difficulties in thermoregulation will also affect bird behavior and welfare. Further research is now studying if this technique can be used on larger numbers of birds in commercial flocks to identify when significant levels of feather damage occur.

References:

1. Tauson R, Ambrosen T, and Elwinger K., 1989. Evaluation of procedures for scoring the integument of laying Hens-Independent scoring of plumage condition. *Acta Agric. Scand.* **34**, 400-408.
2. Richards S.D., The influence of loss plumage on temperature regulation in laying hens. *J. Agric. Sci.*, **89**, 393-398.
3. Emmans G.C., and Charles D.R, Climatic environment and poultry feeding in practice. In: *Nutrition and climatic environment*, (eds.) W. Haresign, H. Swan and D Lewis, Butterworths London, England.
4. Lee B.D, Leeson S, and Morrison, W.D. 1983. Effects of feather cover and insulative jackets on metabolic rate of laying hens. *Poultry Science*, **62**, 1129-1132.
5. Nichelmann M, Baranyiova E, Goll R, and Tzschentke B. 1986. Influence of feather cover on heat balance in laying hens (*Gallus domesticus*). *Journal of Thermal Biology* **11(2)**, 121-126.
6. Holman, J.P., 1986. *Heat transfer*. McGraw Hill, New York.

Combined action of uranium and stress in the rat. Behavioral effects

M. Bellés [a,b], M.L. Albina [a,b], V. Linares [a,b], M. Gómez [a], D.J. Sánchez [a,b], J.L. Domingo [a]

[a] *Laboratory of Toxicology and Environmental Health, School of Medicine, "Rovira i Virgili" University, San Lorenzo 21, 43201 Reus*

[b] *Physiology Unit, School of Medicine, "Rovira i Virgili" University, San Lorenzo 21, 43201 Reus*

Abstract

The influence of restraint stress on uranium (U)-induced behavioral effects was assessed in adult male rats. Eight groups of animals received uranyl acetate dihydrate (UAD) in the drinking water at doses of 0, 10, 20 and 40 mg/kg/day during 3 months. Rats in four groups were concurrently subjected to restraint during 2 h per day throughout the study. At the end of the experiment, the following behavioral tests were carried out: open-field activity, passive avoidance, and Morris water maze. Uranium concentrations in the brain were determined. Restraint stress did not significantly affect the uranium levels in the brain. At 20 and 10 mg/kg/day restraint significantly affected the total distance traveled in the open-field during the first and third periods respectively, while no significant differences between groups were observed on the passive avoidance test. In the Morris water maze test, the influence of restraint was significant on the latency time on Day 3 at 10 mg/kg/day. Although the results of the present study scarcely show uranium-induced behavioral effects at the oral doses of UAD here administered, these effects, as well as the slight influence of restraint stress noted in some tests cannot be underrated.

Keywords

Uranium, restraint, open-field, passive avoidance, Morris water maze test.

1 Introduction

In recent years the use of depleted uranium (DU) in munitions (Persian Gulf War, Kosovo/Balkans, Iraq) has given rise to a new exposure route for this hazardous heavy metal (1). Experimental studies on the neurobehavioral toxicity of uranium have been particularly scarce, especially taking into account that a significant distribution of uranium in the brain after chronic exposure to the metal has been demonstrated (2). Neurobehavioral changes, together with physiological and biochemical alterations, are a common response to stress (3). Information on the combined effects of simultaneous exposure to metals and stress is rather scarce. However, it is evident that humans can be potentially exposed, either occupationally, and/or environmentally, to different levels of metals (uranium for example), while they can also be concurrently subjected to various types of stress. It may be especially evident in specifically grave situations such as some recent wars. Among the animal models to examine the effects of stress on the toxicity of a chemical, restraint was selected as it has been widely used in a number of experimental studies (4).

2 Methods

2.1 Animals and treatment

Adult male Sprague-Dawley rats (220-240 g) were randomly divided into 4 groups of 16 animals per group, which were in turn divided into 2 subgroups of 8 rats. Animals in 4 subgroups received uranium as UAD at 0, 10, 20 and 40 mg/kg/day, while rats in the four remaining subgroups received the same UAD doses concurrently with restraint stress throughout the study. Restraint was daily administered for 2 h by placing the animals in metacrilate cylindrical holders (Letica Scientific Instruments, Barcelona, Spain) (5). The uranium doses were approximately equal to 1/20, 1/10 and 1/5 of the oral LD50 of UAD in adult rats (6).

2.2 Experiment

At the end of the period of UAD exposure, the following behavioral tests were carried out: open–field activity, passive avoidance and Morris water maze test. At the end of the behavioral testing, samples of the brain were taken to determine the cerebral content of uranium.

2.3 Behavioral tests

2.3.1. Open-field activity. General motor activity was measured in an open-field apparatus. Rats were allowed to move freely around the open-field and to explore the environment for 15 min, divided into 3 sessions of 5 min each. The path of the animals was recorded by a video camera (Sony CCD-IRIS model) that was placed above the square and was connected to a VHS videocassette recorder (Panasonic AG-5700 model). The video tracking program EthoVision® from Noldus Information Technology (Wageningen, The Netherlands) was used to measure the total distance traveled and the number of rearings as a measure of vertical activity.

2.3.2. Passive avoidance. The recent memory of animals was tested in a passive avoidance conditioning task. The apparatus consisted of a shuttle box (Ugo Basile model 7550, Comerio, Italy) equipped with a door to restrict access between equal-sized illuminated and dark compartments. In the acquisition trial, rats were individually placed in the illuminated compartment. After 30 seconds the door separating the two compartments was opened. Following some seconds (T1), the animal entered the dark compartment. The door was shut 1 second after the crossing, and the rat was given a 1mA/3 seconds duration foot shock. Twenty-four h later (retention trial), the same procedure was repeated without delay period to open the door and without electric foot shock. The time elapsed to enter the dark compartment was recorded as T2. A latency of entry greater than 5 min was the criterion for successful learning.

2.3.3. Morris water maze test. The movements of the animal in a tank (160 cm diameter and 60 cm height) were monitored with a video tracking system (EthoVision®, Noldus Information Technology, Wageningen, The Netherlands).

Reference memory test. The animals were tested in blocks of five trials during three consecutive days (each trial started from one of four points assigned on different arbitrary quadrants of the circular tank). The total distance traveled and the latency time before reaching the platform were measured and recorded. The maximum trial latency to reach the platform was 60 seconds. The rat climbed the platform (or the rat was placed on the platform if the animal did not find it in 60 seconds), once there remained for 30 seconds and then was returned to the cage. The intertrial interval was 60 seconds.

Probe test. At the fourth day, a single probe trial was conducted. The platform was removed from the pool and each rat was allowed to swim for 60 seconds in the maze. The time and the distance that the animal swam in the quadrant in which the platform had been located, was recorded.

2.4 Uranium analysis

Samples of the brain were taken to determine the cerebral content of uranium. Uranium concentrations were determined by inductively coupled plasma-mass spectrometry (Perkin Elmer, Elan 6000), according to previously reported methods.

3 Results

With respect to uranium, no experimental data concerning the influence of stress on the potential neurobehavioral toxicity of uranium have been reported. In our study, restraint stress did not significantly affect the uranium levels in the brain of rats (Table 1). However, at 20 and 10 mg/kg/day restraint significantly affected the total distance traveled in the open-field during the first and third periods tested, respectively (Figure. 1). While no significant differences between groups were observed on the passive avoidance test. In the Morris water maze test (Figures. 2 and 3), the influence of restraint was only significant on the latency time measured on day 3 in rats exposed at 10 mg/kg/day (Figure. 3).

Table 1. Uranium concentrations in the brain of rats exposed to uranium and/or restraint stress.

Groups of rats	Uranium concentrations
Unrestrained control	ND^{a}
Restrained control	ND^{a}
10 mg/kg/day	4.97±2.35^{bc}
10 mg/kg/day plus restraint	3.58±1.51^{b}
20 mg/kg/day	2.32±1.39^{a}
20 mg/kg/day plus restraint	4.88±1.42^{b}
40 mg/kg/day	13.89±9.37^{c}
40 mg/kg/day plus restraint	4.96±2.12^{bc}

Results are given in ng U/g fresh tissue and expressed as means ±SD. ND: not detected; detection limit: 0.005 ng U/g.

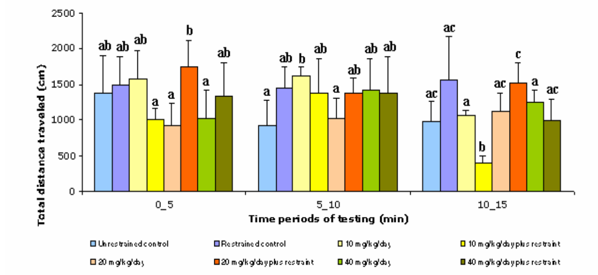


Figure 1. *Total distance traveled in an open-field.*

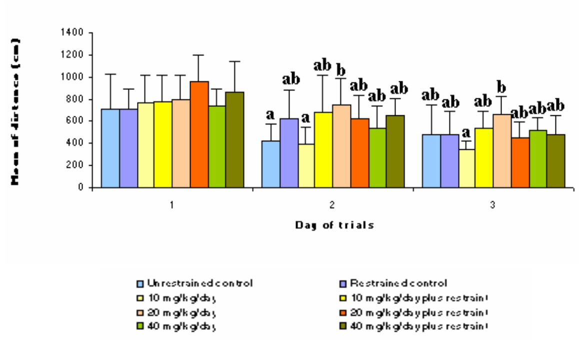
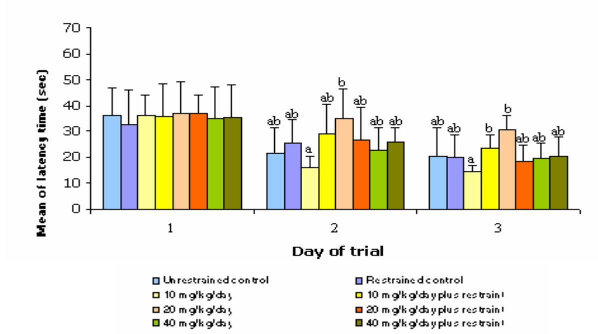


Figure 2. *Effects on the mean distance traveled in a water maze.*



References:

1. McDiarmid, M.A.; Keogh, J.P.; Hooper, F.J.; McPhaul, K.; Squibb, K.; Kane, R.; Dipino, R.; Kabat, M.; Kaup, B.; Anderson, L.; Hoover, D.; Brown, L.; Hamilton, M.; Jacobson-Kram, D.; Burows, B.; Walsh, M. (2000). Health effects of depleted uranium on exposed Gulf War veterans. *Environ Res.*, **82**, 166-180.
2. Lemercier, V.; Millot, X.; Ansoborlo, E.; Ménétrier, F.; Flüry-Hérard, A.; Rousselle, Ch.; Scherrmann, J.M. (2003). Study of uranium transfer across the blood-brain barrier. *Radiat. Prot. Dosimetry*, **105**, 243-245.
3. Vogel, W.H. (1993). The effect of stress on toxicological investigations. *Hum. Exp. Toxicol.*, **12**, 265-271.
4. Albina, M.L.; Bellés, M.; Gómez, M.; Sánchez, D.J.; Domingo, J.L. (2003). Influence of maternal stress on uranium-induced developmental toxicity in rats. *Exp. Biol. Med.*, **28**, 1072-1077.
5. Colomina, M.T.; Albina, M.L.; Domingo, J.L.; Corbella. (1997). Influence of maternal stress on the effects of prenatal exposure to methyl mercury and arsenic on postnatal development and behavior in mice: a preliminary evaluation. *Physiol. Behav.*, **61**, 455-459.
6. Domingo, J.L.; Llobet, J.M.; Tomas, J.M.; Corbella, J. (1987). Acute toxicity of uranium in rats and mice. *Bull. Environ. Cotam. Toxicol.*, **39**, 168-174.

Individual differences in fear-related reactivity in transgenic mice carrying APP and PS1 mutation - random variable, hearing impairment or a personality trait?

H. Iivonen, E. Koivisto, H. Tanila
Dept. Neuroscience and Neurology, Univ. Kuopio, Kuopio, Finland.

Abstract

Alzheimer's disease is associated with a variety of behavioral symptoms in addition to cognitive impairment, and these behavioral symptoms are usually the primary cause for institutionalization of the patients. We have used transgenic mice carrying mutated human amyloid precursor protein (APP) and presenilin-1 (PS-1) genes as a model for the disease. These animals develop amyloid plaques from 5-months of age on and show progressive impairment in their spatial learning ability. We used acoustic startle response (ASR) and pre-pulse inhibition (PPI) as a measure to the innate reactivity of these mice in comparison to their negative littermates. We tested these animals at 4 and 12 months of age, and expected the reactivity to dramatically attenuate with age, as these animals had C57BL/6J background with reported age-associated hearing loss. Unexpectedly, we came across a huge individual variability in the startle responses in both genotypes and age groups. In order to clarify the reason for the individual differences we further studied acoustic startle response using different sound pressure and frequency levels for the pre-pulse and the main pulse. Furthermore, we tested the animals in other fear-related behaviors, the elevated plus maze and novelty suppressed feeding. The results will help the interpretation of acoustic startle responses in genetically manipulated animals with C57B6 background.

Keywords

Acoustic startle response, pre-pulse inhibition, hearing deficit.

Introduction

Transgenic mice carrying different human genes are in common use among neuropharmacological research. Most of these mutant mice have C57BL/6J strain as their background, which may cause some problems. When these mice age, they not only get the symptoms of specific neurodegenerative diseases induced by the transgene, but also other age-related weaknesses such as hearing impairment. Therefore, we tested the hearing abilities and reactivity of these mice in startle response and pre-pulse inhibition.

Methods

We used two age groups (4 and 12 months) of mice carrying human APP and PS-1 gene mutations and their non-transgenic littermates, all with C57BL/6J background. ASR and PPI were tested with TSE-startle-system (www.tse-systems.de), programmed so that the whole test lasted 67 min/mouse. The program included both ASR and PPI with varied inter-trial-intervals. In ASR as the startle stimulus we used white noise (12 kHz) of the following intensities: 70, 80, 90, 100 or 110 dB. The highest intensity of white noise was also used in the PPI test, with a pre-pulse that was a pure tone (12 kHz) with varying intensities (50, 60 and 70 dB). Background noise in all trials was 65dB. Trials were performed randomly. The startle response amplitude was adjusted to the mouse weight.
In addition, we retested a few mice with obvious hearing deficit with no background noise. Some mice responded only to the pre-pulse with 60 or 70 dB; these we tested again using a pre-pulse of lower frequency.
Other fear-related behaviors were tested in elevated plus maze and novelty suppressed feeding.

Results

All 4-month-old mice expressed the expected pattern of changes as a function of test parameters. Their startle response amplitude increased steeply with increasing intensity of the startle stimulus. Correspondingly, the PPI grew as a function of the intensity of the pre-pulse, but the inter-pulse interval had no significant effect. The 12-month-old mice show only a modest trend to respond more vigorously to higher startle stimulus intensities. Furthermore, their PPI was similar to all pre-pulse intensities, although a robust PPI was still present in all 12-month-old mice. Most of the mice in both age and genotype groups responded coherently, but a couple of obvious outliers were found in most groups. When these outliers (responses outside 2 Sd of the group mean) were omitted from the analysis, a highly significant increase in startle response amplitude and a reduction of PPI was observed in APP/PS1 transgenic mice.

Some animals that did not respond to the startle stimulus, yet startled after the presentation of the pre-pulse. This premature reactivity was eliminated by changing the pre-pulse to a lower frequency. Some animals not responding to the startle stimulus, did so when the background noise was turned off.

The APP/PS1 and wild type littermates did not differ in elevated plus maze or novelty suppressed feeding tests, speaking against increased level of anxiety as the underlying factor for increased acoustic startle response.

Conclusion

ASR and PPI are possible tests to conduct even in transgenic mice with C57BL/6J background when certain precautions are taken. First, the age-related hearing loss does not appear at the same time to all mice, and there is a substantial individual variation in the extent of hearing loss. Elimination of statistical outliers is then justified. The lower general reactivity of C57BL/6J mice as compared to most other strains should be compensated for in ASR by lowering the intensity of background noise. Also the hearing loss in these mice according to the literature [1, 2] is most robust in their most sensitive range of frequencies (3 - 9 kHz), whereas the higher frequencies are relatively well spared. Therefore these mice may show abnormal reactivity to high tones. Therefore, the startle stimulus may work best is it limited to high frequencies, whereas the pre-pulse should not combine high intensity and high frequency.

References:

1. Henry, K.R.; Chole, R.A. (1980). Genotypic Differences in Behavioral, Physiological and Anatomical Expressions of Age-Related Hearing Loss in the Laboratory Mouse. *Audiology,* **19**, 369-383.
2. Li, H-S.; Borg, E. (1991). Age-related Loss of Auditory Sensitivity in Two Mouse Genotypes. *Acta Otolaryngol (Stockh)*, **111**, 827-834.

A system for administering and measuring responses to ethanol in rats

G.A. Cousens, L.P. LaBounty, E.P. Wiertelak
Department of Psychology and Neuroscience Studies Program, Macalester College, Saint Paul, Minnesota, U.S.A.

Abstract

Here we describe efforts to develop a new method to permit examination of the effects of acute ethanol exposure and ethanol withdrawal on CO_2 sensitivity in rats. CO_2 sensitivity was determined by brief challenge with CO_2 ranging in concentration from 2.5 to 10%. It was predicted that reduced sensitivity to CO_2 would be observed immediately after ethanol exposure and that enhanced sensitivity would be observed as behavioral measures of ethanol withdrawal emerged. Consistent with the notion that CO_2 hypersensitivity is a consequence of enduring ethanol-induced pathophysiology in respiratory regulation, it was further predicted that hypersensitivity would outlast behavioral measures of ethanol withdrawal. Ultimately, we expect to determine whether CO_2 sensitivity can serve as an acute measure of anxiety in the development of an animal analogue for the study of ethanol-anxiety relationships.

Keywords

CO_2 Sensitivity, Acoustic Startle, Withdrawal

1 Introduction

Anxiety disorders and alcohol dependence occur together frequently [4]. One explanation for this co-morbidity is that alcohol dependence develops as a result of its use to self-medicate chronic anxiety symptoms; however, there is also evidence that anxiety symptoms are produced by chronic alcohol exposure [6]. Identification of potential physiological mechanisms contributing to the effects of alcohol on anxiety symptoms would further our understanding of the relationship between anxiety and alcohol dependence and may lead to novel treatments.

In humans, acute ethanol exposure is associated with reduced respiratory responses to CO_2 challenge (CO_2 hyposensitivity, [3]), and alcohol withdrawal is associated with enhanced respiratory responses to CO_2 challenge (CO_2 hypersensitivity, [4]). Methods described herein have been developed to permit examination of the effects of chronic ethanol exposure on CO_2 sensitivity in rats. The use of a rodent model allows for relative ease both in regulating exposure to ethanol and in assessing of potential physiological mechanisms underlying alterations CO_2 sensitivity.

Ss were exposed to a moderate concentration of ethanol vapor for a 21-day interval. CO_2 sensitivity was determined by challenge with varying concentrations of CO_2 prior to ethanol exposure, immediately after ethanol exposure, and at varying intervals post-ethanol. It was predicted that sensitivity to CO_2 would diminish immediately after ethanol exposure and elevate as behavioral measures of ethanol withdrawal emerged. Consistent with the notion that CO_2 hypersensitivity is a consequence of enduring ethanol-induced pathophysiology in respiratory regulation, it was further predicted that hypersensitivity would outlast behavioral measures of ethanol withdrawal.

2. Methods and Results

2.1 Subjects

To date, pilot experimental procedures have been conducted with four male Sprague-Dawley rats. Future efforts will involve use of control Ss not exposed to ethanol vapor. Ss were maintained on a 12:12hr light/dark schedule and were housed in pairs in clear Perspex cages in a temperature- and humidity-controlled vivarium until study onset. All methods were approved by the Macalester College Institutional Animal Care and Use Committee in accordance with National Institutes of Health guidelines.

2.2 Ethanol vapor exposure

Ethanol exposure was conducted using methods similar to those described by Schulteis *et al*. [7], using a custom-built airtight clear Perspex chamber (58liters) equipped with an air circulation fan. Ss were housed in the chamber in a group of four and were exposed to ethanol vapor continuously for 21days. Ethanol vapor was created by dripping 95% ethanol into a 1 liter Erlenmeyer flask heated to 100degC. Room air was passed through the flask and then to the chamber at 10liters/min. Ethanol vapor concentration can be regulated by varying the rate at which ethanol was delivered to the flask; however, for data described in the present report, the drip rate was maintained at a moderate level, 220mg ethanol per minute.

Chamber ethanol vapor concentration was measured daily by drawing chamber air with a gastight syringe, mixing it 1:4 with room air, and passing the diluted sample at a constant rate through a standard ethanol breath monitor. Ethanol vapor was maintained at a mean concentration of 14.5mg/liter air (range 11.9-16.9mg/liter) over the 21-day period. Tail blood was sampled once for each subject, 24hours prior to the end of the ethanol exposure period, and blood alcohol level (BAL) was determined using a standard NAD^+-alcohol dehydrogenase assay kit (Sigma Chemical Co., St. Louis, MO). BALs ranged from 47-67mg/dl, lower than those shown to produce overt signs of withdrawal after 21day exposure (180-220mg/dl, [7]).

2.3 Acoustic startle

Acoustic startle was assessed using a procedure adapted from that described by Culm *et al.* [2] using a Startle Monitor Behavioral Testing System (Hamilton-Kinder, Poway, CA). Ss were administered six daily startle sessions, two prior to the ethanol exposure period, one immediately after removal from the ethanol chamber, and one each at 24 and 96hours post-ethanol. Thus, only the initial post-exposure session was conducted while blood ethanol levels were elevated. All sessions were conducted between 0800 and 1400hrs. For each session, Ss were continuously exposed to 70dB white noise. Following a 15min acclimation period, 30 startle stimuli (120dB white noise; 20msec width) were presented according to a variable time 15sec schedule (range 10-20sec). In order to assess pre-pulse inhibition of the startle response (data not shown), 15 of these stimuli were preceded by 100msec by white noise pre-pulses (75, 80, or 85dB, 5 trials each; 20msec width) according to a pseudorandom trial sequence.

Previous research has revealed enhancement of acoustic startle amplitude 24hours after cessation of chronic exposure to ethanol vapor [5]. As shown in Table One, in the present study, mean startle amplitude was consistently enhanced 24hours post-ethanol relative to the pre-ethanol baseline. However, peak startle amplitude was also elevated during the immediate post-ethanol test session. Peak startle amplitude was consistently lower 96hours after removal from the ethanol vapor chamber than during the 24hour post-ethanol test.

Table 1. Peak startle amplitude (N) during a 250msec interval following each of 15 120dB white noise pulses over pre-and post-ethanol exposure sessions.

Subject	Pre-EtOH	EtOH	24hr Post-EtOH	96hr Post-EtOH
1	0.206	0.269	0.344	0.276
2	0.158	0.500	0.344	0.238
3	0.159	0.360	0.339	0.147
4	0.172	0.252	0.282	0.223
Mean	0.174	0.345	0.327	0.221

2.4 CO_2 sensitivity

CO_2 sensitivity was assessed using a custom-built dual-chamber body plethysmograph constructed according to the design of Bartlett & Tenny [1]. The apparatus consists of two identical Plexiglas chambers (25 X 18 X 15cm), an animal chamber and a reference chamber, with an intermediary pressure transducer that detects differences in pressure between the two chambers. Continuously measured pressure fluctuations related to respiration were recorded and digitized at 40kHz using a neurophysiological recording system (Recorder, Plexon, Inc., Dallas, TX) and data was saved for offline analysis.

CO_2 sensitivity was assessed during a 90min session conducted 10min after each of the six startle sessions. Ss were placed in the plethysmograph animal chamber with initial exposure to room air (9.4l/min). After a 15-30min acclimation period, Ss were exposed to CO_2 concentrations of 2.5%, 5.0%, 7.5%, and 10.0% for 10 minutes in each condition by adding medical grade CO_2 to the constant-rate room air. The varying CO_2 concentrations were presented according to a pseudorandom sequence and each was separated by an intervening 5min baseline exposure period.

Table 2. Mean tidal volume (ml/min/100g body weight) in response to CO_2 challenge over pre- and post- ethanol exposure sessions.

[CO_2]	Pre-EtOH	EtOH	24hr Post-EtOH	96hr Post-EtOH
0.0%	38.5	42.6	35.2	36.7
2.5%	35.7	34.7	41.6	37.6
5.0%	64.3	56.3	60.6	61.2
7.5%	90.3	94.5	88.1	87.2
10.0%	105.2	98.7	101.6	103.5

Tidal volume was calculated (ml/min/100g body weight) for each subject over 10sec intervals recorded toward the end of each 10min CO_2 exposure period during the final pre-ethanol test and during the immediate, 24hr, and 96hr post-ethanol tests. As depicted in Table Two, steady-state tidal volume increased with CO_2 concentration. However, no consistent alterations in tidal volume were observed across pre-ethanol and post-ethanol sessions.

3 Conclusions

Data collected to date suggests that respiration frequency and tidal volume change consistently in response to CO_2 challenge. However, these data failed to reveal any modulation of respiration during exposure to ethanol or following removal of ethanol following chronic exposure. Given that the concentration of ethanol used thus far produced lower BALs than those known to produce overt withdrawal symptoms and that withdrawal-related behavior alterations in acoustic startle were not observed, it is possible that the present Ss did not experience ethanol withdrawal. Thus, future efforts will involve steadily increasing chamber ethanol concentrations.

Acknowledgements: The authors would like to thank John Payne for assistance with data collection and analysis and Drs. Matthew Kushner and Douglas Wangensteen for discussion and use of select equipment.

References:

1. Bartlett, G, & SM Tenny (1970). Control of breathing in experimental anemia. *Respiratory Physiology,* **10**, 384-95.

2. Culm, KE, Lim, AM, Onton, JA, & Hammer, RP (2003). Reduced G_i and G_o protein function in the rat nucleus accumbens attenuates sensorimotor gating deficits. *Brain Research,* **982**, 12-18.

3. Johnston, RE, & CE Reier (1973). Acute respiratory effects of ethanol in man. *Clinical Pharmacology and Therapeutics,* **14**, 501-8.

4. Kessler, RC, Crum, RM, Warner, LA, Nelson, CB, Schulenberg, J, & Anthony, JC (1997). Lifetime co-occurrence of DSM-III-R alcohol abuse and dependence with other psychiatric disorders in the National Comorbidity Survey. *Archives of General Psychiatry,* **54**, 313-21.

5. Macey, DJ, Schulteis, G, Heinrichs, SC, & Koob, GF (1996). Time-dependent quantifieable withdrawal from ethanol in the rat: effect of method of dependence induction. *Alcohol,* **13***(2)*, 163-70.

6. Rassovsky, Y, Hulriman, E, Abrams, K, & MG Kushner (2004). CO2 hypersensitivity in recently abstinent alcohol dependent individuals: A possible mechanism underlying the high risk for anxiety disorder among alcoholics. *Anxiety Disorders,* **18**, 159-76.

7. Schulteis, G, Markou, A, Cole, M, & GF Koob (1995). Decreased brain reward produced by ethanol withdrawal. *Proceedings of the National Academy of Sciences (USA),* **92**, 5880-4.

Behavioral phenotyping of oxytocin deficient mice exposed to acute shaker stress

M. Mach [1,2], M. Morris [1], M. Dubovický [2], E. Ujházy [2], J.B. Lucot [1]
[1] *Department of Pharmacology/Toxicology, School of Medicine, Wright State University, Dayton,OH, USA*
[2] *Laboratory of Teratology, Institute of Experimental Pharmacology, Slovak Academy of Sciences, Bratislava, Slovakia*

Abstract

Reliable behavioral phenotyping methods are of great importance in understanding the influence of genetic or pharmacological interference on intact organism. Oxytocin (OT) is a stress-responsive hormone which has behavioral actions, including antidepressive, sedative and anxiolytic effects. Behavioral performance assessment was conducted after acute shaker stress in male and female OT knockout (OT -/-) and wild type (OT +/+) mice. A standard battery of behavioral tests using Hamilton-Kinder measuring system was used, including measuring anxiety (elevated plus maze), locomotor activity (open field test) and sensorimotor functions (acoustic startle response and pre-pulse inhibition). Moreover, the plasma corticosterone (Cort) level was established after stress. We conclude 1) there is a gender specific role for OT in the behavioral response to stress, and 2) OT attenuates the locomotor response to stress in males.

Keywords

oxytocin, shaker stress, behavioral phenotyping

1 Introduction

When the targeted gene is not expressed in the brain, the behavioral phenotype of the mutant mice may reveal genetic mechanisms underlying normal behaviors and may increase our knowledge of genetic factors in neuropsychiatric disorders. Knockout and transgenic mice have been developed as animal models for the study of human diseases, such as Alzheimer's disease [1], Down's syndrome [2], Huntington's disease [3], schizophrenia [4], and other neurobiological disorders [5]. In many genetically altered mice, the most noticeable difference between the transgenic mouse and its background strain is a change in behavior. Oxytocin (OT) belongs to a family of peptides that have been identified in all classes of vertebrates and many invertebrate species. OT is important for maternal behavior (labor, lactation, social interaction) [6]. Although found in about equal concentrations in both sexes, the physiological importance of OT in males is still unclear. Previous studies have suggested that OT may have important roles during stressful conditions [7]. The aim was to investigate the role of OT in the behavioral responses to acute stress, using behavioral tools for assessment of locomotor activity in the open field, elevated plus maze, acoustic startle response and levels of plasma corticosterone (Cort).

2 Material and methods

2.1 Animals

Males: OT -/- (n=20); OT +/+ (n=12)
Females: OT -/- (n=24); OT +/+ (n=12)

Colony founders were produced by W.S. Young, III and colleagues (NIH, Bethesda, MD, USA) [8]. Mice were singly housed in plastic cages with wooden shaving bedding in a temperature controlled room (T=23°C) with 12:12-h light:dark cycle (lights off from 1700).

2.2 Experimental & behavioral protocols

Two experiments were conducted. In the first experiment animals were tested without stress. In the second experiment, animals were exposed to acute shaker stress followed immediately by the behavioral tests. Each test was done once, a week apart from one another.

<u>Acute shaker stress</u>

Animals were kept in their home cages fixed to a cage rack with an automated watering system mounted on the shaker. Mice were exposed to acute shaker stress for 15 minutes. The linear horizontal excursion of the shaker was 2.86 cm at 150 rpm.

<u>Elevated Plus Maze</u>

The elevated plus maze (Hamilton-Kinder, Poway, CA, USA) was constructed from black Plexiglas. The maze consisted of two open arms (37.5×5×0.3 cm) and two closed arms (37.5×5×15 cm), which extended from the central platform (5×5 cm). The maze was elevated 63 cm above the floor. Light beam breaks were recorded and analyzed automatically, using Motor Monitor software (Hamilton-Kinder). Animals were tested in 5-minute sessions without previous handling.

<u>Open field</u>

Locomotor activity was evaluated in an automated open field system with infrared photo-beams (Motor Monitor, Version 3.11, 2000, Hamilton Kinder, Poway, CA, USA). The open field (40.6 x 40.6 cm) was divided into central and peripheral zones. The mice were placed in the center of the open field arena and the following variables of motor activity were recorded: locomotor activity, fine movement and rearing. Moreover, distance traveled, total time, rest time, number of entries and head pokes in individual zones were recorded. After the session the number of fecal pellets (defecation) was recorded for assessment of emotional reactivity and the open field arena was cleaned with 70% alcohol solution and let dry. All animals were regularly handled before individual tests in order to minimize handling-related stress.

<u>Acoustic startle response and pre-pulse inhibition</u>

Mice were tested in the SM100 Startle Monitor System Version 4.0 (Hamilton Kinder, Poway, CA, USA) for acoustic startle response (ASR) and pre-pulse inhibition (PPI). The system was programmed for 6 types of white-noise burst stimulus trials: no stimulus (background, 60 dB), pre-pulse (70 dB), pulse (100 dB and 120 dB), pre-pulse plus pulse (70 dB+100dB and 70 dB+120 dB). Each trial type was presented 10 times in 10 blocks. Stimuli were presented in random order to avoid order effects and habituation. The inter-trial interval varied from 9 to 16 sec. Mice were loosely restrained in holders placed on a sensing plate transforming movements of the body (jerks) into an analog signal through an interface. Finally, percentage pre-pulse inhibition measures were calculated as the difference between the pulse alone and the pre-pulse + pulse trials, divided by the pulse alone and multiplied by 100. Percentage scores are typically used to minimize the effect of individual variation of startle amplitude on pre-pulse inhibition.

2.3 Plasma Cort levels

Animals were decapitated 5 min. after shaker stress and plasma Cort levels were measured using radioimmuno-assay.

2.4 Statistical analysis

2-way ANOVA with two factors, condition (stress/non-stress) and genotype (OT +/+ vs. OT -/-). Software used was STATISTICA, 6.1 (StatSoft, Inc., Tulsa, OK, USA). The results are presented as a means ± S.E.M. ($p<0.05$). The exponential function $Y_{(t)} = Y_0e^{-kt}$ was used as a model of the habituation course of locomotor activity in individual animals (k-individual rate of habituation).

3 Results

3.1 Behavioral tests

Elevated plus maze

Acute stress increased the time spent in the closed arms in both genders and genotypes ($F_{Males}(1,28)=9.609$, $p<0.0044$; $F_{Females}(1,32)=14.024$, $p<0.0007$).

Open field

Acute stress changed behavior (basic movements, periphery total time, rearing) in both genders and genotypes. The interaction between stress and genotype was significant for basic movements ($F(1,28)=4.155$, $p<0.05$; Fig. 1) in males only. Further analyzing revealed changed intrasession habituation rate in males after stress (Fig. 2).

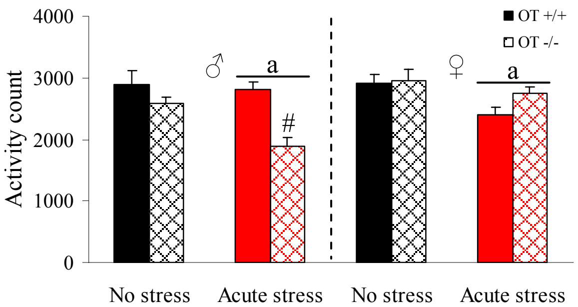


Figure 1. Locomotor activity tested in open field.
a - different from non-stress condition; # - different from OT +/+.

Acoustic Startle Response

ANOVA revealed that stress did not have any effect on the startle response to the 100 and 120 dB stimulus. Pre-pulse inhibition was also not changed. (Data not shown)

3.2 Plasma Cort levels

Acute stress increased plasma Cort in both genotypes and genders, with higher stress levels in females than males (560.3±39.9 vs. 158.1±9.2 ng/ml). No differences were observed between OT-/- and OT +/+ mice.

Supported by US Department of Defense (Grant # DAMD 17-00-C-0020) and VEGA, Bratislava, Slovakia (Grant # 2/5052/25).

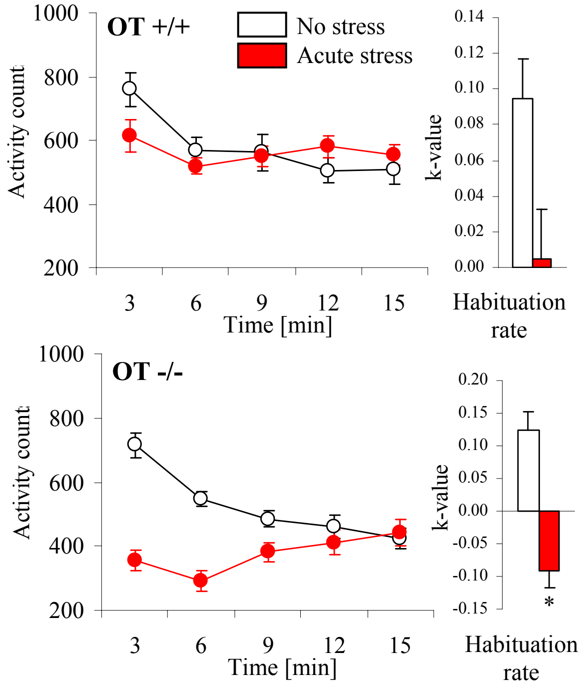


*Figure 2. Intrasession habituation of the locomotor activity in open field.*p<0.05 compared to No stress condition.*

4 Conclusions

Acute shaker stress altered the behavior of mice, as seen by the reduction in locomotor activity and habituation rate. While we predicted differences between OT +/+ and OT -/-, differences were found only in males, specifically in locomotor activity. There were no differences in females between the genotypes. The acoustic startle response and PPI did not reveal any changes after acute stress. We conclude: **1)** there is a gender specific role for OT in the behavioral response to stress; **2)** attenuation of locomotor activity was exaggerated in OT -/- males, suggesting possible involvement of OT in locomotor response to stress.

References:

1. Nalbantoglu, J. et al. (1997). Impaired learning and LTP in mice expressing carboxy terminus of the Alzheimer amyloid precursor protein. *Nature*, **29**, 500-05.
2. Schuchmann, S. et al. (1998). Altered Ca^{2+} signaling and mitochondrial deficiencies in hippocampal neurons of trisomy 16 mice: a model of Down's syndrome. *J. Neuroscience*, **18**, 7216-31.
3. Bates, G.P. et al. (1997). Transgenic models of Huntington's disease. *Hum Mol Genet*, **6**, 1633-37.
4. Mohn, A.R. et al. (1999). Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. *Cell*, **98**, 427-36.
5. Nelson, R.J.; Young, K.A. (1998). Behavior in mice with targeted disruption of single genes. *Neurosci Biobehav. Rev.*, **22**, 453-62.
6. Insel, T.R. et al. (2001). Oxytocin: who needs it? In: Russel, J.A. et al. *Prog in Brain Research*, **133**, 59-66.
7. Onaka, T. (2004). Neural pathways controlling central and peripheral oxytocin release during stress. *J. Neuroendocrinol.*, **16**, 308-12.
8. Young, W.S. 3rd; et al. (1996). Deficiency in mouse oxytocin prevents milk ejection, but not fertility or parturition. *J. Neuroendocrinol.*, **8**, 847-53.

Light deprivation differently influences spatial and stereotyped navigations in the radial arm maze

M. Salami and M. Anvari
Department of Physiology and Pharmacology, Kashan University of Medical Sciences, Kashan, I. R. Iran

Abstract
Sensory experience plays an important role in development of brain. During a critical period of early postnatal age lack of sensory inputs leads to some irreversible changes in the nervous system. In this study dark (DR) and light reared (LR) rats were trained for spatial working memory (WM) and strategy selection (SS) tasks on an eight arm radial maze. The results from the WM tests express that the DR rats make higher performances than their LR counterparts; however, the LR rats display a better improvement through the experiments. Conversely, in the SS experiments the LR animals outperformed the DR ones. The present findings demonstrate that early exposure to light deprivation differently affects spatial and stereotyped maze searching. Since, in contrast to spatial tasks, the adjacent arm tasks do not require memory it can be concluded that the visual deprivation positively underlies only the tasks require memory.

Keywords
Radial maze; Visual deprivation; Strategy selection; Working memory

1 Introduction
It is well known that the hippocampus plays a crucial role in some aspects of learning and memory. This area receives sensory inputs indirectly, converged on the entorhinal cortex [6] as well as directly [9] from the neocortex. Development of both cortical sensory systems [8] and the hippocampus [3] is influenced by environmental experience during a critical period before onset of behavioural function.

Hippocampal-cortical interactions lead to strong and persistent memories for events and their constituent elements and interrelations, together with a capacity for flexibly producing memories across a wide range of circumstances [4]. The present study explores effect of early visual deprivation on spatial working memory (WM) and strategy selection (SS) in young adult rats.

2 Materials and Methods

2.1 Apparatus
An eight arm radial maze was used in this study. In the WM experiments Plexiglas guillotine doors were located at the entrance of arms. To test the SS no door separated the arms and the central platform. Food rewards were placed in a small cup at the end of each arm.

2.2 Animals
The experiments were performed on two groups of male and female Wistar rats at 40 days of postnatal age. One group were the animals housed in a 12 hr light/dark cycle (light reared, LR) and another group were the rats reared in complete darkness since birth through end of the experiments (dark reared, DR). The animals had free access to water and food. During the experiments water was available ad libitum but the animals were fed for only one hour after the second trial.

2.3 Shaping
The animals were acclimatized to the testing room two days prior to the experiments.

Then they were subjected to the shaping and training phases during dark period of the light/dark cycle of housing. The rats were shaped with the maze in four sessions; two sessions/day. The animals were placed in the maze to obtain food rewards scattered throughout the maze (on the first day) or from the food cups (on the second day). In the WM tests, the guillotine doors were also closed and opened regularly to accustom the rats with the movements of the doors.

2.4 Experimental procedures
The rats were trained with the maze in a room with numerous visible cues. In the WM experiments the animals had to learn to visit each of eight baited arms once within a trial. Each trial started with the animal placed in the central platform with all the doors closed. After 30 s all the doors were opened simultaneously and the animal was free to choose an arm and enters. Then, all the doors except that of the selected arm were closed. When the animal explored the arm and came back into the central area the door of the arm was closed, confining the animal for 10 s. Then, all the doors were opened again and the same procedure was repeated. In the SS experiments the animal placed in the central platform and could navigate the maze freely. Generally, the rats were given two trials/day with 3-4 hours intertrial interval. Each session continued until all the baited arms were entered or 10 min had elapsed. Experiments for each animal were continued at least until two consecutive sessions in which the animal entered all 8 baited arms in 8 or maximum 9 selections (the criterion for WM task learning). Number of the errors (re-entries into the previously visited arms) within each session was considered for assessment of the WM tasks. In the SS experiments the animals' choices were measured in terms of the strategy of arm selection. The first entry was scored 0. If the second entry was an adjacent arm the animal received score 1. For next adjacent arm selections scores were 2, 3 and so forth. The results were analyzed using student t-test or analysis of variance. All the data are presented as mean ± S.E.M. A P value of less than 0.05 considered significant.

3. Results

3.1 Working memory experiments
Training the LR and DR rats for the WM task indicated that the DR group acquire the criterion in fewer trials compared to the LR group (13.37 ± 1.46 vs. 18.37 ± 0.98 trials for the DR and LR rats, respectively; $P < 0.05$). Although both groups similarly declined the error numbers

through the experiment, however, the DR rats outperformed the LR ones during the first half of the experiment.

3.2 Strategy selection experiments

The results from SS experiments reveal a marked superiority of the LR rats on their DR counterparts. The higher performance of the control group, however, was appeared from the middle of the training where they obtained scores of about 3 times than did the DR group.

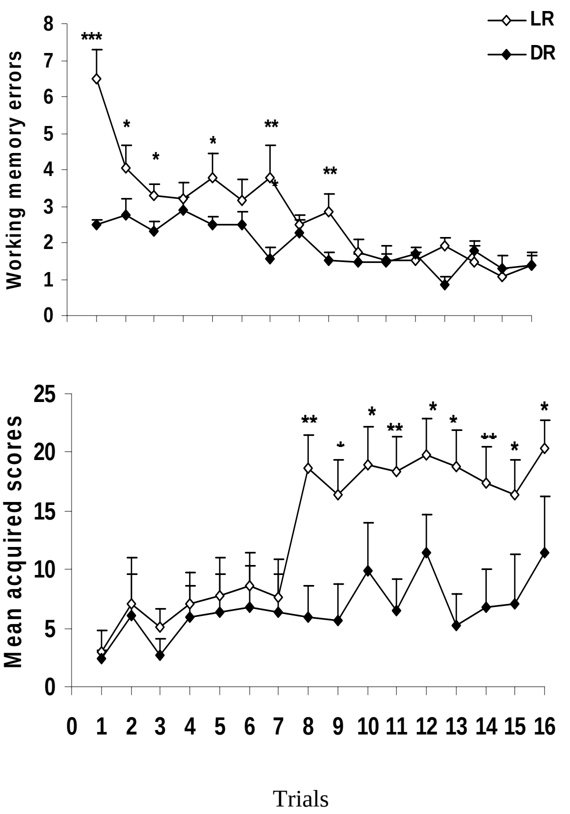


Figure 1. Graphs to show the WM errors (upper) and scores acquired from strategy of arm selection (lower) for each trial and group. The animals were subjected to 16 trials over 8 days. The upper curve indicates that early exposure to dark rearing improve the maze navigation. Instead, in the stereotypic behaviors the LR group acquired significantly further scores than the DR group as illustrated in the lower curve.
$^*P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.001$

4. Discussion

Two general strategies can be used by rats to solve the eight arm radial maze, highly variable patterns of arms or selection of adjacent arms [1]. In solving the WM tasks, our data indicate that the DR rats have more capability than the LR rats. Important point is that although the DR group begins with higher performances than the LR group, however, the latter display a better improvement through the proceeding trials so that the LR and DR rats displayed a similar positively affects spatial representations. Inconsistent results are also reported by Tees et al. where they found that the DR animals make significantly more errors than the LR animals performance in the second half of the experiments. From these results it is concluded that the visual experience may [7]. In the SS experiments, the LR rats outperformed the DR ones; indicating that the light deprivation makes the animals to be failed in performing stereotypic behaviors. It is reported that the stereotypic behavior is not produced but revealed by hippocampal lesions which remove the memory-guided behavior masking of such behavior in the intact animal [2,5]. From these considerations one possibility is that the hippocampal function is affected by the light deprivation. Another possibility could be direct influence of the sensory deprivation on the visual system so that the DR animals weakly perform the strategy selection task on basis of visual cues. Taken together, our findings express that the light deprivation differently influences the spatial and stereotypic searching; improving the former but declining the latter.

This work was supported by grant No. 8209 from research deputy of Kashan University of Medical Sciences.

References

1- Berman R.F.; Hannigan J.H. (2000). Effects of prenatal alcohol exposure on the hippocampus: spatial behavior, electrophysiology, and neuroanatomy. *Hippocampus*, **10**, 94–110.

2- Devenport L.D.; Merriman V.J. (1983). Ethanol and behavioral variability in the radial-arm maze. *Psychopharmacology* (Berl), 79, 21-24.

3- Dumas T.C., Foster T.C. (1995). Developmental increase in CA3-CA1 presynaptic function in the hippocampal slice. *Journal of Neurophysiology*, **73**, 1821-1828.

4- Eichenbaum H.; Otto T.; Cohen N.J. (1994). Two functional components of the hippocampal memory system. *Behavior and Brain Science*, **17**, 449-518.

5- Lanke J.; Mansson L.; Bjerkemo M.; Kjellstrand P. (1993). Spatial memory and stereotypic behaviour of animals in radial arm mazes. *Brain Research*, **605**, 221-228.

6- Lavenex P.; Amaral D.G. (2002). Hippocampal-neocortical interaction: a hierarchy of associativity. *Hippocampus*, **10**, 420-430.

7- Tees R.C.; Midgley G.; Nesbit, J.C. (1981). The effect of early visual experience on spatial maze learning in rats. *Developmental Psychobiology*, **14**, 425-438.

8- Waters N.S.; Klintsova A.Y.; Foster T.C. (1997). Insensitivity of the hippocampus to environmental stimulation during postnatal development. *Journal of Neuroscience*, **17**, 7967-7973.

9- Yukie M. (2000). Connections between the medial temporal cortex and the CA1 subfield of the hippocampal formation in the Japanese monkey (Macaca fuscata). *Journal of Comparative Neurology*, **423**, 282-298.

Intrahippocampal interactions of nicotine and cyclo-oxygenase-2 on spatial memory retention in the Morris water maze

M. Sharifzadeh[1*], M. Tavasoli[1], N. Naghdi[2], A. Ghanbari[1], M. Amini[1], A. Roghani[3,4]

[1]*Department. of Toxicology and Pharmacology, School of Pharmacy, Pharmaceutical Sciences and Medicinal Plants Research Centers, Tehran University of Medical Sciences, Tehran, Iran*

[2]*Department of Pharmacology, Iran Pasteur Institute, Tehran, Iran*

[3]*Departmwent of Pharmacology and Neuroscience, Texas Tech University Health Sciences Center, Lubbock, TX, USA*

[4]*Texas Institute of Environmental and Human Health, Texas Tech University, Lubbock, TX, USA*

Abstract

In the present work, we have investigated the effects of nicotine, infused in the rat dorsal hippocampus several minutes after infusion of celecoxib, on memory retention in the Morris water maze. Rats were trained for 3 days; each day included two blocks, and each block contained 4 trials. Test trials were conducted 48 h after surgery. Bilateral intrahippocampal infusion of celecoxib (0.1 M) increased escape latency and travel distance in rats, indicating significant impairment in spatial memory retention. We also examined effects of bilateral infusion of nicotine (0.5, 1.0, and 2.0 µg/side) on memory retention. Infusion of 1 µg nicotine significantly decreased escape latency and travel distance but not swimming speed, compared to controls. In separate experiments, bilateral infusion of nicotine 5 min after 0.1 M celecoxib reversed the memory deficit induced by celecoxib significantly. These results suggest the important roles of nicotine and COX-2 pathway on spatial memory retention and protected or restored the immunostaining pattern of COX-2 neurons in the rat dorsal hippocampus.

Keywords

Stereotaxic surgery; Dorsal hippocampus; Morris water maze; COX-2-containing neurons

Introduction

Cholinergic hippocampal neurons are considered to be critical for memory formation. In particular, nicotinic acetylcholine receptors (nAChR) localized on these neurons have been widely shown to be important for memory and cognition [3]. Some studies indicate that nicotine binding sites in the human brain were affected by aging or by Alzheimer's disease (AD), and that in the brains of AD patients, large reductions in the number of nicotinic binding sites have been reported [4]. According to the published data, nicotine improved cognitive function, including memory and attention, in a variety of experimental animal and human studies [3].

Cyclooxygenase (COX) catalyses the first committed step in the biosynthesis of prostaglandins (PGs), prostacyclin, and thromboxanes (TXs), which are collectively known as prostanoids [6]. COX exists as constitutive and inducible isoforms [1]. COX-1 is the constitutive form of cyclooxygenase and COX-2 is the inducible isoform [2].

Materials and methods

Subjects

Male albino Wistar rats (200-250 g), obtained from Pasteur Institute of Iran, were used.

Drugs

All the drugs used in this work were purchased from Sigma (St. Louis, MO, USA). Celecoxib and (-) nicotine were dissolved in DMSO (100%) and saline (0.9%), respectively.

Surgery

Rats were anesthetized with a mixture of ketamine (100 mg/kg, i.p.) and xylazine (25 mg/kg, i.p.) and placed in a stereotaxic instrument (Stoelting, USA). Stereotaxic surgery was done based on Paxinos and Watson's atlas coordinates [5].

Behavioral training and testing

For assessing the interactive effects of nicotine and celecoxib on memory retention, we infused nicotine (1µg/side) 5 min after celecoxib (0.1 M) into the CA1 region of the hippocampus immediately after the training trials.

Six to seven days after recovery from surgery the animals were trained for 3 days in the Morris water maze. Each training day included two blocks, and each block consisted of four trials. Each trial was initiated by placing the animals in one of the four quadrants randomly. Animals were allowed to find the hidden platform during a period of 60 sec. The rats rested 5 min between two consecutive blocks, and the inter-trial interval time was 30 sec in each block. The interval between drug infusions and testing trials was 48 hours. The testing included 1 block of 4 trials.

Results

Effects of celecoxib and DMSO on time and distance of finding hidden platform during the test trials

Post-training bilateral intrahippocampal infusion of DMSO (0.5 µl/side) into the CA1 of the hippocampus did not cause any significant changes (i.e. in escape latency, traveled distance or swimming speed) compared to saline-infused and sham-operated groups (Fig. 1). Post-training bilateral infusions of celecoxib (0.1 M) into the CA1 region of the hippocampus caused significant differences (***, $P<0.001$) in escape latency and (***, $P<0.001$) traveled distance compared to a control group (DMSO-treated animals) (Fig. 1B and 1C). However, the swimming speed was not significantly affected by the celecoxib infusion in comparison with control groups (Fig 1C).

Protective effects of nicotine against celecoxib-induced spatial memory deficits

Post-training bilateral infusion of nicotine (1 µg/side) caused improvement in escape latencies and traveled distance. Bilateral post-training intrahippocampal infusion of nicotine (1 µg) 5 min after celecoxib (0.1 M) infusion prevented the onset of spatial memory deficit normally observed after celecoxib infusion in the Morris Water Maze. The celecoxib-induced increase in escape latency and traveled distance was reversed to a level similar to the control group by nicotine infusion (Fig 2A and 2B). The swimming speed was very similar in control, celecoxib-infused and combination nicotine with celecoxib-infused groups (Fig 2C).

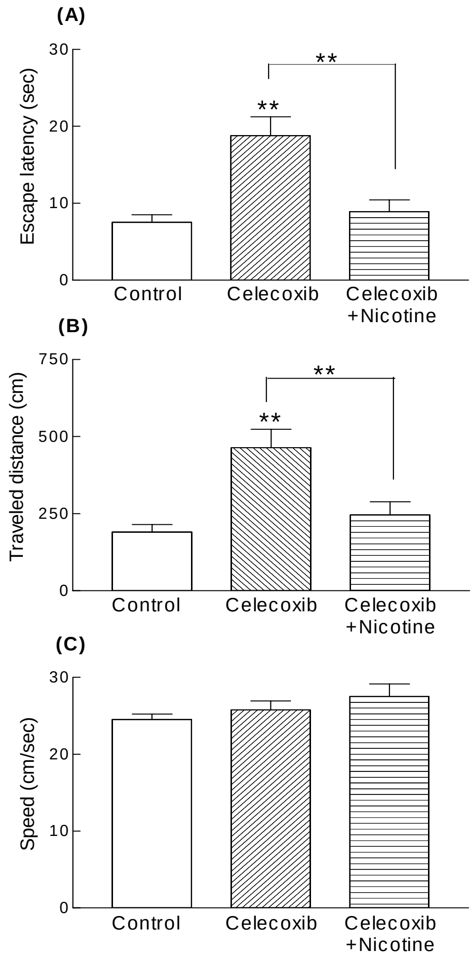


Figure 2. Protective effects of nicotine against celecoxib-induced spatial memory deficits

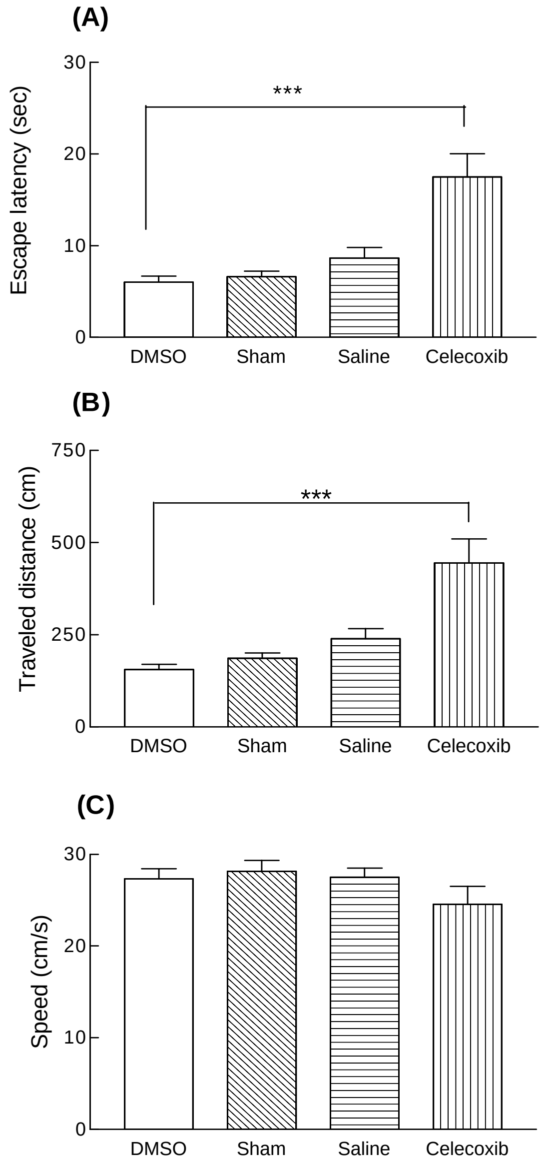


Figure 1. Effects of celecoxib and DMSO on time and distance of finding hidden platform during the test trials

References:

1. Chen C. C., Sun Y. T., Chen J. J. and Chang Y. J. (2001) Tumor necrosis factor-alpha-induced cyclooxygenase-2 expression via sequential activation of ceramide-dependent mitogen-activated protein Kinases, and IkappaB kinase1/2 in human alveolar epithelial cells. *Mol Pharmacol* **59**, 493-500.
2. Herschman, H.R. (1996). Prostaglandin synthase 2. *Biochem Biophys Acta*, **1299**, 125-140.
3. Marti Barros D., Ramirez MR., Dos Reis E. A. and Izquierdo I. (2004) Participation of hippocampal nicotinic receptors in acquisition, consolidation and retrieval of memory for one trial inhibitory avoidance in rats. *Neuroscience* **126**, 651-656.
4. Nordberg A., Alafuzoff I. and Winblad B. (1992) Nicotinic and muscarinic subtypes in the human brain: brain changes with aging and dementia. *J Neurosci Res* **31**, 103-111.
5. Paxinos, G. and Watson W. (1997). *The rat brain in stereotaxic coordinates*, Academic Press: San Diego.
6. Teather L. A., Packard M. G., and Bazan N. G. (2002) Post-training cyclooxygenase-2 (COX-2) inhibition impairs memory consolidation. *Learn Mem* **9**, 41-47.

Allothetic and idiothetic navigation in triangular pool and in three different sizes of circular pool

S. Wittnerova, J. Bures
Institute of Physiology, Academy of Sciences, Prague, Czech Republic

Abstract

Research into allothetic and idiothetic navigation of laboratory rats did not yet lead to definite assessment of the importance of hippocampus for these modes of spatial cognition. The purpose of the present study is to develop a task based on navigation to a single intramaze location that can be implemented with the same success from any point at the circumference of the circular pool and from three angles at the equilateral triangular pool either by allothesis (in light) or by idiothesis (in darkness). Convenient test location is centrally located hidden platform. Adult male Long-Evans rats were tested. The rats were trained from 3 start positions in the large and middle sized circular pools (South, near to NorthEast, near to NorthWest) corresponding to the positions of the angles of the triangular pool. The escape latency depends not only on the distance between the start position and the submerged platform but also on intramaze cue (the wall of the pool) and the area of the pool.

Keywords

Morris Water Maze, the pools of different geometry, idiothetic navigation, allothetic navigation

1 Introduction

The water maze [3] serves 25 years as an efficient tool in the studies of animal cognition based on the assumption [1] that the rat can find the hidden goal by allothesis, because it remembers its location relative to visible landmarks, or by idiothesis, because it remembers the track already passed from the same starting point to the same goal. The present paper examines these possibilities by studying the navigation of rats to an idiothetically easily detectable goal in light or in darkness [2]. Particularly suitable for this purpose is the center of a circular pool which is equally distant from all points at pool´s circumference and can thus be approached by a multitude of routes of identical shape. The results were evaluated not only by the length and duration of the escape swim but also by the logical sequence of the navigation steps, illustrating the animal´s approach to the solution of the task.

ORIENTATION	allothetic	idiothetic
extramaze	**cues** the walls of the room, windows	
intramaze	**cue** the wall of the pool	**stimuli from** proprioreceptors, vestibular system, memory of motor commands

Table 1. *Spatial orientation*

Navigation in the Morris water maze is mediated by allothetic mechanisms based on the relationship of the hidden escape platform to remote extramaze cues and on idiothetic mechanisms based on the memory of proprioceptive and vestibular signals received and of motor commands emitted during the previous navigation trials. The wall of the pool is a complex allothetic cue, but at the same time the departure point of idiothetic records (Table 1).The present paper examines its significance for navigation to the center of circular pools of 3 different sizes and of a equilateral triangular pool in light and in darkness.

2 Methods

Eight male Long-Evans hooded rats were 4 month old at the start of the experiments. They were trained to find a hidden escape platform in the center of pools of different geometry.

2.1 Pools of different geometry

We used large circular pool (diameter 188 cm), middle sized circular pool (diameter 120 cm), small circular pool (diameter 88 cm) and equilateral triangular pool fitting inside the large circular pool. The side of the triangular pool measures 162.5 cm. The size of the large circular pool corresponds to the circumscribed circle of the triangular pool.

2.2 Modification of the Water Maze procedure

The rats were trained from 3 start positions in the large and middle sized circular pools (South, near to NorthEast, near to NorthWest) corresponding to the positions of the angles of the triangular pool. The escape platform was located centrally because we can easily define principles, used by the rat when searching the platform. The centrally located platform is on the line perpendicular to the wall of the circular pool at the start and equidistant from each place of the wall. In triangular pool is the centrally located hidden platform equidistant from the three angles. Escape latency to pool center in light is learned with same efficiency as to other locations (typically to center of a pool quadrant). All manipulations were done in accord with the Law on Animal Protection of Czech Republic and with the appropriate directive of the European Communities Council (86/609/EEC).

3 Results

3.1 Comparison of performance in light and in darkness

We trained the rats in large circular pool, middle sized circular pool and triangular pool (the middle sized circular pool has equal area as the triangular pool) 9 days in light and after 6 days in darkness. Each day each rat swam 6 trials from the following start positions (S, NE, NW, S, NE, NW). The 7th trial of the day was a 1 min probe trial.

Comparison of asymptotic values in two different circular pools and triangular pool

- The asymptotic value of learning curve in the large circular pool is 8 s in light and is twice longer in darkness 20 s.
- The asymptotic value of learning curve in the middle sized circular pool is 5 s in light and 10 s in darkness.
- The asymptotic value of learning curve in the triangular pool is 4 s in light and 9 s in darkness.

3.2 Alternating training in the large and small circular pools

At the start of this experiment four male Long – Evans hooded rats were 4 months old. They were trained to an escape platform in the center of the large circular pool (diameter 188 cm) for 6 days in light. They reached asymptotic value 8s. Beginning on the 7th swimming day they were trained in the small circular pool (diameter 88 cm) [4].

Small circular pool in 5 different sectors of the large circular pool

Placing the small pool into one of 5 different sectors of the large pool (to the central sector of the large circular pool, into the south, north, west and east sectors of the large circular pool) showed that the changed room frame position of the small pool center does not interfere with the animals performance. This suggests that the task is solved on the basis of idiothetic memory and wall derived allothetic information but is not much influenced by remote allothetic cues.

Use of two memory traces after their alternating acquisition

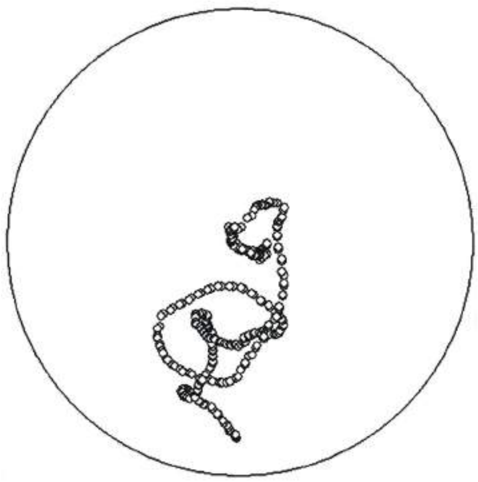


Figure 1. *Darkness, Day 19, Swim 3, 22 s*
Test in the large pool in darkness after alternating training in large and small pools in light.

After alternating acquisition of navigation to the centrally placed escape platform in the large and small circular pools, retention was tested by a probe trial performed in darkness in the large pool. The rat was obviously unaware of the size of the pool and attempted in most cases (Figure1) to find the escape platform first in the position corresponding to the small circular pool, i.e. perpendicularly to and 44 cm away from the wall. When this attempt failed some rats continued to swim in the same direction and concentrated their search around a point 94 cm away from the wall. Only when no platform was found there, the rat started to visit wider areas of the pool in a pattern resembling the earlier phases of search for a randomly located goal.

3.3 Equilateral triangular pool

The rats were trained to swim from the angels of the equilateral triangle placed into the large circular pool, to the escape platform in the center. The asymptotic value of the learning curve was 4 s in light and 9 s in darkness and was thus significantly shorter than the escape latency in the large circular pool. This indicates that this value depends not only on the start-goal distance, but also on the intramaze cues represented by the divergent walls of the triangle which may guide the rat to the goal. This assumption was confirmed by the probe trial when the animal was started in darkness from a point 20 cm from the pool´s center. Typical trajectories were formed by a straight swim toward the wall followed by swim along the wall into one of the angles and by subsequent approach to the escape platform. Less frequently encountered was wall contact triggering a random search in the direction of the diverging walls.

4 Discussion

The principles of searching the hidden platform

- The distance from start position at the wall in the large circular pool and from the angle of the triangular pool to centrally located platform is the same, but intramaze cue (the wall of the pool) and the area of the pool are different. So it means, the escape latency depends not only on the distance between the start position and the submerged platform.
- It seems to be little bit easier for the rat to find the platform in the triangular pool, because the walls of triangle can serve as guidelines for the rat.
- After alternating training in the large and small circular pools in light or in darkness, the animal swims perpendicularly to the wall of the pool, starts search at the presumed position of the small pool center and after short time swims to the large pool center and continues searching there.

Supported by GAČR grant No 309/03/0715, by MSMT ČR project No 1M0002375201 and by AV0Z 50110509.

References:

1. Mittelstaedt,M.L.; Mittelstaedt, H. (1980). Homing by path integration in a mammal. *Naturwiss*, **67**, 566.
2. Moghaddam, M.; Bures, J. (1996). Contribution of egocentric spatial memory to place navigation of rats in the Morris water maze. *Behav Brain Res*, **78**, 121-129.
3. Morris, R. G. M.; (1981). Spatial localization does not require the presence of local cues. *Learning Motiv*, **12**, 239-261.
4. Wittnerova, S.; Bures, J. (2004). Idiothetic navigation of rats in pools of simple geometry. *Fyziologicke dny,* Prague, Czech Republic. **Feb 3-5**.

Strain differences in the Hole-board and Fear conditioning tasks

M. Al Banchaabouchi, K. Gale and N. Rosenthal
Phenotyping Core facility, EMBL-Mouse Biology Unit, Monterotondo, Italy

Abstract

We are part of the Eumorphia Project, a program dedicated to developing and standardizing procedures for comprehensive phenotyping studies in the mouse to identify and understand the genetic basis of human diseases. Our facility is particularly interested in mouse behavioral phenotyping.

After the successful establishment of the first line screening, we are currently involved in developing specialized secondary tests designed to more fully characterize complex behavior phenotypes. We will be validating behavioral tasks such as the hole-board and the fear conditioning protocols. The hole-board maze is used to assess spatial learning and memory whereas fear conditioning is used to measure emotional and spatial learning as well as memory. Four male inbred mouse strains (C57BL/6J; BALB/c; C3H and 129S2/SvPas), at twelve weeks of age, will be studied and results will be further analyzed and compared.

Keywords

Holeboard task, fear conditioning, mice, strains, learning and memory.

Methods & Materials

1. Holeboard task—spatial discrimination learning

This task was originally described by Oades and Isaacson (1978) and first used to test deficit in attention and thought disorder (Oades RD, 1982). It is a hippocampal-dependant learning task that requires the search for food location based on spatial orientation and discrimination learning (Oades, 1981; Van der Zee et al., 1992). However, it has also been used as an explorative-behavior test (Monnier and Lolande, 1995) and a model of anxiety (Adamec et al., 2004; Kanari et al, 2005).

The hole-board is an enclosed square arena (L: 70cm; W: 70cm; H: 45cm) made of grey PVC containing four rows of four equidistant holes (2.5 cm diam.; 2 cm deep) in the floor plate. Food-restricted mice are required to locate small pieces of sweet pellets which are rodent treats (Fruit Crunchies, Bio-serv, NJ) in 4 of 16 holes.

Mice are housed 3 per cage and kept on a 12-hr light-dark cycle. Two days prior to training, food restriction is started from 5:00PM. Food is removed from their home cage and mice are familiarized with 2 pieces of sweet pellets in their home cage.

Body weight is monitored daily and the amount of chow food is pre-measured: 1.6 g regular laboratory food per mouse plus one piece of sweet pellet. This procedure will continue until the last day of the experiment.

The hole-board test is divided into three phases: habituation, acquisition and retention. From training day 1, food and sweet pellets are given only at the end of their last test session until the next day.

Habituation

Mice are habituated to the maze by allowing them to explore and eat freely in a baited maze over 2 days. The habituation consists of 2 sessions per day lasting 5 min each on day1 followed by two 3-min session on day 2. On day 1 during session 1, to reduce anxiety, mice are put into the center of the arena in groups of 3 and allowed to search a fully baited maze (on the floor and in every hole) for 5 min. Right after the last group/cage of animals, session 2 starts and animals are now placed individually in the maze which contains one pellet in each hole (all 16). On day 2, mice are put singly in the starting box; the 3 minutes session starts only when the mouse enters the arena with its four paws and the door is slid down. Only the holes are baited in the maze.

Note that the time is reduced to 3 min on day 2 as the acquisition and retention will be given for 3 min only.

Acquisition Training

From day 3, mice are subjected to a morning and afternoon training session (between 8:00 and 16:00hr). Each session consists of 5 trials (3 minutes each) after which animals are rewarded with one sweet pellet in their home cage.

A total of 6 training sessions are given over a 3-day period. Several different patterns are used, but each animal always receives a fixed pattern throughout the whole test.

The mouse is put in the starting box. Once it enters the arena, it has 3 minutes to search for and find four sweet pellets placed consistently in the same four holes. The trial is terminated as soon as the 4 baits are found and consumed or when the 3 minutes trial has elapsed. After each trial, the floor and the holes are cleaned with 70 % alcohol to remove possible odor cues.

The sequence is as follows: Session 1 (5 trials) is first done with cage 1, then the same procedure with cage 2 and so on until all cages go through the hole-board test. In the afternoon, we start Session 2 for each cage again and we proceed with the same order as in the morning.

The sequence of hole visits and the duration of the trial are recorded manually. Note that a second poke right after the first one in the same hole where the mouse has retrieved its bait is not counted.

Retention

The retention test, consisting of 5 trials of 3 min each, is given 48 hrs later.

Spatial learning and memory performance of the task are assessed by recording the latency to complete the task

(Searching Time), the number of reference errors (number of visits and revisits to non-baited holes considered as incorrect visits) and the reference memory ratio (RMR) which is the total visits (visits and revisits) of baited holes divided by the total visits of baited and non-baited holes.

Animals learn the task by orienting themselves in space using extra-maze visual landmarks.

2 The Pavlovian Fear Conditioning

Fear conditioning is a valuable behavioral paradigm for studying the neural basis of emotional learning and memory (Le Doux 2000; Maren 2001).

A polymodal operant chamber (Coulbourn Instruments, Allentown, PA, USA) is used. The chamber (18.5 x 18 x 21.5 cm) consisted of aluminium side walls and Plexiglas rear and front (the door) walls. A loudspeaker and a bright light are inserted in the opposite side walls 15 cm and 18 cm from the floor respectively, and constitute the sources of the cues during conditioning and cue-testing. The general activity of animals is recorded through the infrared cell placed at the ceiling of the chambers and is directly recorded on a PC computer using the Graphic State software (Coulbourn Inst.).

Conditioning

Mice are allowed to acclimate for 4 min, then a light/tone (10 kHz) CS is presented for 20 s and co-terminated by a mild (1 s, 0.4 mA) footshock (US). Two minutes later another CS-US pairing is presented. Mice are returned to their home cages 2 min later.

Testing

Testing is performed 24 h following conditioning session. Testing for the context is performed in the morning. The mouse is placed back into the same chamber that was used for the conditioning and allowed to explore for 6 minutes without presentation of the light/auditory CS. Testing for the cue is performed in the afternoon (about 5 h after the context testing). The contextual environment of the chamber is changed (wall color, odor and floor texture. The mouse is placed in the new chamber and allowed to habituate for 2 minutes then presented with light/auditory cues for 2 minutes. This sequence is repeated once again. At the end of testing, animals are returned to their home cages.

Freezing is scored automatically as the animal show no movement except breathing.

This project is supported by the EU-funded Eumorphia Mouse Phenotyping project, Grant number 50105.

References:

1. Adamec, R.; Walling S; Burton P. (2004). Long-lasting, selective, anxiogenic effects of feline predator stress in mice. *Physiol. Behav.*, **83**, 401-10.
2. Kanari, K.; Kikusui, T.; Takeuchi, Y.; Mori, Y. (2005). Multidimensional structure of anxiety-related behavior in early-weaned rats. *Behav. Brain Res.*, **156**, 45-52.
3. Le Doux, J.E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.*, **23**, 155–184.
4. Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.*, **24**, 897–931.
5. Monnier, C.; Lolande, R. (1995). Elevated (+)-maze and hole-board exploration in lurcher mutant mice. *Brain Res.*, **702**, 169-72.
6. Oades, R.D.; Isaacson, R.L. (1978). The development of food search behavior by rats: the effects of hippocampal damage and haloperidol. *Behav. Biol.*, **24**, 327-37.
7. Oades, R.D. (1982). Search strategies on a hole-board are impaired in rats with ventral tegmental damage: animal model for tests of thought disorder. *Biol. Psychiatry* , **17**, 243-58.
8. Oades, (1981). Type of memory or attention? Impairments after lesions of the hippocampus and limbic ventral tegmentum. *Brain Res. Bull.*, **7**, 221-6.
9. Van der Zee, E.A.; Compaan, J.C.; de Boer, M.; Luiten, P.G. (1992). Changes in PKC gamma immunoreactivity in mouse hippocampus induced by spatial discrimination learning. *J. Neurosci.*, **12**, 4808-15.

Towards a rodent model of the Iowa Gambling Task

R. van den Bos, W. Lasthuis & B. Spruijt
Ethology & Welfare, Utrecht University, Utrecht,The Netherlands

Abstract
The Iowa Gambling Task in humans is in principle suited to study long-term efficiency of behavior in a biologically relevant context. Animal models allow for studying the underlying neurobiology of decision-making processes and long-term efficiency of behavior in more detail and at greater depth than would be possible in humans. Therefore we set out to develop an animal analogue of this task in rodents and describe the first results in this paper.

Keywords
Iowa Gambling Task, human, animal model, rodent, radial maze

1 Introduction
As outlined elsewhere [2,3] three stepping-stones have emerged over the last decade allowing for studying human and animal long-term efficiency of behavior in a coherent biologically relevant framework. Thus far no animal models exist of the Iowa Gambling Task that is in principle suited to study long-term efficiency of behavior in a biologically relevant context [1,2,3]. Therefore we decided to develop a model that captures the essence of the Iowa Gambling Task, i.e. uncertainty of outcomes and a conflict between different options in terms of immediate and long-term pay off. As the rat and mouse model are essentially the same, we only describe the mouse model.

2 Materials and methods
We used an eight-arm radial maze for mice of which two arms were used as start arm and four as goal arms (Figures 1 and 2). In the centre of the maze mice could make their choice once they left one of the start arms. A trial started by lifting the octagonal slide door of the centre of the maze. Once mice hopped into one of the goal arms the octagonal slide door was lowered and the animals were allowed to sample the arm's contents.

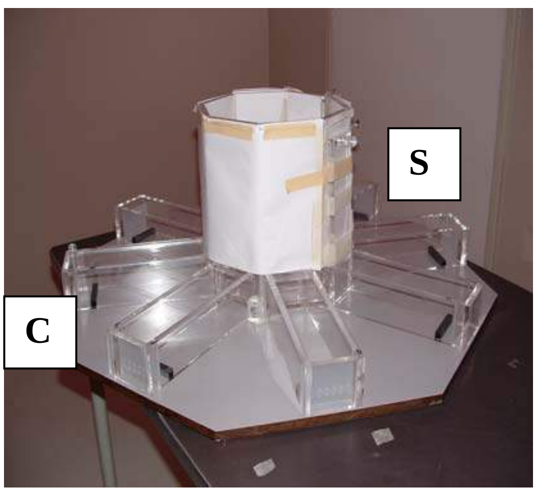


Figure 1. *Radial maze used for mice. S: 2 start arms; C: 4 choice or goal arms.*

To avoid the development of routines the order of start arms during sessions was randomized. We noticed in pilot-studies that mice were less accurate in their choices when the centre of the maze was used as start box. Mice appeared to choose more often the arm where they happened to be at the moment the octagonal slide door was lifted.
The walls of the centre of the maze were covered by white paper. This was done so that mice were not distracted, and could not directly see the experimenter while he/she rebaited arms. We noticed in pilot-studies that without internal cues mice could hardly differentiate the goal arms as no strong external cues were present to guide their choices. We therefore attached internal cues above the arms that helped mice to differentiate the goal arms. These cues consisted of a cross or circle (5*5 cm), that were either black or white. Colors rather than forms were positioned adjacent to one another.

Figure 2. *View inside maze. Above the arms cues were positioned (black/white circles and crosses, not shown in this picture).*

The goal arms differed in the extent to which they contained rewards ('sugar pellets') and punishments ('quinine saturated sugar pellets'). The 'quinine saturated sugar pellets' were made by mixing pieces of sweetened corn (Kellogg®) with a 180 mM quinine (Sigma-Aldrich, Schnelldorf, Germany) solution for a few minutes. They were subsequently dried to the air on paper towels. This procedure led to small and hard brownish pieces. Hereafter they were stored in small vessels and shielded from light until use. These pellets were unpalatable but not uneatable as judged by the behavior of the mice. For rats commercially available sugar pellets (Bioserve Inc., Frenchtown (NJ), USA) were used and treated with quinine solution in the same way. The quinine treated pellets were bitter/sweet tasting.
As in the Iowa Gambling Task [1] two arms contained occasionally big rewards (20-30 mg), but were not profitable in the long run, as they contained 9 times out of 10 'quinine saturated sugar pellets'. Two other arms

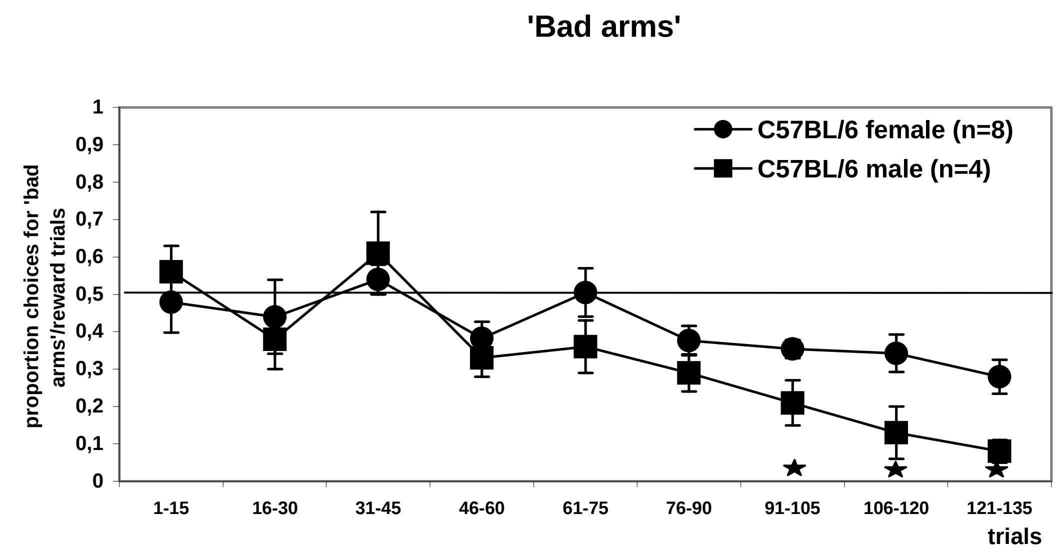


Figure 3. *Performance of male and female mice in the modified version of the experiment on choosing the long term 'bad arms'. Shown are means±SEM. See text for statistics. *: p≤0.05 (males versus females; two-tailed t-test).*

contained regular small rewards (5-10 mg) and were profitable in the long run, as they contained 2 times out of 10 'quinine saturated sugar pellets'. The difference between the two sets of arms in terms of the amount of sugar pellets was: 25 mg ('bad arms') versus 60 mg ('good arms'; i.e. a factor 2.4). Per block of 10 pre-programmed choices of each arm the position of 'sugar pellets' and 'quinine saturated sugar pellets' varied [1]. Thus, there was no rule to learn for mice when to run into rewards or punishments. 'Sugar pellets' and 'quinine saturated sugar pellets' were placed at the end of the arm behind a small black bar.

Before the start of the first trial mice were confined for 30 seconds in one of the start arms. Inter-trial interval was 15 seconds thereafter. Animals were allowed to explore the arms and eat the pellets for 2 minutes at maximum. Each mouse was given its own configuration of 'good' and 'bad' arms to avoid bias in the experimental design.

Hungry male and female mice (85-90% free feeding weight) of the C57BL/6 strain were given 15 trials a day during 9 days. Food regimen was such that they received a specified amount of food between 16.00-07.00 hrs based on free feeding hourly consumption. During weekends they had unrestricted access to food. Water was freely available throughout.

Mice were given one habituation session (5 min) before the first experimental day. The animals were 6-7 months old when tested. They were kept on a 12 hr day-night cycle (lights on: 07.00-19.00 hrs) in a temperature (21 degrees Celsius) and humidity controlled (50%) room in Macrolon type II cages (2-3 individuals) with a shelter and tissues as enrichment. Experiments took place in the dark phase between 10.00-15.00 hrs. To avoid aggression between male mice, they were put into separate cages during testing, and released together in their home cage after testing.

3 Results

Two features stood out in the first series of experiments: (i) mice selected more often the 'good' arms across days, i.e. they learned the best option ($F(8,64)=3.357$, $p \leq 0.003$); (ii) male mice tended to perform better across days with less variation between individuals than female mice ($F(8,64)=1.841$, $p \leq 0.086$). Based on these results we made a modification of the original test to assess whether female mice would be able to discriminate the arms of the maze and to enhance the performance of both male and female mice. In this modified version two arms were left empty while one arm was used as the 'good' arm and one as the 'bad' arm. The remainder of the protocol was left the same. When males and females were tested in this way, female mice performed worse than male mice, but only so for the difference between 'good' and 'bad' arms (day 5-9: $F(1,10)=15.602$, $p \leq 0.003$; Figure 3). Female DBA/2 mice were also tested in this modified version. They tended to perform worse than CB57BL/6 mice but again only so for the difference between 'good' and 'bad' arms ($F(8,136)=1.877$, $p \leq 0.068$).

4 Discussion

The present study is a first step towards developing an animal analogue of the Iowa Gambling Task. The set-up we have chosen seems promising as we observe a profile in choice behavior that resembles that in humans [1]. Furthermore we observe differences in this behavior between strains and gender, specifically related to rewards. The gender differences may be due to the effect of the estrus cycle on reward sensitivity and/or long-term efficiency of behavior.

References

1. Bechara, A.; Damasio, A.R..; Damasio, H.; Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, **50**, 7-15.
2. Bos, van den R. (2004). *Emotion and Cognition.* In: M. Bekoff (Ed.) The Handbook of Animal Behavior, 554-557, Westport (CT), Greenwood Press.
3. Bos, van den R.; Houx, B.B.; Spruijt, B.M. (2005). The effect of reward magnitude differences on choosing disadvantageous decks in the Iowa gambling task. *Biological Psychology* (in press).

Measuring learning and memory defects in lysosomal α-mannosidase deficient mice

K. Caeyenberghs[1], P. Saftig[2] & Rudi D'Hooge[1]
[1]Laboratory of Biological Psychology, University of Leuven, Belgium
[2] Biochemical Institute, University of Kiel, Germany

Abstract

Deficiency of α-mannosidase causes the autosomal recessive lysosomal storage disease, α-mannosidosis. Mice with targeted disruption of the lysosomal α-mannosidase gene were generated as the first mouse model for human α-mannosidosis. Performance on tests of learning/memory was assessed in α-mannosidase deficient mice and their wild type littermates at 3 months of age. Passive avoidance learning was not impaired in α-mannosidase mice. In the Morris water maze, α-mannosidase deficient mice displayed impaired hidden-platform acquisition performance, and retention deficits during the probe trials. Impaired learning/memory, reported here in α-mannosidase deficient mice, may relate to the decline of cognitive functions in α-mannosidosis patients. α-Mannosidase deficient mice provide a valid model for investigation of new therapeutic strategies for α-mannosidosis, such as enzyme replacement.

Keywords

α-Mannosidosis; Lysosomal disorder; Transgenic mouse model; Learning; Memory

Introduction

α-Mannosidosis is a lysosomal storage disorder that is caused by the deficiency of lysosomal α-mannosidase, leading to the multisystem accumulation of mannose-rich oligosaccharides [8]. The disease can be divided in two types, the infantile phenotype (or type I) and the juvenile-adult phenotype (or type II) according to its clinical manifestations. It is not known, however, whether the existence of the two different forms has any physiological evidence [1]. The mental abilities are slightly to moderately impaired in most human α-mannosidosis patients, and repeated IQ testing shows a consistent cognitive deficit [3,4]. Some authors have also reported psychotic symptoms and emotional changes in some α-mannosidosis patients [2,5]. Mice with targeted disruption of α-mannosidase were generated as a mouse model for α-mannosidosis [7].

Learning and memory tests

A cohort of female mice (14 α-mannosidase knockouts, 16 wildtype littermates) were examined at the age of 3 months.

Passive avoidance learning was tested is a step-through box. During training, dark adapted mice were put in the small illuminated compartment of the box and 5 s later, the sliding door connecting both compartments was opened. Upon entry, they received a slight electric footshock (0.3 mA, 1 s) using a constant current shocker (MED Associates Inc., St Albans, Vermont, USA), and 24 h later they were again placed in the box for a testing trial. Step-through latency was recorded during the training and testing trials with a cut-off time of 5 min.

During the training trial of the step-through task, knockout mice entered into the dark compartment just as quickly as their wildtype littermates (Figure 1). During the testing trial, several knockouts and controls did enter the dark compartment during the 300s observation period. Step-through latency revealed no difference between knockout mice and controls.

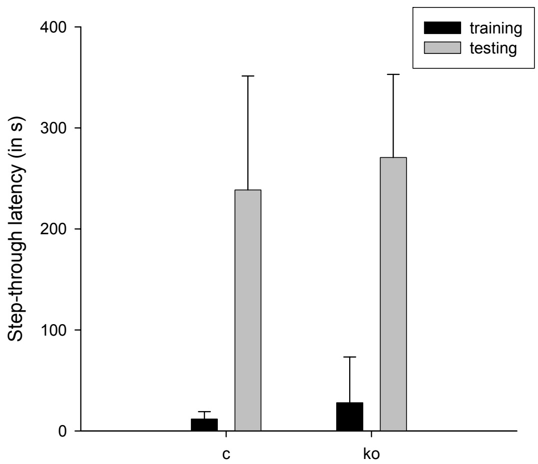


__Figure 1.__ Passive avoidance learning in 3-month-old wild type (c, n=16) and α-mannosidosis mice (k, n=14). Black bars depict mean step-through latency (and SEM) during training, grey bars during testing. Step-through latency during the testing was not significantly reduced in the knockout group (p>0.05).

The Morris-type water maze consisted of a circular pool (diameter: 150 cm; height: 30 cm) filled with opacified water (26°C). A round perspex platform (15 cm diameter) was placed at the centre of the target quadrant, 1 cm beneath water surface. One daily trial block consisted of four swimming trials randomly starting from one of the four different positions with 15 min intertrial intervals. If the animal could not find the platform within the maximum swimming time of 120 s, they were placed on the platform, and had to stay there for 15 s. Eight acquisition trial blocks were followed by a probe trial during which mice had to swim for 100 s in the pool without a platform. During acquisition and probe trials, the animals were tracked using *EthoVision 3.1* video tracking equipment and software (Noldus bv, Wageningen, The Netherlands).

During the acquisition trials of the first week, no significant difference in escape latency was found between knockouts and controls. Also, path length and velocity were not significantly different between deficient mice and controls. Only the effect of interaction between genotype and trial block on escape latency ($F_{4,149}$=10.3; p=0.003) and path length ($F_{4,149}$=14.2; p=0.034) was significant. In the second session, acquisition training did reveal a consistent difference in escape latency between deficient mice and controls. Escape latency of knockouts was consistently increased on all trial blocks (Figure 2). Two-way repeated measures ANOVA demonstrated a highly significant effect of genotype on escape latency ($F_{1,149}=9.9$; $p=0.004$). Genotype had also a highly significant effect on path length ($F_{1,149}=16.0$; $p<0.001$). The effect of interaction between genotype and trial block on path length was significant ($F_{4,149}=7.7$; $p=0.006$).

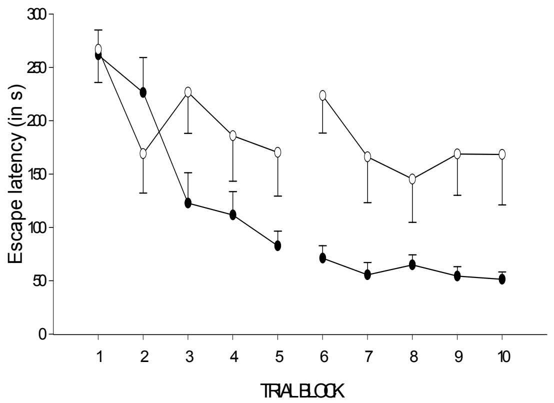


Figure 2. Water maze performance in 3-month-old wildtype and α-mannosidase knockout mice. Acquisition of the task consisted of 10 trial blocks. Plots show mean escape latency (and SEM) for wildtype (closed circles) and knockout mice (open circles). Escape latency was consistently longer in the knockouts the trials blocks at the end of training (p<0.05).

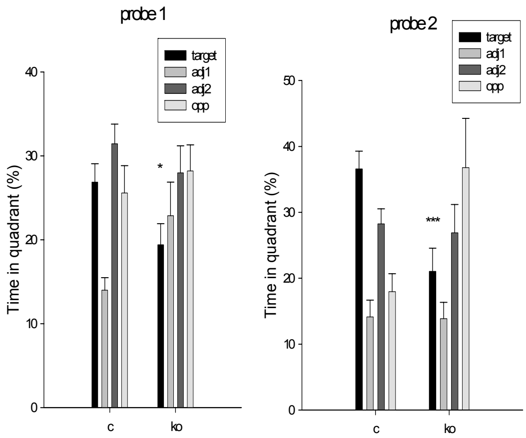


Figure 3. Probe trial performance following hidden-platform acquisition training in the same control and α-mannosidase deficient mice shown in Fig. 4. Data are mean percentage of time (and S.E.M.) spent searching for the platform in the each of the 4 pool quadrants (i.e., target, adjacent 1 and 2, and opposite quadrant) (see insert) during probe trail 1 and 2. There was no difference between genotypes in probe trial performance. However, knockouts spent les amounts of time in the target quadrant than controls during probe 1 and probe 2 (p<0.05).

Comparing the percentage spent in the different quadrants of the pool of probe 1 between α-mannosidase deficient mice and their wildtype littermates, showed no difference between genotypes in probe trial performance (Figure 3). There was a significant effect of genotype x quadrant interaction on mean percentage of search time in the different quadrants ($F_{3,112}=3.2$; $p=0.026$ for the first probe trial, $F_{3,112}=7.2$; $p<0.001$ for the second probe trial). Time searching the target quadrant was significantly different between knockout and control mice: knockouts spent significantly less time in the target quadrant than controls ($p=0.033$ for probe 1; $p=0.001$ for probe 2). Knockouts remained in probe 1 on average 54.1 ± 3.6 cm removed from the target area compared to 42.8 ± 2.3 cm in controls ($p=0.012$). In probe 2, knockouts crossed the target area 3.2 ± 0.8 times compared to 5.9 ± 0.8 times in controls ($p=0.019$). The latency of the first approach of the target was 40.2 ± 10.4 s in knockouts vs. 15.6 ± 2.9 s in controls ($p=0.022$). Also, knockouts remained on average 56.8 ± 5.7 cm removed from the target area compared to 35.7 ± 2.0 cm in controls ($p<0.001$).

Active avoidance was tested in an operant chamber. Mice were put on a food restriction schedule: 1h of feeding per 24h for 4 days prior to training, which reduced their body weight by 10-20%. Food restriction was continued throughout, but body weight was monitored daily to prevent it to drop below 80% of the initial value, and trained by gradual shaping to nose poke for food pellets (Noyes precision pellets, Research Diets Inc., New Brunswick, New Jersey, USA) in Coulbourn operant cages for mice (Coulbourn Instruments, Allentown, USA). Different reinforcement schedules were programmed, and data were collected with Programme State (Coulbourn Instruments, Allentown, USA). During the 30-min training sessions, mice were able to obtain food pellets by using the nose poke device inside the cages, but the reinforment schedule was gradually changed to arrive at a stable response rate by intermittent reinforcement. Shaping started with 8 continuous reinforcement (CRF) sessions. This was followed by 7 fixed-ratio sessions (FR-5: 5 responses/reinforcement), during which they had to nose poke a fixed number of times to receive a food pellet. These FR trials were followed by a variable-interval constant-probability protocol (VI-30s trials), during which lever presses produced on the average a food pellet every 30 s. There were 7 VI-30s sessions. The next 8 trials consisted of an adapted Sidman active avoidance task [6]. A 0.2-mA electric shock, scrambled by Coulbourn shocker, was delivered through the grid for 0.3 s. The mouse active avoidance response was to use the nose poke device. The shock-shock interval was 5 s, and the response-shock interval was 25 s.

In the instrumental learning task (Figure 4), animals were trained to nose poke for food pellets. Acquisition curves over 22 shaping sessions were highly different in knockouts and controls. Repeated-measures ANOVA demonstrated a highly significant main effect of genotype ($F_{1,615}=7.98$; $p=0.009$). There was also a highly significant main effect of session ($F_{21,615}=63.51$; $p<0.001$) and a significant genotype x session interaction ($F_{21,615}=2.47$; $p<0.001$).

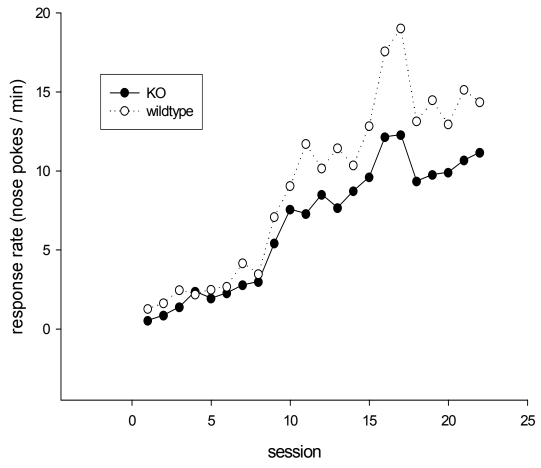


Figure 4. Instrumental conditioning sessions in 3-month-old wildtype (open circles), and α-mannosidase knockout mice (closed circles). During the acquisition, response rates (expressed as mean number of nose pokes/min) gradually increased. The rates were significantly different between the genotype groups.

In the Sidman trials (Figure 5), repeated-measures ANOVA on the response rate revealed an effect of session ($F_{7,223}$=4.52; p<0.001), but neither a main effect of genotype ($F_{1,223}$=0.78; p=0.39), nor a genotype x session interaction ($F_{7,223}$=0.50; p=0.83).

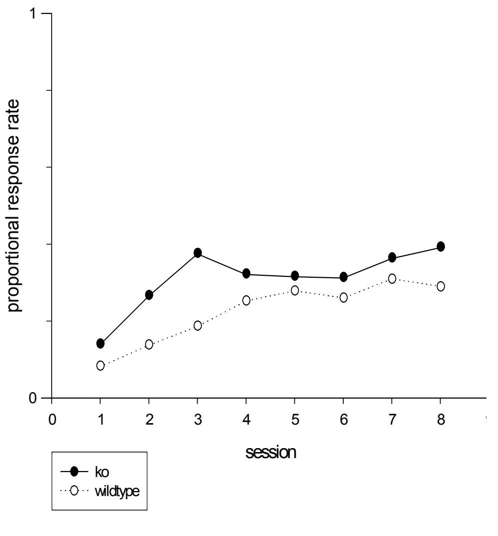


Figure 5. Sidman sessions in 3-month-old wildtype (open circles), and α-mannosidase knockout mice (closed circles). The response rates (expressed as the number of nosepokes in the Sidman session divided by the mean number of nosepokes of the previous 3 VI-30 s sessions) gradually increased, but rates were not significantly different between the genotype groups.

Conclusion

In the present study, we have compared the performance of 3-month-old α-mannosidase deficient mice on several behavioural tests to that of control animals. We have found some behavioural abnormalities in these animals, which could relate to observations in human patients. Cognitive impairment is a characteristic feature of α-mannosidosis, and it was also found in our animal model. Performance of the α-mannosidase deficient mice was impaired in the acquisition of the hidden-platform water maze as well as in the retention during the probe trials. α-Mannosidase deficient mice did not show impaired retention performance in passive avoidance learning, nor did they show impairments in active avoidance Sidman conditioning. α-Mannosidase deficient mice were only impaired during the shaping phase of the latter instrumental learning task, where mice learned to associate nose poking with food rewards.

References

1. Daniel, P.F., Winchester, B., & Warren, C.D. (1994). Mammalian α-mannosidases-multiple forms but a common purpose? *Glycobiology, 4*, 551-566.
2. Gutschalk, A., Harting, I., Cantz, M., Springer, C., Rohrschneider, K., & Meinck, H.M. (2004). Adult alpha-mannosidosis: Clinical progression in the absence of demyelination. *Neurology, 63*, 1744-1746.
3. Noll, R.B., Kulkarni, R., & Netzloff, M.L. (1986). Follow-up of language and cognitive development in patients with mannosidosis. *Archives of Neurology, 43*, 157-159.
4. Noll, R.B., Netzloff, M.L., Kulkarni, R. (1989) Long-term follow-up of biochemical and cognitive functioning in patients with mannosidosis. *Archives of Neurology, 46*, 507-509.
5. Seidl, U., Giesel, F.L., Cantz, M., Schmidbauer, M., Schroder, J., & Pantel, J. (2004). Unusual course of alpha-mannosidosis of paranoid-hallucinatory psychosis. *Nervenarzt, epub 2004*.
6. Sidman, M. (1953). Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science, 118*, 157-158.
7. Stinchi, S., Lüllman-Rauch, R., Hartmann, D., Coenen, R., Beccari, T., Orlaccio, A., von Figura, K., & Saftig, P. (1999). Targeted disruption of the lysosomal α-mannosidase gene results in mice resembling a mild form of human α-mannosidosis. *Human Molecular Genetics, 8,* 1365-1372.
8. Thomas, G.H.. (2001). Disorders of glycoprotein degradation: α-mannosidosis, β-mannosidosis, fucosidosis and sialidosis. In C.R. Scriver, A.L. Beaudet, W.S. Sly, & D. Valle (Eds.), *The Metabolic and Molecular Bases of Inherited Disease*, 8th Ed, McGraw-Hill, New York, pp. 3507-3533.

The significance of multiple stressors in prehistory for the specificity in the learned helplessness model of depression

S. Chourbaji, C. Dormann, C. Zacher, P.Gass
Central Institute of Mental Health (CIMH), Mannheim, Germany

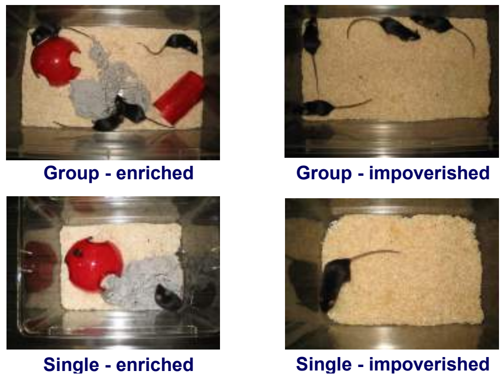
Abstract
Animal models represent the only possibility to gain insights into the pathogenesis of particular diseases. Depression constitutes a major challenge, since characterization of disease-specific traits is complicated due to multifactorial derivation. The understanding of associations of diverse factors represents a major target of behavioral animal models. Working with learned helplessness consequently requires exact consideration of modulating aspects. This study focuses on potential distorting factors like housing conditions. To guarantee a specific readout, we compared the effects of different social and environmental factors, regarding general and helplessness-specific effects, confirmed by a pharmacological validation. In studies of depression and emotionality it is important to establish standardized protocols, involving the animal's environment, to precisely assess potential sources of stress and exclude artifacts. The design/modification of the learned helplessness bears the advantage not only to detect genetic aspects by investigating mice carrying mutations of target genes, but also exploiting fundamental causes of depressive-like phenotypes.

Keywords
Learned helplessness, evaluation, mice, housing, depression

1 Introduction
Working with behavioral animal models for disease states requires exact knowledge and control of "internal" and "external" factors that may influence the outcome of the experiments. Internal factors comprise the animals' gender, age, hormone state, satiety and genetic background; external factors are housing, and social conditions as well as many other factors, which are essential to consider. The learned helplessness paradigm, a rodent model of depressive-like states with appropriate construct and face validity (Maier, 2001; Vollmayr, 2001), may reflect the described housing-induced physiological changes at the behavioral level. This prompted us to design a reliable helplessness procedure in C57BL/6N mice, excluding and ranking possible artifacts.

2 Materials and Methods

2.1 Animals
All behavioral tests were conducted in 4-months old male C57Bl6/N mice. The animals were purchased from Charles River (Sulzfeld, Germany) at an age of 8 weeks and were then housed for 7 weeks under specific conditions in type II cages for single and type III cages for group rearing (4 mice per cage). Four cohorts of mice were investigated: 1.) single housed, impoverished conditions (n=16), 2.) single housed, enriched conditions (n=16), 3.) group housed, impoverished conditions (n=16), 4.) group housed, enriched conditions (n=16). Enrichment consisted of red, transparent plastic mouse igloos and tunnels (EMSICON Jung GmbH, Forstinning, Germany) that were cleaned weekly, when the cage was changed. Additionally, nesting material (tissue) was provided. Impoverished cages simply contained bedding material. All mice were kept in the same room, in a 12h:12h reversed dark-light cycle, lights on at 6.00 p.m. Water and food pellets were available *ad libitum*. Body weight was assessed once a week when the cages were cleaned. All experiments were approved by German animal welfare authorities.



2.2 Behavioral procedures
The mice were subjected to tests for locomotion, exploration, and anxiety, followed by the learned helplessness procedure. Following earlier recommendations for repetitive behavioral testing, animals were initially tested in the experiments ranked as less stressful (McIlwain, 2001). Considering that the determination of the pain threshold by the hot plate test could have a direct influence on helpless behavior, we assessed the effect of housing conditions on pain sensitivity in a separate group of mice reared under identical housing conditions. Test procedures were essentially performed as described earlier (Reif, 2004). Between individual tests was a pause of at least 24 h. Prior to each test, mice were acclimatized to the experimental room for 30 min. All behavioral tests were conducted during the dark cycle.

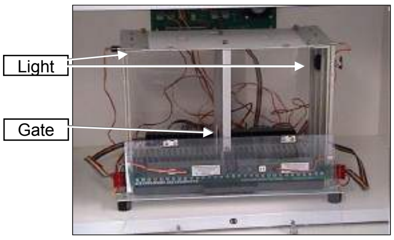
Open field test
Activity monitoring was performed in a square, white open field, measuring 50x50 cm^2 and illuminated from above by 25 Lux. Mice were placed individually into the arena and monitored for 15 min by a Video camera (Sony CCD IRIS). The resulting data were analyzed using the image processing system EthoVision 2.3 (Noldus Information Technology, Wageningen, the Netherlands). For each sample, the system recorded position and the occurrence of defined events. Parameters assessed in the present study were total distance moved, velocity, and thigmotaxis (*i.e.* the percentage of time spent in a corridor with maximal distance of 10 cm to the walls).

Dark-Light Box test
The dark-light box consisted of two plastic chambers, connected by a small tunnel. The dark chamber measured 20x15 cm^2 and was covered by a lid. The light chamber, measuring 30x15 cm^2, was white and illuminated from above with an intensity of 600 Lux. Mice were placed into the dark compartment, and the latency to the first exit, the number of exits, and the total time spent in the light compartment were recorded for 5 min.

Learned helplessness

The learned helplessness paradigm was conducted as described earlier (Reif, 2004). Mice received 360 footshocks (0.150 mA) on two consecutive days in a transparent plexiglas shock chamber (18x18x30 cm^3), equipped with a stainless steel grid floor (Coulborn precision regulated animal shocker, Coulborn Instruments, Düsseldorf, Germany). The footshocks applied were unpredictable with varying shock-durations (1-3 s) and interval-episodes (1–15 s), amounting to a total session duration of approximately 52 min. 24 h after the second shock exposure, learned helplessness was assessed by testing two-way avoidance in a shuttle box (Coulborn Instruments, Düsseldorf, Germany). Spontaneous initial shuttle activity from one compartment to the other was monitored during the first two minutes by red-light beams at the bottom of each of the two compartments. Thereafter, specific avoidance learning was tested over 30 trials. Each trial started with a light stimulus of 5 s, announcing a subsequent footshock (0.150 mA; maximal duration: 10 s). The duration of this procedure varied (total testing time averaged 20 min) according to each animal's individual performance in terms of failures, *i.e.* not attempting to escape from the shock, and escape latencies, *i.e.* latency to escape after onset of the shock.

Unshocked mice provided a control for a possible effect of housing conditions *per se* on the performance in this test.



Hotplate test

To exclude altered pain sensitivity as a modulating factor for learned helplessness, an additional batch of mice (n=8 per condition) was tested on the hotplate test (ATLab, Vendargues, France). Temperature was set at 53°. Latency to first reaction, *i.e.* licking hind paws or jumping, with a cut-off at 45 s, was assessed.

Statistical Analyses

Statistical analyses were performed with two-way ANOVAs, followed by Fisher's (LSD) *posthoc* tests (XLstat program Version 7.5, Addinsoft). Comparisons of escape deficits in the learned helplessness paradigm between shocked animals and non-shocked controls were performed by t-tests.

3 Results

- Group housing affects exploratory behavior, but not locomotion
- Group housing and enrichment evoke reduced anxiety-like behavior
- Housing conditions affect the coping behavior in the learned helplessness paradigm
- Housing conditions do not affect pain sensitivity

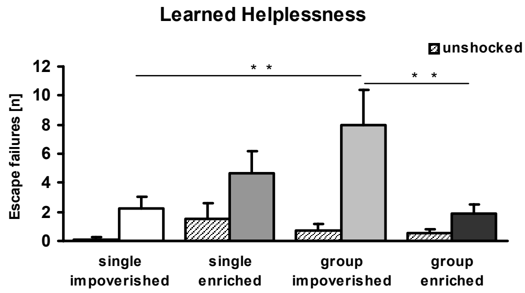


Figure 1. *In the learned helplessness task, no difference is seen between the different rearing conditions in unshocked animals, as seen in the respective escape failures. Following shock exposure, however, ANOVA shows a significant interaction of the two housing factors for failures and latencies (data not shown). Post-hoc analyses demonstrate that group impoverished mice exhibit significantly more helplessness behavior than group enriched or single impoverished animals.*

4 Discussion

The increased helplessness observed under impoverished group housing conditions could have resulted from alterations of general home cage behaviors, which are important to consider with respect to the performance of the test, or from interference of stress with different stages of the complex paradigm, which are determined to assess helplessness. First, the elevated number of failures and increased escape latencies could have resulted from locomotor deficits produced by the specific housing conditions. However, as assessed in the open field (total distance moved, locomotor velocity), locomotion was not influenced by the various housing conditions. Second, housing could affect the pain threshold, thus changing the impact of the electroshocks used for the induction of helplessness. However, the pain threshold as measured by the hotplate test was not altered by housing. Third, shuttle-box performance, *i.e.* the behavioral read out in the helplessness paradigm following the uncontrollable foot shock exposures, could have been influenced by the respective animal husbandry. However, non-shocked animals displayed similar avoidance behavior irrespective of their housing conditions, indicating that the mode of housing did not affect two-way avoidance *per se.*

References

1. Maier, S.F. (2001). Exposure to the stress environment prevents the temporal dissipation of behavioral depression/learned helplessness. *Biological. Psychiatry*, **49**.
2. McIlwain, K.L.; Merriweather, M.Y.; Yuva-Paylor, L.A.; Paylor, R. (2001). The use of behavioral test batteries: effects of training history, *Physiol. Behav.*, **73**: 705-17.
3. Reif, A.; Schmitt, A.; Fritzen, S.; Chourbaji, S.; Bartsch, C.; Urani, A.; Wycislo, M.; Mossner, R.; Sommer, C.; Gass, P.; Lesch, K.P. (2004). Differential effect of endothelial nitric oxide synthase (NOS-III) on the regulation of adult neurogenesis and behaviour, *Eur. J. Neurosci.*, **20**: 885-95.
4. Seligman, M.E.; Weiss, J.; Weinraub, M.; Schulman, A., (1980). Coping behavior: learned helplessness, physiological change and learned inactivity, *Behav. Res. Ther.*, **18**: 459-512.
5. Vollmayr, B.; Henn, F. (2001). Learned helplessness in the rat: Improvements in validity and reliability. *Brain Research Protocols*, **8**, 1-7.

A new behavior test for distinguishing spatial learning deficits from complex learning deficits in aged-rodents

J.T. Drott and F.A. Haun
NeuroDetective Int., Quakertown PA, U.S.A. 18951

Abstract

The study reported here is an initial attempt to describe age-related differences in performance of foraging/hoarding behavior by pigmented rats. The traditional Barnes task, which assays spatial memory deficits in rodents, was modified to reveal a separate learning deficit in aged rats, in addition to the memory deficit. The modification exploits a naturally occurring, complex behavior in rats, foraging for food.

Both 9 mo. and 33-36 mo. Fischer/Norway rats were allowed to forage for large food pellets placed in 3 of 15 possible food cups in the center of a circular table, and return with them to one of 8 boxes on the periphery. Location of the three pellets remained constant throughout testing; however the animal's home box periodically changed location. All boxes were visible to the animal and were identical except for panels hidden behind felt that blocked entry to the 7 non-home boxes.

Using image analysis software, changes in foraging strategy as well as accuracy in returning to the home cage were measured over multiple trials. Specific measures included number of visits to all food cups, repeat visits to individual food cups during a foraging session before successfully finding a food pellet, and length of the return path to the animal's home location following acquisition of a food pellet. From the first location of the home box, the old animals' foraging strategy was less efficient (more empty cups explored) compared to young animals, but both groups learned to return to the home location equally well. When the home location was moved, the old animals were initially less accurate in returning to the new location, but improved their accuracy over several trials. However, foraging strategy in the aged animals never improved throughout the test. Thus this foraging/homing test provides an opportunity to detect drug effects on two very different cognitive deficits in old rats, one involving memory for a home location and the other involving complex learning.

Keywords

Foraging; spatial learning; aged rats,

1 Introduction

Foraging/hoarding behavior is a complex, naturalistic behavior common to many species, with distinct components. The animal chooses a location to search for food, conducts the search, determines whether to consume the food where it is found or bring it to another location for either consumption or hoarding, and if the latter, navigates to either a food storage point or its home location. Current research into this complex behavior is focused on optimal foraging theory (for reviews see Stephens & Krebs, 1986; Commons et al., 1987; Shettleworth, 1998) and the neural bases of returning to a home location where the food is either hoarded or consumed (Whishaw & Tomie, 1989; Whishaw & Gorny, 1999).Thus far there has been no report of age-related deficits in this behavior, which might be predicted because of the multiple cognitive functions that are involved, including spatial orientation and memory (for previously visited search areas and for the home location). This study is an initial attempt to describe age-related differences in performance of foraging/hoarding behavior by pigmented rats.

2 Materials and Methods

2.1. Animals

Fischer/Norway F1 hybrid female rats aged 9-11 and 34-36 months at the onset of the study were used. Rats were housed by pairs in standard polycarbonate cages with water available ad libitum. One week prior to the beginning of testing, rats were food restricted to 85% of their normal body weight.

2.2 Foraging/Homing Apparatus

Animals were tested on a circular platform (178 cm diameter), the surface of which consisted of rough, opaque Plexiglas. Eight identical black boxes, one of which was designated as the animal's home box, surrounded the perimeter of the platform equi-distant from each other. The boxes were constructed of black Plexiglas (23 cm x 23 cm x 23 cm), enclosed on three sides, with a guillotine door comprising the fourth side; the top was covered with a removable roof. A piece of black felt was placed in front of the guillotine door of each box. In the seven non-home boxes the guillotine door remained down during all trials, while in the home box the (unseen) guillotine door was always in the up position. Fifteen white plastic food cups were affixed to the platform in an irregular pattern, such that the summed straight-line distances from three of the cups to each of the eight boxes varied by less than 3 cm. On each trial only this triad of cups contained food pellets. The testing apparatus used here is a modified version of the Barnes maze (Barnes, 1979), originally designed to be a dry land version of the watermaze and employing aversive motivation to stimulate homing behavior.

2.3 Procedure

2.3.1 Training

Animals were initially acclimated to their home box by placing it in a separate open-field box (constructed of black Plexiglas, 60 cm x 75 cm x 22 cm).

On the day following acclimation, the animal was placed inside its home box, which had been affixed to one of the 8 possible locations (randomly selected) on the periphery of the foraging/homing platform. The home box contained bedding from the residence cage but no food pellets. There were 15 food cups on the platform, and a large, reward food pellet (1000 mg, Bio-Serv, Noyes) was placed in each cup. The pellets varied in flavor; during this training phase each animal's pellet preferences were noted, and those preferred pellets used during subsequent testing. The animal received daily 10-minute acclimation sessions, to familiarize the animal with the testing apparatus. Acclimation continued until the animal exited the home box, retrieved a pellet, and carried it back to the home box, repeating this pattern three times within a session. If the animal began to eat a pellet without first returning to the home box, it was immediately removed from the platform and returned to its residence cage. Any animal that did not meet training criterion after 10 trials was not tested further.

2.3.2 Initial Learning

On the day after meeting training criterion, the animal was again placed in its bedding-containing home box, which was then placed in a new peripheral location 90 degrees away from the training location. Food pellets were placed in the 3 specific cups for that home location (see 2.2). The physical location of the food pellets did not change during the initial learning or re-learning portions of the study. While the pellets' positions remained constant for each animal throughout its entire testing, the sum of the straight-line distances from the three pellet-containing cups to both home-box locations (initial learning and re-learning) varied by less than half a body-length. No two baited cups were adjacent to each other, nor was any baited cup the closest cup to the home box.

The initial learning phase ended once the animal retrieved all three food pellets (one at a time) and returned with them to its home box without first approaching a non-home box, on two consecutive daily trials. An incorrect box approach was considered to have occurred if the animal came within 4 cm of any part of a non-home box while carrying a food pellet in its mouth. A trial was terminated when (a) all three pellets had been returned to the home box; (b) 12 minutes had elapsed; or (c) the animal began eating a pellet outside the home box. Each animal received 1 trial/day until reaching criterion, up to a maximum of 10 trials.

2.3.3 Re-learning

After initial learning, the home box was moved to a new location, 135 degrees away from the previous location. Trials continued using the same procedure and criterion as during initial learning, again up to maximum of 10 trials. For this re-learning session, the physical location of the food-containing cups was the same as in initial learning for a particular animal, but their locations relative to the home-box location were different.

2.4 Measures

Each trial was videotaped from an overhead camera, and the following measures obtained from the videotape.

2.4.1 Return Path Distance

The distance traveled by the animal (in cm) from a food-containing cup to its home box with the food pellet in its mouth was recorded separately for each pellet-retrieval, and included any attempts to enter a "non-home" box along the way. For data analysis the return-path distances from all three food-containing cups were summed for each animal.

2.4.2 Total Number of Food Cups Searched

Total number of food cups searched (repeat searches included) was recorded for each animal, separately for each trial. A "search" was considered to have occurred if the animal's snout was directly over the cup for 1 second.

2.4.3 Repeat Visits to Pellet-Containing Cups

Total number of repeat visits to the three pellet-containing cups was also recorded, separately for each trial. The definition of "visit" is the same as for "search," above. In order for an animal to make a repeat visit, it must first have retrieved a pellet from that cup earlier in the same trial.

2.4.4 Repeat Visits to Empty Food Cups

Total number of repeat visits to empty cups was also recorded, separately for each trial. "Empty cups" are the twelve that never contained a food pellet throughout testing of a particular animal at the particular home-box location.

3 Results

Both the young and old rats reached the criterion of returning directly to their home box with each of three food pellets on two consecutive trials, within the same number of trials. This was true when the home box was in its initial location and when it was moved to a new location. However in reaching this criterion, the aged animals consistently searched more food dishes than the young animals before retrieving a food pellet, and did not become more efficient in their search with increasing practice. In addition, within a trial the old animals made more return visits to food cups from which they had already retrieved a pellet on that trial, compared to the young animals; however this effect occurred only while retrieving the first three pellets from the new location of the home box. By the last trial, the old animals made few if any such repeat visits. Similarly, the old animals took longer routes returning to their home box with the food pellets, often attempting to enter several non-home boxes, compared to young animals, which rarely attempted a wrong-box entry. However this only occurred when the home box had been moved to a new location and then only in the initial three return trips with the food pellets. By the last trial both young and old animals were taking direct routes back to their home box.

References

1. Barnes, C.A., 1979. Memory deficits associated with senescence: A neurophysiological and behavioral study in the rat. *J. Comp. Physiol. Psychol*., **93**(1), 74-104.
2. Commons, M.L., Kacelnik, A., & Shettleworth, S.J. (Eds.) (1987). *Quantitative analyses of behavior: Foraging*. Lawrence Erlbaum Associates, Hillsdale, NJ.
3. Shettleworth, S.J., 1998. Foraging and measuring rate. In: Shettleworth, S.J. (Editor), *Cognition, evolution, and behavior*, Oxford University Press, New York, pp. 379-424.
4. Stephens, D.W. and Krebs, J.R., 1986. Foraging Theory. In: Krebs, J.R. and Clutton-Brock,, T. (Eds.), *Monographs in behavior and ecology*, Princeton University Press, Princeton, NJ, pp. 1-11.
5. Whishaw, I.Q. and Gorny, B., 1999. Path integration in scent-tracking fimbria-fornix rats: Evidence for hippocampal involvement in "sense of direction" and "sense of distance" using self movement cues. *J. Neurosci*., **19** (11), 4662--4673.
6. Whishaw, I.Q. and Tomie, J.A., 1989. Food-pellet size modifies the hoarding behavior of foraging rats. *Psychobiology*, **17** (1), 93--101.

Figure 1. *An image of the foraging-hoarding apparatus discussed in this paper.*

The Object Recognition test in mice and the importance of the Object characteristics

C. Jacquot and H.J. Little
Department of Basic Medical Sciences: Pharmacology, St George's University of London, London, United Kingdom

Abstract
The Object Recognition test used to assess working memory in rodents is based on the tendency of rodents to explore novel objects. During the use of this test, we realized that slight changes to the methodology could make considerable differences to the results and, consequently, could lead to misinterpretation. The importance of these parameters is difficult to assess when reading the literature. To provide a reliable test and to decrease the variability, we examined the importance of various parameters, including object size, shape of objects, and previous handling of animals.
The aim of this presentation is therefore to provide information concerning the object recognition test in mice, based on our studies that will increase the reliability of this test.

Keywords
Memory test, object recognition, methodology, mice,

1 Introduction
The spontaneous Object Recognition test, first developed by Ennaceur in rats [1,2], is a behavioral test used to assess working memory in rodents that does not involve reinforcement. The test is based on the tendency of rodents to explore novelty. It starts with "habituation" to an arena, then a first trial or "sample" phase where the experimental animal are placed in the arena with one or two identical objects followed, after a certain delay, by a second trial or "choice" phase, in which the animal is exposed to one of the objects used during the first trial plus a novel object. A normal animal will explore more intensely the novel object, and an absence of difference in the exploration of the two objects during the choice phase is interpreted as a memory deficit.
This test is now widely used in mice, although the original methodology is usually revisited. During extensive use of this test, we realized that slight changes to the methodology could make considerable differences to the results. The importance of such changes is difficult to assess when reading the literature. In this paper we assess the importance of modifying pre-test factors (handling, objects present during habituation) and object parameters.
The aims of theses studies were to improve the acquisition of information by increasing the total exploration of the objects and to find object forms of comparable attractiveness to decrease bias and avoid misinterpretation.

2 Method

2.1 General method
All mice tested were naïve, C57 strain and homebred. Tests were carried out under dim red light, a minimum of 1 hour after the start of the nocturnal phase. Experiments were recorded on video and the time spent exploring each object analyzed *a posteriori* using the ethological analysis software, Hindsight 1.5 (developed by Dr. Scott Weiss, UK). *Exploration of an object* corresponds to the time the mouse spent with its nose directed toward the object when less than 1cm away or touching the object with the nose. Turning around or sitting on the object was not considered as exploratory behavior. N values were 9 - 18.
All results are expressed in Mean ± SEM.

2.2 Protocol
All experiments described below started on Day 1 with either two (Sections 2.3b, 2.4a, and 2.4b) or three (Sections 2.3a and 2.4c) 10min sessions of habituation to the arena (40x40x25cm).
Depending on the parameter to be assessed, on Day 2
- either each mouse was placed in the arena for a sample phase of 10 min, with two identical objects (*object recognition experiment*). The choice phase was then 15 min later, when one familiar and one novel object were present (Sections 2.3a, 3.2c). In these experiments, the Difference (*Difference Ratio*) between the time spent exploring the objects (Novel Object Expl - Familiar Object Expl) was calculated as a proportion of the total time spent exploring both objects (Novel Object Expl + Familiar Object Expl).
- or each mouse was allowed to explore the arena and 2 different objects for 10 min (*choice experiment*, sections 2.3b, 2.4a, 2.4b).

2.3 Pre-test factors

2.3a Previous handling
Whether or not mice were accustomed to being handled prior to test (decreases stress) was assessed in this section. To examine this aspect, one group of mice was handled once per day on the three days preceding the habituation (three sessions), while the other group was unhandled. On Day 2, mice were tested in an object recognition experiment.

2.3b Presence of an object in the arena during the habituation sessions
The importance of an object present during the habituation on the objects exploration during the experiment was assessed in this section. For one group of animals an object was present in the center of the apparatus, for the other group no object was present. On day 2, mice were tested in a choice experiment.

2.4 Object form

2.4a Size of objects
The attractiveness of large and small objects was tested in a choice experiment on Day 2. Mice were exposed for 10 min to two identical objects differing only by their size, large 10x10x5cm and small 6x6x3cm.

2.4b Shape of the objects
We tested, in a choice experiment, the effects of presenting objects in a different configuration, or in a

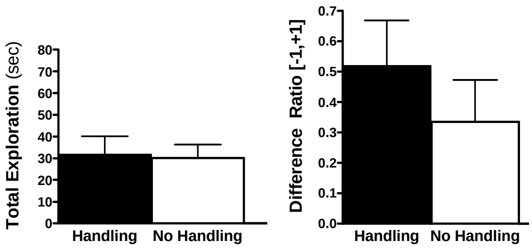
different position so that they looked different shapes. The objects used were two Lego® objects with a different shape but equivalent size (~10x10x5cm) and identical measuring cylinder plastic stands (~10x10x5cm) either upright or upside down.

2.4c Different objects

In this part, we examined the effects of the extent of differences between objects using an object recognition experiment. In each experiment, a pair of the following objects was used: Lego® vs cylinder stand, Lego® vs metal pellet holder, Lego® vs plastic beaker. The choice of Novel object was made randomly, so depending on the Novel object, mice were divided into two groups.

3 Results

3.1 Pre-test factors



(a) *(b)*

Figure 1. *Comparison of the Total Time spent exploring the objects (a) and the Difference Ratio (b) when mice that had been handled or not – Unpaired t-tests, NS*

Handling the animals before the test did not significantly affect the Total Exploration time of the objects (figure 1a) or the Difference in Exploration of the objects (figure 1b).

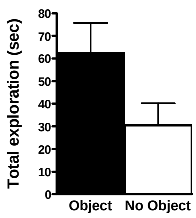


Figure 2. *Comparison of Total Time spent exploring the objects when mice had an object in the arena or not during the habituation – Unpaired t-test, NS*

Although mice that had an object in the arena during the habituation spent longer exploring the objects during the test, this was not statistically different (figure 2).

3.2 Object form

3.2a Size of objects

Larger objects were explored significantly more than smaller objects (figure 3).

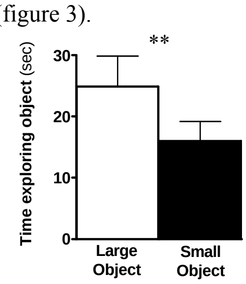


Figure 3*. Time spent exploring the larger object or the smaller object in seconds - Paired t-test, ** P<0.005*

3.2b Shape of objects

Object attractiveness was not dependent on the configuration or position (figure 4).

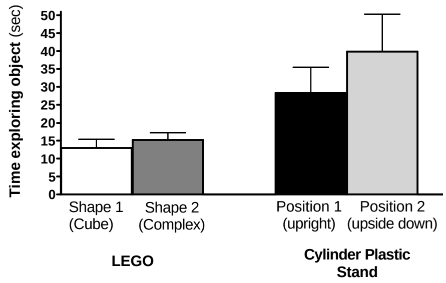
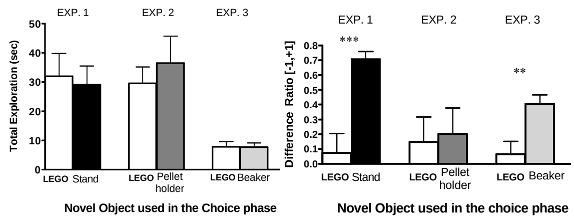


Figure 4*. Comparison of the Time spent exploring each object according to the Shape of the Objects – Paired t-tests, NS*

3.2c Different objects



(a) *(b)*

Figure 5. *(a) Total exploration time and (b) Difference Ratio - Unpaired t-tests, ** P<0.005 *** P<0.0005*

In all experiments, the Total Exploration of the objects did not depend on the Novel object type (figure 5a), but the Difference in Exploration was affected (figure 5b). The Difference Ratio for the cylinder stand as the novel object, with the familiar one made of Lego® was considerably higher than when Lego® was used as the novel object. This difference was mainly due to the difference in attractiveness of these two objects, the multifaceted shape of the cylinder stand was more attractive than the simpler-shaped Lego®. On the other hand, the metal pellet holder and Lego® with closed shape gave comparable results so would form a suitable choice.

4 Conclusion

Our results highlighted the importance of particular parameters in the Object Recognition test. Pre-test factors, such as handling the animals prior to the test, or adding an object during the habituation sessions, can slightly increase the exploration of the objects. However, the object choice plays a more important role. Use of large size objects increased the exploration of these and, with increased acquisition time, control mice then should perform better during the choice phase. It is also important, to avoid bias and misinterpretation, to choose objects of similar attractiveness for the novel and familiar object. These could be the same objects in different configuration or position, or objects with a similar shape but made of different materials.

Supported by National Institute on Alcohol Abuse and Alcoholism (NIAAA)

References

1. AbdelLab. http://abdellab.sunderland.ac.uk/. University of Sunderland - UK
2. Ennaceur, A. and Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural Brain Research*,**31,**47-59.

The Object Location memory task: a test not suitable for use in mice?

C. Jacquot and H.J. Little
Department of Basic Medical Sciences: Pharmacology, St George's University of London, London, United Kingdom

Abstract
The Object Location memory task is a behavioral test often used in rats to assess their spatial memory. During a project assessing brain damage in mice caused by chronic alcohol treatment, we used this test to see if the spatial memory of these mice was affected. As the results showed that normal mice did not display any differences in exploration between an object that had been moved and an object that was retained in its original position, we repeated it in different conditions: with another mouse strain, different arena, range of different objects, range of delay between sample and choice phase, in different light conditions and assessed by a different experimenter. None of these changes resulted in differences in control mice in exploration of an object that had been moved, compared with one that remained in the same location. Our conclusion was that this test may not be suitable to assess spatial memory in mice.

Keywords
Memory test, object location, methodology, mice

1 Introduction
The Object Location memory task is a behavioral test often used in rats to assess their spatial memory. In this task, the animals are first familiarized with the position of two identical objects during a first trial or sample phase. During the subsequent choice phase, one object stays in its previous location while the other is moved to a new location within the arena. Normal rats spontaneously spend more time exploring the object in the novel location [1].
During a project assessing brain damage in mice caused by chronic alcohol treatment, we were interested to see if the spatial memory of these mice was affected, as assessed using an object location memory task. The first results showed that control mice did not display any differences in exploration between an object that had been moved and an object that was retained in its original position. In order to examine this further, we repeated the test in different conditions in control animals.

2 Method

2.1 General method
Each experiment was divided in 3 parts:

- 1- HABITUATION

On day 1, all mice received 2 or 3 habituation sessions in the arena, each of 10 min.

- 2- SAMPLE PHASE

On day 2, each mouse was placed in the arena for a sample phase of 10 min, with two identical objects present.

- 3- CHOICE PHASE(s)

After a delay, each animal was placed back in the arena for a 5 min testing period, then returned to the home cage. During this period, one of the objects from the sample phase stayed in its original position while the other was moved to a new location. In some experiments there was a second choice phase of 5 min, with one object second moved to a location different from that in the sample phase and choice phases.

2.2 Experiments
The differences in the method and protocol of the various experiments are summarized in Table 1. N values were 8-16.

2.3 Analysis
Time (in sec) exploring each object (familiar vs. novel position) was recorded for the duration of the testing period.
Exploration of the object corresponded to directing the nose toward the object when less than 1cm away or touching the object with the nose. Turning around or sitting on the object was not considered to be exploratory behavior.
The difference (*Difference Ratio*) between the time spent exploring the objects (Explo Object Novel position - Explo Object Familiar position) was calculated as a proportion of the total time spent exploring both objects (Explo Object Novel position + Explo Object Familiar position).

3 Results

3.1 Exploration of the objects

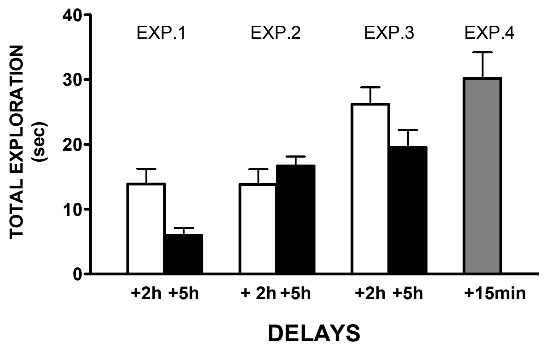


Figure 1. *TOTAL EXPLORATION of the OBJECTS during the choice phase(s) in each experiment*

Mice in all the experiments had a reasonably high level of exploration of the objects during the choice phases (figure 1). These times were similar in range to the time the mice spent exploring the objects during the sample phase (data not shown) and that for the sample time in the sibling test, the object recognition memory task [2].

3.2 Difference in exploration
In all experiments, the mice displayed no difference in exploration between the moved object and the object that

EXPERIMENT	Exp 1	Exp 2	Exp 3	Exp 4
STRAIN	TO male mice	TO male mice	TO male mice	C57 male mice
EXPERIMENTER	A - Junior	B - Experienced	B - Experienced	B - Experienced
ARENA	24 x 40 x 10 cm Transparent plastic cage	24 x 40 x 10 cm Transparent plastic cage	24 x 40 x 10 cm Transparent plastic cage	40 x 40 x 45 cm Non transparent
Cycle phase	Light	Light	Light	Dark
Test design	Mice moved to separate room 10-15min before experiments. Test under low light	Mice moved to separate room 10-15min before experiments. Test under low light	Mice moved to separate room 10-15min before experiments. Test under low light	Test in the housing room Test under dim red light
Treatment prior to sample phase	YES. Intraperitoneal injection of saline 15min before sample phase.	NO	NO	NO
Handling prior to sample phase	NO	NO	NO	YES. Mice gently handled for 3 days prior to test.
Arena cleaned between mice	Feces and urine spots removed	YES	Feces and urine spots removed	Feces and urine spots removed
Objects cleaned / copies for choice phase	NO / NO	YES / YES	YES / YES	YES / YES
Objects used	Bottle plastic cap (4.5x4.5x3.5cm)	Plastic vial (3.5x7cm)	Measuring cylinder plastic stand (10x10x5cm)	Plastic beaker (5x10cm) or Lego (5x5x5cm).
Habituation sessions	2 individual 10min sessions Object in centre of the arena	2 individual 10min sessions Object in centre of the arena	2 individual 10min sessions Object in centre of the arena	3 sessions of 10min: 1 in group and 2 individual. No object in arena.
Sample phase	10min	10min	10min	10min
Delay(s) between sample and choice phases	2h + 5h	2h + 5h	2h + 5h	15min

***Table 1**. Summary of the difference between the various experiments*

stayed in its original position (figure 2).

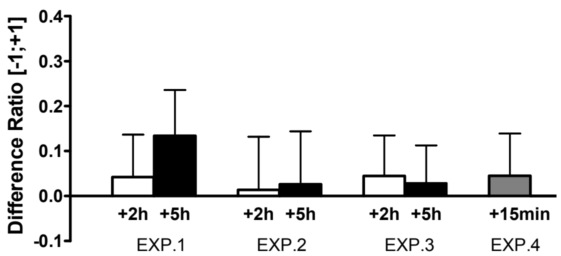


***Figure 2.** DIFFERENCE RATIO of the OBJECTS*

4 Discussion

In all the experiments, control mice did not display any difference in exploration time between the object that had been moved and the object that remained in its original position, although there was a reasonably high level of total exploration of the objects. Repetition with altered experimental parameters, a different strain, and different delays between the sample phase and the choice phase did not affect the results.

As no reference in the literature except one [3] is using this test directly adapted from Dix and Aggleton [1], we concluded that it may not be suitable for assessment of spatial memory in mice without largely revisiting the methodology as done by Roullet et al [4].

Supported by National Institute on Alcohol Abuse and Alcoholism (NIAAA)

References

1. Dix, S.L. and Aggleton, J.P. (1999). Extending the spontaneous preference test of recognition: evidence of object-location and object-context recognition. *Behavioural Brain Research*, **99,** 191-200.
2. Jacquot, C. and Little, H.J. (2005). Of the Object Recognition in mice and the importance of the Object characteristics. In: *5th International Conference on Methods and Techniques in Behavioral Research*. p.
3. Massey, P.V., Warburton, E.C., Wynick, D., Brown, M. W. and Bashir, Z.I. (2003). Galanin regulates spatial memory but not visual recognition memory or synaptic plasticity in perirhinal cortex. *Neuropharmacology*, **44,** 40-48.
4. Roullet, P., Mele, A. and Ammassari-Teule, M. (1997). Ibotenic lesions of the nucleus accumbens promote reactivity to spatial novelty in nonreactive DBA mice: implications for neural mechanisms subserving spatial information encoding. *Behavioral Neuroscience*, **111,** 976-984.

Motor-training-specific improvement of morphological and functional integration of dopaminergic intrastriatal grafts in the 6-OHDA rat model of Parkinson's disease

A. Klein[1], V. Kloth[1], A. Papazoglou[1], G. Kempermann[2], and G. Nikkhah[1]

[1] *Lab of Molecular Neurosurgery, Dept. of Stereotactic Neurosurgery, University Hospital Freiburg – Neurocentre, Freiburg, Germany*

[2] *Research group "Neural stem cells", Max-Delbrück-Centre for Molecular Medicine, Berlin, Germany.*

Abstract

Parkinson's disease (PD) is one of the most common neurodegenerative diseases and leads to severe motor and cognitive impairments. It is mainly caused by the degeneration of the dopaminergic cells within the nigrostriatal pathway. In the 6-hydroxydopamine rat model of PD we could previously show that the survival of ectopic grafts of rat embryonic day 14 neural progenitor cells, which were derived from the ventral mesencephalon, can be influenced by the extent of motor training. In the present study, we want to examine if enriched environment and/or specific motor training enhance functional recovery after transplantation of dopaminergic cells in hemiparkinsonian rats. Lesion and graft effects were evaluated by drug-induced rotation. Graft survival, graft volume, fibre density, and neurogenesis will be assessed by immunocytochemistry and bromodeoxyuridine labelling. This is an ongoing study, and further results will be presented at the conference. Our previous results and the present study may highlight the neuromodulatory effects of environmental stimuli and sensorimotor exercise on dopaminergic graft survival and function, which may ultimately lead to the optimization of clinical transplantation trials in PD.

Keywords

Skilled forelimb use, BrdU, motor behaviour, enriched environment, transplantation

1 Introduction

Parkinson's disease (PD) is linked to a hypofunctional dopamine (DA) system, which is mainly located in the nigrostriatal pathway. The development of the symptoms is generally considered to be due to a certain degree of DA loss. To substitute this loss of cells and their functions cell replacement approaches have been established in the past but -despite some promising signs of achieving recovery in animal models and clinical trials [1,4,8]- they all failed to restore and rewire the basal ganglia circuitry completely. To optimize and improve the outcome of these transplantation-experiments many attempts of *in vitro* manipulation of cells [6] have been launched in addition to *post* grafting treatment such as physiotherapy in humans [5] or motor training in experimental models [8]. It is known that physical exercise promotes adult neurogenesis and has neuroprotective effects [3]. Not only motor training but also a stimuli-enriched environment has a beneficial effect on brain development in rat experiments [10]. Besides, in our previous work (Kloth, manuscript in preparation) we could show that -after intrastriatal transplantation of neural progenitor cells in hemiparkinsonian rats- different intensities of sensorimotor training had a strong positively correlating effect on cell survival (the more training the rats performed, the better the graft survival was). This means that motor training enhances brain plasticity, and aids to change the neuronal environment that promotes DAergic graft survival. One way to mimic PD and to test motor behaviour in an experimental animal model is the complete unilateral 6-hydroxydopamine (6-OHDA) lesion model in rats. This model can imitate some of the major symptoms of PD patients (for review, see [2]). The unilateral model gives the opportunity of not only to assessing morphological data *post mortem* but also of registering and evaluating behaviour *in vivo*: the degree of lesion or of graft-related effects can be tested -as we did in this study- by drug-induced behaviour (rotational behaviour caused by lesion-induced asymmetry), or by "drug-free" behavioural analysis such as the spontaneous behaviour [9] or the staircase test [8]. Nikkhah *et al.* (1998) [8] demonstrated that VM-derived grafts could improve skilled forelimb movements in the staircase test.

In this study we chose the unilateral 6-OHDA lesion rat model, and we transplanted E14 VM-derived cells. By using the microtransplantation technique [7] a single cell suspension was stereotactically implanted into the DA-depleted striatum. Our goal in the present project is to test the effects of enriched environment and different motor training schedules on (i) graft integration into the host tissue, (ii) cell survival, (iii) proliferating processes (neurogenesis, proliferative activity within the graft), and (iv) improving effects of the motor performance itself.

2 Methods

Experimental design

Eight experimental groups of rats (n=12) were used in this study: five groups with motor training were lesioned and transplanted: 1. enriched environment-housed rats (EE rats), 2. rats trained in spontaneous behaviour, 3. rats trained in the paw-reaching-task, 4. rats trained in the paw-reaching-task to grasp pellets only with their impaired paw ("forced choice"), 5. rats that had voluntary access to a running wheel. There were three control groups without motor training: 1. healthy rats (n=8), 2. lesioned but sham-transplanted rats, 3. lesioned and transplanted rats. Training continued throughout the whole experiment. Rotational behaviour assessment was regularly performed two and six weeks after 6-OHDA lesion and after transplantation of E14 VM-derived cells. After the last training session the animals were perfused and the brains were submitted to morphological analysis (for general overview see figure 1). However, in a preceding study four different groups of rats were trained to different intensities in the staircase test and the spontaneous behaviour: rats in the Pre-Tx group were trained even before the lesion, rats in Pl-Tx group started *post lesionem*, the Ptx-Tx group *post transplantationem* and the NT-Tx group had no training until the very last training session started. As we wanted to observe learning behaviour due to repeated motor testing, which could potentially lower the improvements displayed in the transplanted groups, each group had an equivalent sham-transplanted group with the same intensity of training but no injection of E14 VM derived cells (only cell culture medium). So not the type of motor training (present study) but the time in which the rats were trained was the main focus of that preceding experiment.

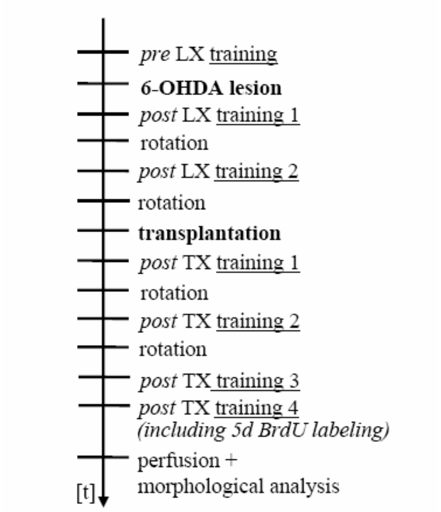


Figure 1. Experimental design (LX = lesion; TX = transplantation)

Subjects

Subjects were female Sprague-Dawley rats (Charles River, Germany), weighing 250g ± 30g at the beginning of the experiment. Rats were housed with respect to their experimental group in either enriched environment cages, or in standard cages with or without running wheel in groups of maximum 6 animals in a temperature-controlled room (23 ± 3°C) on a 12h light : 12h dark schedule. Each animal was fed with 12g standard laboratory chow (Altromin, Lage, Germany); water was freely available. Additionally, the EE rats received some biscuits and oat bran every day.

Surgery

6-OHDA lesion. All rats (except the control rats) were anaesthetized with ketamine (10%; 0.1ml/kg bodyweight; Essex, Germany) and rompune (2%; 0.01ml/kg bodyweight; Bayer, Germany), and were stereotactically injected with 6-OHDA hydrobromide (3.6µg 6-OHDA / µl in 0.1% L-ascorbic acid-saline) into the right medial forebrain bundle (MFB). Lesion coordinates were set according to bregma and dura in mm: tract 1: 3µl; tooth bar (tb) +3.4, anterior/posterior (ap) -4.0, lateral (lat) -0.8, dorso-ventral (dv) -8.0; tract 2: 2.5µl; tb -2.3, ap -4.4, lat -1.2, dv -7.8). The injection rate was 1µl/min, and the cannula was left in place for two minutes before slowly retracting it.

Transplantation. All rats (except the control rats) were stereotactically operated and anaesthetized (see above). Except for the sham-group, which received only cell culture medium, all other groups were injected with a single cell suspension of E14 VM-derived rat neural progenitor cells into the DA-depleted striatum. E14 rat embryos were dissected, and the VMs were mechanically and enzymatically digested in order to produce the single cell suspension. The cells were transplanted with a glass capillary (for review see [7,9]). Transplantation coordinates were set according to bregma and dura in mm: two tracts, four deposits (1µl each): tb 0.0, ap +1.0, lat -2.5/-3.3, dv -5/-4. The injection rate was 1µl/min and the glass capillary was slowly retracted after 30sec. The concentration of the cell suspension was 100,000 cells/µl so that each rat received 400,000 cells in total. The cell viability remained stable between 95-98% during the whole implantation procedure.

Behavioural tests

Drug-induced rotation. Drug-induced rotational behaviour was monitored in rotometer boxes two and six weeks *post lesionem* and *post transplantationem*. First apomorphine-induced rotation was tested for 40 minutes after subcutaneous injection of 1ml/kg apomorphine solution (0.05mg apomorphine + 0.2mg L-ascorbate in 1ml 0.9% saline). Three days later, amphetamine-induced rotation was tested for 90 minutes after intraperitoneal injection of 1ml/kg amphetamine solution (2.5mg D-amphetamine in 1.0ml saline). After the 6-OHDA lesion only rats that exhibited > 6.0 full body turns ipsilaterally to the lesioned side after amphetamine injection and > 4.0 full body turns contralaterally to the lesioned side after apomorphine injection have been included in this study, which meant that the lesion was successful and the animals were well prepared for cell transplantation.

Staircase test. The staircase test is a motor test to evaluate skilled forelimb movements. The modified set up of the staircase test was used as described by Nikkhah *et al.* (1998) [8]. The rats were put in a box where they could reach the food pellets from a staircase. Importantly, the rats could grasp the food pellets on the right stair only with their right forepaw and on the left stair only with their left forepaw. Right and left stairs were baited with 40 pellets each from step number 2 to 5. After a period of 15 minutes the animals were removed from the staircase box and the remaining pellets were counted. Test performance was evaluated for two different parameters: *"pellets remained"* for pellets left on the steps they had originally been placed on, and *"pellets missed"* for pellets dropped elsewhere, i.e. on steps 1 or 6 or outside the regular staircase in other parts of the staircase box. While the parameter *"pellets remained"* is considered as a measure for general reaching activity and motivation, the parameter *"pellets missed"* rather reflects the rats' forelimb skills in reaching and grasping movements, and is thus directly dependent on the rats' sensorimotor performance status. One training session consisted of nine days (d) in a row within which the staircase test was repeated: on d1-d3 the phase of *acquisition* to the test set-up, and on d4-d6 the *free choice test* with bilateral filling of the left and right steps were performed. On d7-d9 the *forced choice test* with filling the steps first of the right staircase and then of the left one was carried out. During the *forced choice test* the rats were tested twice 15min for each side/per day. There were two groups that were trained in the staircase test: one group did the experiment as described above, and the rats of other group received pellets only for their left (impaired) paws (twice 40 pellets for 15min for the left/impaired paw).

Spontaneous behaviour. Three tests were performed over a period of nine days in a row per training session: the stepping test, the postural imbalance test and the table lift test. In all three tests the forepaws were tested independently one after the other. The experimenter allowed the rats to use either the left (impaired) or the right (healthy) paw to perform the experiment. For the stepping test the rats were guided over a distance of 60cm on a regular table surface and the adjusting steps of each side were counted. On each experimental day the rats were tested twice for forehand stepping and twice for backhand stepping. During the postural imbalance test the rats were turned around their body axis five times per body side. The quality of the performance of an adjusting step was evaluated. The right or left paw had always contact with the table surface. There was a 4-point-scale: 0 = no adjusting step, 1 = some but incorrect movement, 2 = slow but less deficient movement and 3 = correctly-performed adjusting step. For the table lift test the rats were allowed to use only one paw, which was freely movable, and the rats were lifted up to the edge of a table so that only the vibrissae would touch the table. A normal control rat would immediately move its free paw to the table surface. There was a 4-point-scale for the evaluation of this behaviour: 0 = no reaction of the paw, no contact with the table, 1 = some reaction and movement, no or very late contact with the table, 2 = slow but less deficient movement, table contact, 3 = correct movement, immediate contact with the table. This test was performed five times per side every experimental day.

Enriched environment. The EE rats received no active motor training instead they were housed in cages in a weekly changing stimuli-enriched environment. Several plains and bridges, and some boxes for the rats to hide were present in big cages (1.4m x 0.7m x 0.8m). There was one running wheel per cage inside.

Running wheel. The rats of this group ("runners") were housed in standard cages with a slightly elevated lid and a running wheel inside. The rats had free access to the running wheel. In contrast to the EE rats there were no other stimuli for these rats.

Morphology

Immunohistochemistry. The rats were terminally anaesthetized and perfused through the ascending aorta with ice-cold 0.1M (pH 7.4) phosphate-buffered saline (PBS) and 4% paraformaldehyde (PFA, diluted in PBS). After post-fixation overnight in PFA the brains were dehydrated in 20% sucrose. With a freezing microtome the brains were cut in four series of coronal sections of 40µm each. One series was stained for tyrosine hydroxylase (TH) by using the free-floating TH-immunohistochemistry technique (for details see [8,9]).

Grafted cell counting. Within the graft, all cell bodies were counted under bright field illumination using a microscope (Leica, Germany) with an X-Y motor stage. With the software "stereoinvestigator" (Microbrightfield, USA) a meander-like scan through each section of the graft was performed. The total number was estimated using the Abercrombie's formula.

Fibre density. At four coordinates within the striatum (ap +1.7, +1.0, +0.5, -0.4) the optical density of TH+ immunoreactive fibres was analyzed as a measurement for graft integration and fibre outgrowth. The non-lesioned (healthy) side served as control side. With the exception of the graft core (with its cell bodies) the whole striatal area per section was evaluated under bright field condition using an Olympus AX70 microscope and a self-programmed macro in the frame of the analySIS software (Soft Imaging System, Germany). Data were collected from medial and lateral striatum.

Graft volume. In order to measure the graft volume whole-graft pictures were taken under bright field illumination with Olympus AX70 microscope and analySIS software. The borders of the graft core (separated in a medial and lateral part) were marked on every section where cell bodies of DAergic neurons were found. Then the software calculated the area in µm^2. The result of one section was multiplied by the original thickness of the section (40µm) and by the number of series (4). In the end the results of all sections were added and the total graft volume in µm^3 was calculated.

BrdU labelling. Bromodeoxyuridine (BrdU) (Sigma, St. Louis, USA) was dissolved in 0.9% saline. All rats received a daily dose of 50mg/kg body weight intraperitoneally for five days running.

Statistical analysis

For statistical evaluation, the data were subjected to one-factor ANOVA followed by Student-Neuman-Keuls post hoc test (StatView 4.5, Abacus Concepts Inc., Berkeley, USA) with p set at < 0.05 as level of significance. Results are expressed as means ± standard error of the mean (SEM).

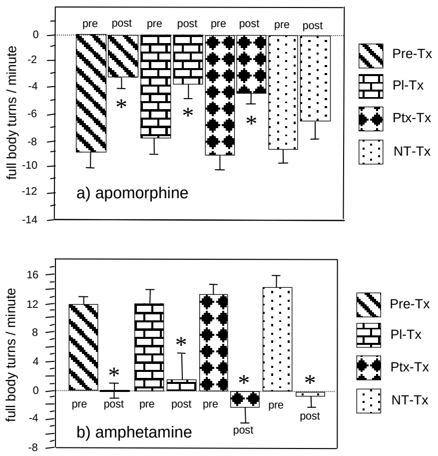


*Figure 2. Drug-induced rotation before ("pre") and after ("post") the transplantation for apomorphine (a) and amphetamine (b) injection. Except for the apomorphine rotation of the NT-Tx group all "post" transplantation data are significant different to the "pre" transplantation data (marked for all groups with *).*

3 Results

As this is an ongoing study we can present data only from the current status of the experiment. So far, in the present study all rats have been lesioned, transplanted, and rotated. All four *post transplantationem* training sessions have been carried out including the BrdU-labelling. The rats are ready for perfusion, and the morphological analysis will have been finished by the time the conference starts. The full project will be presented there. Figures 2 and 3 briefly show the most important results of the preliminary project where not the type of motor training (present study) but the duration of exercise during which the rats were trained was the main focus of that experiment.

Drug-induced rotation. In order to test *in vivo* if the lesion and the transplantation were successful, drug-induced rotation was performed. As shown in figures 2 a and b the grafts had a dramatic effect on the rotational behaviour, but no correlation between reduction of rotation and more motor training could be found. The results of the drug-induced rotation five weeks *post* lesion and ten weeks *post* transplantation are shown after apomorphine (figure 2a) and amphetamine (figure 2b) injection. The control rats did not rotate at any time, and the lesioned but sham-transplanted groups (with equivalent training to their corresponding transplanted group) remained on a stable level with no rotation-reducing effect (data not shown). Under apomorphine administration (figure 2a) the rats rotated per minute [rpm] at the following levels: (i) before the transplantation ("*pre*"): Pre-Tx -9.1rpm ±1.1SEM, Pl-Tx -8.4rpm ±0.1SEM, Ptx-Tx -9.2rpm ±1.0SEM, and NT-Tx -8.7rpm ±1.5SEM. (ii) ten weeks after the transplantation ("*post*") the rotational values were significantly reduced (except for the NT-Tx group; $p<0.05$): Pre-Tx -3.7rpm ±1.0SEM, Pl-Tx -3.9rpm ±0.9SEM, Ptx-Tx -4.6rpm ±0.8, and NT-Tx -6.0rpm ±1.1SEM. In figure 2b the results of the amphetamine-induced rotation are shown: (i) before the transplantation ("*pre*"): Pre-Tx 11.4rpm ±1.0SEM, Pl-Tx 11.9rpm ±1.5SEM, Ptx-Tx 12.8rpm ±1.3SEM, and NT-Tx 13.8rpm ±1.4SEM. (ii) ten weeks after the transplantation ("*post*") the rotational values were significantly reduced, or the rats even overcompensated to the other side ($p<0.05$): Pre-Tx 0.0rpm ±1.6SEM, Pl-Tx 1.0rpm ±3.2SEM, Ptx-Tx -2.6rpm ±2.5SEM, and NT-Tx -0.3rpm ±1.5SEM.

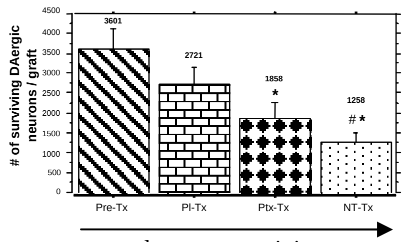


Figure 3. Effect of different amount of motor training on survival of DAergic neurons within the graft. Graft survival was enhanced in Pre-Tx compared to Ptx-Tx and NT-Tx (*) and Pl-Tx compared to NT-Tx (#).

Grafted cell counting. Figure 3 demonstrates the effect of motor training on the number of surviving DAergic neurons within the graft. TH+ cell bodies were counted in the sections where a transplant could be found. The total number was calculated with Abercrombie's formula. The results show that the more training was performed the more cells survived: Pre-Tx 3601 TH+ neurons ±519SEM, Pl-Tx 2779 TH+ neurons ±447SEM, Ptx-Tx 1858 TH+ neurons ±388SEM, and NT-Tx 1245 TH+ neurons ±334SEM.

4 Discussion and Conclusions

The 6-OHDA lesion leads to severe motor impairments in rats. Transplantation of E14 VM-derived DAergic neurons has clear beneficial effects on these motor deficits. The main focus of the present study is to find out if there are specific motor tests which promote best the functional recovery, or if general exercise is sufficient for convalescence. Together with our previous results the outcome of this study may be taken into consideration for future clinical trials and prove that physical exercise may be beneficial for the success of cell replacement strategies.

Supported by Deutsche Forschungsgesellschaft (DFG) Graduate School 843 "Mechanisms of Neuronal Signaltransduction" and DFG Ni 330/4-2.

The authors want to thank Johanna Wessolleck and Brunhilde Baumer for excellent and outstanding technical support.

References:

1. Bjorklund,A., Dunnett,S.B., Brundin,P., Stoessl,A.J., Freed,C.R., Breeze,R.E., Levivier,M., Peschanski,M., Studer,L. & Barker,R. (2003) Neural transplantation for the treatment of PD. *Lancet Neurol.*, **2**, 437-445.
2. Cenci,M.A., Whishaw,I.Q. & Schallert,T. (2002) Animal models of neurological deficits: how relevant is the rat? *Nat.Rev.Neurosci.*, **3**, 574-579.
3. Dobrossy,M.D. & Dunnett,S.B. (2004) Environmental enrichment affects striatal graft morphology and functional recovery. *Eur.J.Neurosci.*, **19**, 159-168.
4. Freed,C.R., Greene,P.E., Breeze,R.E., Tsai,W.Y., DuMouchel,W., Kao,R., Dillon,S., Winfield,H., Culver,S., Trojanowski,J.Q., Eidelberg,D. & Fahn,S. (2001) Transplantation of embryonic dopamine neurons for severe Parkinson's disease. *N.Engl.J.Med.*, **344**, 710-719.
5. Keus,S.H., Bloem,B.R., Verbaan,D., de Jonge,P.A., Hofman,M., van Hilten,B.J. & Munneke,M. (2004) Physiotherapy in Parkinson's disease: utilisation and patient satisfaction. *J.Neurol.*, **251**, 680-687.
6. Kim,J.H., Auerbach,J.M., Rodriguez-Gomez,J.A., Velasco,I., Gavin,D., Lumelsky,N., Lee,S.H., Nguyen,J., Sanchez-Pernaute,R., Bankiewicz,K. & McKay,R. (2002) Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature*, **418**, 50-56.
7. Nikkhah,G., Olsson,M., Eberhard,J., Bentlage,C., Cunningham,M.G. & Bjorklund,A. (1994) A microtransplantation approach for cell suspension grafting in the rat Parkinson model: a detailed account of the methodology. *Neuroscience*, **63**, 57-72.
8. Nikkhah,G., Rosenthal,C., Falkenstein,G. & Samii,M. (1998) Dopaminergic graft-induced long-term recovery of complex sensorimotor behaviors in a rat model of Parkinson's disease. *Zentralbl.Neurochir.*, **59**, 97-103.
9. Rodter,A., Winkler,C., Samii,M. & Nikkhah,G. (2000) Complex sensorimotor behavioral changes after terminal striatal 6-OHDA lesion and transplantation of dopaminergic embryonic micrografts. *Cell Transplant.*, **9**, 197-214.
10. Van Praag,H., Kempermann,G. & Gage,F.H. (2000) Neural consequences of environmental enrichment. *Nat.Rev.Neurosci.*, **1**, 191-198.

The Concentric Square Field: a multivariate test arena for ethoexperimental analysis of explorative strategies

Meyerson, B.J.[1], Augustsson, H.[2], Berg, M.[1, 3], Roman, E.[3], Nylander, I.[3]
[1] *Department of Neuroscience, Uppsala University, Uppsala, Sweden*
[2] *Department of Animal Environment and Health Section of Animal Welfare, Swedish University of Agricultural Sciences, Uppsala*
[3] *Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden*

Abstract
The Concentric Square Field (CSF) is a multivariate test arena that includes sheltered, open and elevated areas, a hole board, areas of different light conditions and wall enclosed corridors. The model aims to provoke explorative strategies with respect to risk assessment and risk taking performance.

Keywords
ethoexperimental, multivariate, rat, principal component analysis (PCA).

1 Introduction
The Concentric Square Field (CSF) contains open areas, sheltered space, enriched areas to explore, zones of different lightning conditions and elevated versus ground floor regions. The various items included in the physical setting should, as far as laboratory conditions allow, have relevance with regard to what the animal might meet in its natural habitat. The model allows for seeking shelter versus exploring the surrounding on a risk and benefit basis. Validation of risky versus safe areas was based on the retrieval behavior in lactating females. We here report on the performance of three strains of laboratory rats: Wistar (WI), Lister Hooded (LIHO) and Sprague Dawley (SD). The method has earlier been applied in an attempt to relate the localization and extent of brain lesions to functional effects [3, 6], in comparative studies of explorative strategies of wild and laboratory mice [1, 2] and to study effects on the adult behavior in male rats after early maternal separation [5]. Further data on performance over time, strain differences, effects of various stressors, repeated trials, sex differences in the laboratory rat are in progress [4].

2 Methods

2.1 The Concentric Square Field (CSF)
The apparatus (Fig.1) consisted of a square field (72x72 cm) surrounded by an outer wall (25 cm high) and subdivided by inner walls into one central squared arena (CENTRE, 40x40 cm) and peripheral corridors (width 15 cm). Access to the corridors from the CENTRE was by openings (A 10 cm or corridors B, C circular opening φ= 8cm) located in the middle of the surrounding walls. One of the corner areas of corridor A/B was closed off by walls and covered with a PVC lid, thereby providing a dark corner room (DCR, 18x18 cm). A wall closed off the other end of corridor A. In the corner of corridors B/C was provided another corner room the HURDLE (18x18cm) to which the animal had access by a hole (φ= 8cm) situated 10 cm (entrance from corridor B) or 7 cm (entrance from corridor C) above the floor. The floor of the HURDLE holds a hole board (two holes φ= 2.5 cm) [4]. A photocell device (photocell 2) located under the floor provided recording of head dips into the holes. A stainless steel wire mesh construction (15 mm between bars) that bridged over an illuminated opening in the floor was situated in corridor D (BRIDGE). The BRIDGE started and ended by a slope ascending at an angle of 45 degrees from the arena floor (SLOPE). Between the SLOPE and BRIDGE areas was located a photocell device (photocell 1). This arrangement provided measures of the activity on the border of the BRIDGE entrance. The floor and walls of the CSF were built in black PVC (4mm thick) except for the outer wall of the BRIDGE that was made of clear Plexiglas. Approximate light conditions (lux): DCR: <1, CENTRE: 30, CORRIDORS: 10- 15, BRIDGE: 125-190.

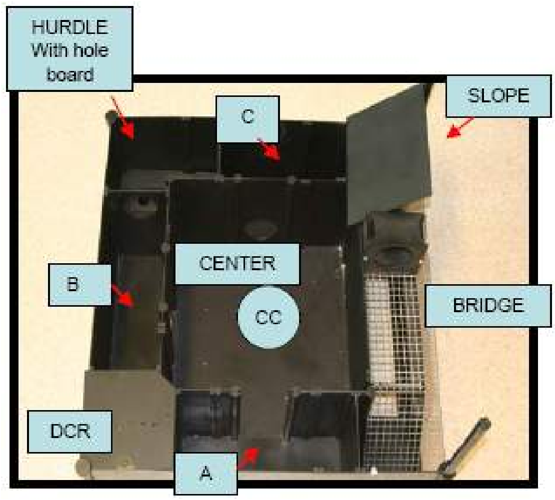


Figure 1. The Concentric Square Field

2.2 Recording
The observation arena was divided into 11 zones (some of them illustrated in Fig. 1), which constituted the basis for the dependent variables i.e. the animal's performance. The behavior was videotaped (Panasonic BP550 camera, AG-TL 300E VHS recorder). The analyses were based on direct observation using the software SCORE (Pär Nyström, Copyright Soldis), the photocell device records and the processing by the Ethovision system (Noldus Inc, The Netherlands version 2:3). Statistical comparisons were perfomed using Stat Soft Statistca 7 and for the principal component analysis (PCA) we used SIMCA-P+10.02 software, Umetrics.

3 Results

3.1 Establishing areas of risk versus safety

In designing the CSF apparatus, the assumption was that the animal would consider the elevated and illuminated BRIDGE as an area of risk whereas the DCR possessed qualities of safety. Retrieving: Five female rats with 10-13 pups were tested 6-8 days after delivery. The female to be tested was removed from the home cage to the transport bucket. The pups were then, still in their home cage, brought to the adjacent laboratory and placed on cellulose cotton at the end of the BRIDGE opposite to the entrance slope. The female was placed in the CENTER zone of the CSF. After visiting some of the CSF zones the female soon located the pups and immediately started to retrieve them. She carried the pup directly to the DCR area. In all cases the place to which the pups were brought was the DCR. The latency of first retrieval was 85 s. An analogous experiment was carried out with another two females but now the pups were placed in the DCR. The initial explorative activity was about the same but no retrieving occurred.

Conclusion: Taken together these results are the reason for considering the BRIDGE as an area that the animal reacts to as of risk. Relative to the BRIDGE the DCR implies an area of security.

3.2 Strain differences

PARAMETERS	WI			SD		
LAT LEAVE	24	±	11	10	±	3
LAT CTR CIRCLE	425	±	77**	169	±	68
LAT DCR	232	±	70	58	±	25
LAT BRIDGE	318	±	93	343	±	111
LAT SLOPE	245	±	81	253	±	68
LAT HURDLE	287	±	80	184	±	77
TOTAL ACTIVITY	127	±	14	171	±	20
CORR. ENTRIES	20	±	3*	33	±	4
FRQ CENTRE	22	±	3*	37	±	4
FRQ CTR CIRCLE	7	±	2	11	±	2
FRQ DCR	10	±	1	11	±	1
FRQ BRIDGE	3	±	0,6	4	±	1
FRQ SLOPE	6	±	2	9	±	1
FRQ HURDLE	4	±	1	6	±	1
DUR CTR CIRCLE	7	±	2	8	±	1
DUR DCR	334	±	70	196	±	29
DUR BRIDGE	77	±	19	111	±	26
DUR SLOPE	89	±	18	150	±	30
DUR HURDLE	88	±	14	76	±	6
PHOTOCELL 1	37	±	9	47	±	8
PHOTOCELL 2	9	±	1	8	±	2
DUR/VISIT DCR	45	±	19	19	±	3

Table 1. Behavioral response of male Sprague-Dawley (SD) and Wistar (WI) rats in the CSF. Data are presented as mean ± S.E. M, *p<0.05 ** p<0.01,(Mann-Whitney U test).

WI, LIHO (data not shown) and SD male rats differed in TOTAL ACTIVITY (visits to all zones), CORRIDOR ENTRIES and visits to the CENTRE (Kruskal-Wallis oneway analysis of variance: H=9.70 p=0.02, H = 9.82 p=0.007 H=12.7 p=0.002). WI was less active than the two other strains. The WI group had a longer latency for visit to CENTRAL CIRCLE (H=7.7 p=0.02) than the SD and LIHO rats. Latencies to other zones did not differ. WI stayed longer and LIHO shorter time per visit in the DCR than the SD and moved (velocity cm/sec.) more slowly than the SD males in the CENTRAL CIRCLE zone (p<0.05). The distance moved (cm/20 min) by the WI males was less than for the SD (p<0.001). A PCA score plot distribute the animals in different regions around the origo in relation to strain affiliation. We conclude that the WI is less exploring and differs in its approach and avoidance pattern as to open areas (CTRCI) from the SD and LIHO rats.

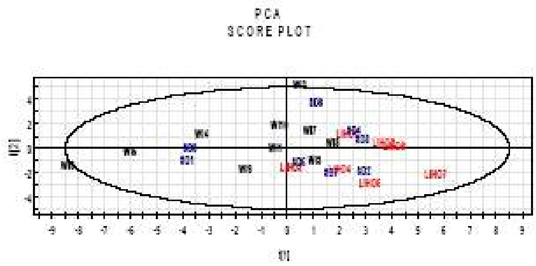


Figure 2. PCA chart showing Score plot for Sprague Dawley (SD), Wistar (WI) and Lister Hooded (LIHO) male rat performance in the CSF.

4 General conclusions

The CSF is a multivariate test arena mainly aimed for analysis of explorative strategies and studies on risk assessment and risk taking. The variety of measures also provides means to access changes in perceptive, cognitive or locomotor functions.

References

1. Augustsson, H.; Dahlborn, K.; Meyerson, B.J. (2005). Exploration and risk assessment in female wild house mice (*Mus musculus musculus*) and two laboratory strains. *Physiol Behav*, **84**, 265-277.
2. Augustsson, H.; Meyerson, B.J. (2004). Exploration and risk assessment: a comparative study of male house mice (*Mus musculus musculus*) and two laboratory strains. *Physiol Behav*, **81**, 685-698.
3. Clausen, F.; Meyerson, B.J.; Hillered, L. (2001) A novel tool for analysing functional outcome in rodents. *Society for Neuroscience*, San Diego, CA, USA.
4. Meyerson, B.J.; Augustsson, H.; Berg, M.; Roman, E.; Nylander, I. (2005) The concentric Square Field: a multivariate test arena for analysis of explorative strategies. *Manuscript to be submitted.*
5. Roman, E. (2004) Maternal separation in rats, *Acta Universitatis Upsaliensis, Comprehensive summaries of Uppsala Dissertations from the Faculty of Pharmacy*, **313**, 1-81.
6. Roos, M.W.; Ericsson , A.; Berg, M.; Sperber, G.O.; Sjöquist, M.; Meyerson, B.J. (2003) Functional evaluation of cerebral microembolization in the rat. *Brain Res*, **961**, 15-21.

The Concentric Square Field test for monitoring behavioral profiles in AA and ANA rats

Roman E.[1], Berg M.[1], Hyytiä P.[2], Meyerson B.J.[3], Nylander I.[1]
[1] *Department of Pharmaceutical Bioscience, Division of Pharmacology, Uppsala University, Uppsala, Sweden*
[2] *Department of Mental Health and Alcohol Research, National Public Health Institute, Helsinki, Finland*
[3] *Department of Neuroscience, Division of Pharmacology, Uppsala University, Uppsala, Sweden*

Abstract
The ethanol-preferring AA and ethanol-avoiding ANA rats are selectively bred for high and low voluntary ethanol intake, respectively. Besides the differences in voluntary ethanol intake there exists behavioral differences where the AA rats are considered to be less emotionally reactive. The aim of the present experiment was to compare the behavioral profiles in these two lines of rats using three different test methods.

Keywords
ethoexperimental, multivariate, open field, elevated plus-maze

1 Introduction
The ethanol-preferring AA (Alko, Alcohol) and ethanol-avoiding ANA (Alko, Non-Alcohol) rats are selectively outbred for high and low voluntary ethanol intake, respectively [2, 3]. Besides the differences in voluntary ethanol intake there exists behavioral differences where the AA rats are considered to be less emotionally reactive compared to ANA rats [e.g. 6, 7].

The Concentric Square Field (CSF) test provides multiple areas for the animal to explore including sheltered, open and elevated areas. The idea underlying this test is that the various areas should provide different elements of risk versus shelter and explorative incentives enabling measurements of explorative strategies with respect to risk assessment and risk taking [4, 5].

2 Aim
We here compare the data achieved from the multivariate test arena, i.e. the CSF, and the more traditional Open Field (OF) test and the Elevated Plus-Maze (EPM) test.

3 Methods

3.1 Animals
Twelve adult male AA and ANA rats, respectively, were supplied from the National Public Health Institute, Helsinki, Finland. The AA rats weighed 225-320 g and the ANA rats weighed 220-390 g at the first day of testing.

3.2 Behavioral tests
The test battery consisted of the CSF test as well as the OF and EPM tests. The rats did not have previous experience of the test arenas prior to testing. Each animal was run in the three behavioral tests during three consecutive days for a period of two weeks starting with the CSF, then the OF and finally the EPM test. The sequential order of the three tests was based on a pilot study in mice using the same battery of behavioral tests [1]. The rats in each experimental group were tested using a running schedule in order to avoid time and order bias. Observations were carried out during the dark period of the light/dark cycle.

3.3 The CSF test
The method provides multiple areas for the animal to explore with different physical design. The arena has been described in detail elsewhere [4, 5]. The test sessions were run in dimmed light and lasted for 20 min.

3.4 The OF test
The OF arena consisted of a black circular stainless steel arena with a diameter of 90 cm surrounded by 35 cm high walls and a black stainless steel wire mesh floor (10 mm between the bars). Two cages were attached to the wall opposite to each other. One cage was empty and wire bars blocked access to this cage. The other cage was used as a start box (25x25x25 cm with an opening ∅=8 cm). The rat was transferred from its home cage and placed in the start box. Thereafter the test session started and the animal was free to explore the arena with the possibility to return to the start box. The arena was divided into zones as shown in Figure 1. The test sessions were run in dimmed light and lasted for 30 min.

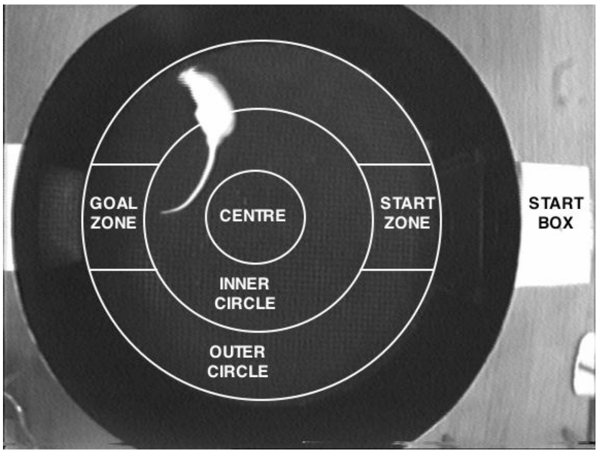


Figure 1. *The Open Field test with its defined zones.*

3.5 The EPM test
The EPM test began with each rat being placed in the centre facing an open arm, after which the experimenter left the room. Each rat was tested for 10 min and dimmed light was used during the testing.

3.6 Behavioral scoring
The behavior in the three different tests was videotaped. The analyses were based on direct observation using the software SCORE (Pär Nyström, Copyright Soldis). The latency (LAT) to first visit, the frequency (FRQ) of visits and duration (DUR) of visits were measures taken for each defined zone. The total number of frequencies in each test was used as a measure of TOTAL ACTIVITY. The CSF

arena was divided into nine zones that constituted the basis for the performance of the animal. The OF arena was divided into six zones (Figure 1) and finally the EPM was divided into open and closed arms as well as centre. To consider the animal being in a certain zone, both of the animals' hind legs must have crossed over into that section.

4 Results

4.1 The CSF test

The main findings from the CSF test are summarized in Table 1.

PARAMETERS	ANA			AA		
TOTAL ACTIVITY	138	±	9	79	±	2****
LAT CENTRAL CIRCLE	145	±	39	268	±	75
LAT DCR	37	±	14	117	±	24**
LAT BRIDGE	201	±	57	344	±	68
LAT HURDLE	165	±	49	183	±	19
OCCURRENCE BRIDGE	8			12		
DUR/VISIT CENTRAL CIRCLE	2	±	1	1	±	0.3
DUR/VISIT DCR	19	±	3	22	±	2
DUR/VISIT BRIDGE	29	±	6	68	±	4**
DUR/VISIT HURDLE	16	±	2	29	±	2***

Table 1. *Behaviors recorded during 20 min in the Concentric Square Field (CSF) test in male AA and ANA rats. Values represent mean ± SEM. **p < 0.01, ***p < 0.001, ****p < 0.0001 (Mann-Whitney U test).*

4.2 The OF test

The main findings from the OF test are summarized in Table 2.

PARAMETERS	ANA			AA		
TOTAL ACTIVITY	246	±	15	172	±	7***
LAT OUTER CIRCLE	7	±	1	44	±	6****
LAT INNER CIRCLE	28	±	7	81	±	20**
LAT CENTRE CIRCLE	116	±	22	132	±	26
DUR/VISIT OUTER CIRCLE	10	±	1	8	±	1
DUR/VISIT INNER CIRCLE	3	±	0.2	6	±	1***
DUR/VISIT CENTRE CIRCLE	2	±	0.2	3	±	0.5

Table 2. *Behaviors recorded during 30 min in the Open Field test in male AA and ANA rats. Values represent mean ± SEM. **p < 0.01, ***p < 0.001, ****p < 0.0001 (Mann-Whitney U test).*

4.3 The EPM test

The results from the EPM test are summarized in Table 3.

PARAMETERS	ANA			AA		
TOTAL ACTIVITY	36	±	3	29	±	2*
LAT OPEN ARMS	8	±	3	60	±	14**
LAT CLOSED ARMS	12	±	4	16	±	5
DUR/VISIT OPEN ARMS	9	±	1	15	±	2
DUR/VISIT CLOSED ARMS	38	±	3	39	±	4
DUR/VISIT CENTRE	3	±	0.4	11	±	2****
OCCURRENCE OPEN ARMS	10			11		

Table 3. *Behaviors recorded during 10 min in the Elevated Plus-Maze test in male AA and ANA rats. Values represent mean ± SEM. *p < 0.05, **p < 0.01, ****p < 0.0001 (Mann-Whitney U test).*

5 Conclusion

Taken together, results from the three tests showed that the ANA rats were more active. Furthermore, the ANA rats generally had shorter latencies to the first visit of a certain zone while the AA rats generally had longer duration per visit. The multivariate CSF test provided a more comprehensive analysis with special regard to approach/avoidance patterns associated with risk assessment and risk taking. The ANA rats were characterized by a higher explorative activity and differed clearly from the AA males in their approach/avoidance pattern as to open versus sheltered areas indicating a difference between the two lines in risk assessment and risk taking strategies. From a methodological point of view it is concluded that the CSF test arena provides an interesting complement to the OF and EPM tests.

Supported by grants from the Facias Foundation (ER), AFA, the Alcohol Research Council of the Swedish Alcohol Retailing Monopoly and the Swedish Medical Research Council (IN).

References:

1. Augustsson, H. (2004). Ethoexperimental studies of behaviour in wild and laboratory mice. *Acta Universitatis Agriculturae Sueciae, Veterinaria*, **174**, 7-62.
2. Eriksson, K. (1968). Genetic selection for voluntary alcohol consumption in the albino rat. *Science*, **159**, 739-741.
3. Hilakivi, L.; Eriksson, C.J.P.; Sarviharju, M.; Sinclair, J.D. (1984). Revitalization of the AA and ANA rat lines: effects on some line characteristics. *Alcohol*, **1**, 71-75.
4. Meyerson, B.J.; Augustsson, H.; Berg, M.; Roman, E.; Nylander, I. (2005). The concentric Square Field: a multivariate test arena for analysis of explorative strategies. Manuscript to be submitted.
5. Meyerson, B.J.; Augustsson, H.; Berg, M.; Roman, E.; Nylander, I. (2005). The Concentric Square Field: a multivariate test arena for ethoexperimental analysis of explorative strategies. *Measuring Behavior*, Wageningen, The Netherlands.
6. Möller, C.; Wiklund, L.; Thorsell, A.; Hyytiä, P.; Heilig, M. (1997). Decreased measures of experimental anxiety in rats bred for high alcohol preference. *Alcohol Clin Exp Res*, **21**, 656-660.
7. Tuominen, K.; Hilakivi, L.A.; Päivärinta, P.; Korpi, E.R. (1990). Behavior of alcohol-preferring AA and alcohol-avoiding ANA rat lines in tests of anxiety and aggression. *Alcohol*, **7**, 349-353.

A refined behavioural analysis for the genetic dissection of food exploration strategies in mice

J.G. Mooij-van Malsen[1], B. Olivier[1,2], B.M. Spruijt[3], J..M. van Ree[1] and M.J.H. Kas[1]
[1]*Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, The Netherlands*
[2]*Department of Psychopharmacology, University Utrecht, The Netherlands*
[3]*Department of Animal, Science and Society, University Utrecht, The Netherlands*

Abstract

In nature, an animal establishes a balance between finding new resources and protecting itself. We introduce a novel behavioural phenotyping method that dissects the composition of exploration strategies by assessing behaviours directed at obtaining food (approach) and the innate preference to shelter (avoidance), regardless of motor activity levels. Combined with the Ethovision® video-tracking system (PhenoTyper® PT10S/P /N Version 1.01, Noldus Information Technology, Wageningen) mice are monitored for several days and novelty-induced, baseline and circadian behavioural variation can be studied without experimenter interference. In our initial studies, distinct behavioural phenotypes were found, affecting mouse behaviour in either motor activity levels or shelter preference. The data shows that continuous multi-day recordings in designed environments can dissociate complex behaviours into refined behavioural components that involve physiological processes with different genetic origin. Potentially, characterization of a panel of chromosome substitution strains (CSS) in this modified home cage environment (MHE) introduces a powerful genetic strategy for identifying quantitative trait loci affecting different behavioural processes.

Keywords

Avoidance, exploration, modified home cage environment, multi-day recording, chromosome substitution strains

1 Introduction

Innate survival strategies of animals depend on the natural balance between different behavioural strategies, such as approach and avoidance behaviour. These strategies evolve as response to the naturally conflicting motivations related to fear and exploration the environment produces. Conventional laboratory tests for exploration strategies in rodents reveal marked behavioural differences between inbred strains of mice, reflecting genetic variation. Tests like the elevated plus maze, open field and light/dark box are commonly employed to study avoidance and motor activity levels of animals in novel environments. However, the nature of these tests makes it hard to interpret the behavioural variation observed in different mouse strains (Kas, M.J. and Van Ree, J.M., 2004). The individual locomotor activity levels, short duration of the experiment, experimenter interference and novelty responsiveness of the animal can interfere with the identification of genotype – behavioural phenotype relationships.

In order to find quantitative trait loci (QTL) and possibly pathways in the brain involved in avoidance and approach behaviour, these behavioural strategies should be studied within the overall observed behaviour. Monitoring the animal in a designed living environment can be used to dissect the complex behaviours involved in exploration strategies and monitor the separate components simultaneously. Therefore, we developed a Modified home cage environment (MHE). In the MHE, the animal can be monitored for avoidance and approach behaviour for several days without experimenter interference. By evaluating three days of continuous recording, measures for novelty as well as baseline behaviour can be obtained. A shelter is provided for sheltered sleeping. In order to eat, the animal can choose between different food resources.

Different inbred strains show marked differences in their behavioural expression. For example, avoidance behaviour is found to be strain dependent, reflecting the genetic influence on this behaviour. Many investigations have shown that C57BL/6 mice display low levels of anxiety and the A/J strain to be relatively anxious (Reviewed by Crawley, J.N. et al., 1997). Most of these behavioural experiments however, are mainly based on locomotor activity during exploration of a novel environment. However, novelty responsiveness levels are expressed by comparing one strain to the other, not taking into account possible differences in baseline or circadian variation between strains. These strain differences in avoidance/approach behaviours can contribute to the identification of biological substrates underlying complex behaviours. For example, mouse genetic mapping panels for these strains have recently been designed.

With our modified home cage environment combined with novel genetic strategies, we will attempt to show distinct phenotypes by dissecting the behaviour in refined behavioural phenotypes, taking into account novelty-induced and baseline behaviour, as well as circadian variation in behaviour as a function of genetic background.

2 Materials and methods

2.1 Subjects

Male C57BL/6J mice were housed in groups, with tap water and chow (CRM(E), Special Diets Services, England) provided *ad libitum*. A minimum of two weeks prior to the start of the experiment the animals were moved from the stables to the adaptation room next to the experimental room with a shifted 12:12 hr dark: light cycle (lights off at 1 PM).

2.2 Experimental procedure

A modified home cage environment (MHE) was used in these experiments, containing a shelter, drinking spout and two different feeding platforms. One hour prior to the dark period, C57BL/6j male mice (10-14 weeks old, n = 13) were weighed and placed individually in the MHE for 73 hours. No experimenter interference occurs during these three days. At the end of the experiment body weight change and food- and water intake were measured. The distance moved, enter frequency and duration spent in the several zones were registered using the video-tracking system (PhenoTyper® PT10S/P /N Version 1.01, Noldus

Information Technology, Wageningen) (Visser, L et al., 2005) and the data were analysed using Ethovision® (Noldus, L.P. et al., 2001).

Infrared sensors placed in front of the food- and water supplies measured frequency and duration of feeding and drinking.

3 Preliminary results

3.1 Movement

An increase in percentage movement was seen during the first day in the MHE compared to the second day (p=0.0001; figure 1), reflecting a hyperactive response to the novel environment. A higher level of movement during the dark phases compared to the light phases illustrates the circadian rhythm of this nocturnal species.

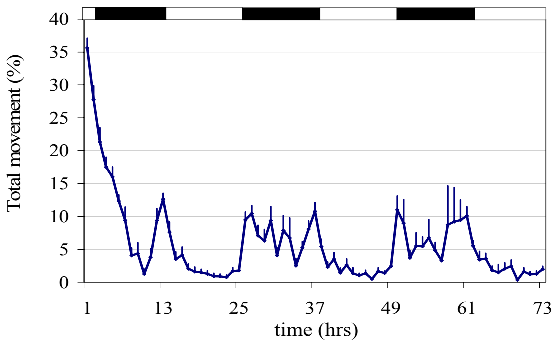


Figure 1. *Percentage movement per hour. Graph illustrates the movement of C57BL/6 male mice (n=13) during three days of testing. Light and dark phases are indicated by black (dark) and white (light) bars at the top of the graph.*

3.2 Drinking behaviour

Drinking activity shows similar circadian variation as observed for movement (figure 2). Contrary to the movement of the animal, no novelty effects were observed in drinking behaviour.

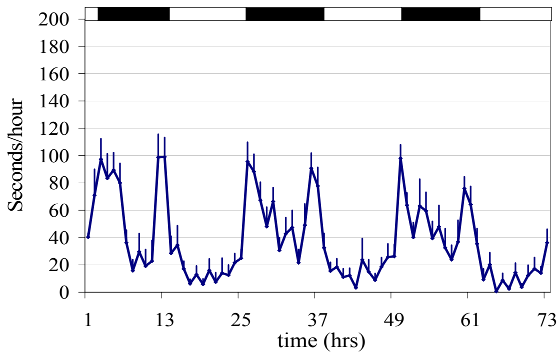


Figure 2. *Duration of drinking per hour. Graph illustrates recording of drinking behaviour of C57BL/6 male mice (n=13) during three days of testing. Light and dark phases are indicated by black (dark) and white (light) bars at the top of the graph.*

4 Discussion

Our modified home cage environment can measure novelty, baseline and circadian variation in behaviour. In this way, we are able to show distinct phenotypes based on several parameters and measurements. Locomotor activity and the exploration of resources can be dissected and compared. Shelter preference of different mouse strains unconfounded by locomotor activity is expected to be a refined measurement for avoidance behaviour.

Our next goal is to find genes underlying food exploration strategies and avoidance behaviour. Several chromosomes have been proposed to contain QTLs involved in anxiety-related behaviours. Repeatedly, chromosome 1, 4 and 15 are reported to have an effect on anxiety related behaviour (Henderson, N.D. et al., 2004; Turri, M.G. et al., 2001). Recently, a novel method to examine the genetic influences on behaviour has been introduced (Nadeau, J.H. et al.2000). A panel of chromosome substitution strains (CSS) was developed, derived from the C57BL/6J host strain and the A/J donor strain. Each of these substitution strains has a single chromosome from the donor strain substituting for the corresponding chromosome in the host strain. This has proven to be a powerful way to detect QTL's on the mouse chromosome (Singer, J.B. et al., 2004; Singer, J.B. et al., 2005). Concurrent with the mentioned results, using the sensitive mapping CSS-approach, Singer *et al.* found evidence for QTLs on chromosome 1,4, and 15, but also on chromosome 6 and 17 affecting anxiety-related behaviours demonstrated in an open field and a light-dark box (Singer, J.B. et al., 2004; Singer, J.B. et al., 2005).

Based on the genetic differences between the host and donor strain and the effects of substitution of chromosomes in a different genetic background, we expect to find multiple gene regions differentially regulating behaviour. By combining the CSS and the modified home cage environment, we will attempt to show distinct phenotypes dependent on chromosomal strain and we postulate that multiple chromosomes and QTLs will be involved in distinct behavioural domains within the food exploration strategies observed.

References:

1. Crawley, J.N.; Belknap, J.K.; Collins, A.; Crabbe, J.C.; Frankel, W.; Henderson, N.; Hitzemann, R.J.; Maxson, S.C.; Miner, L.L.; Silva, A.J.; Wehner, J.M.; Wynshaw-Boris, A.; Paylor, R. (1997). Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology (Berl)*, **132**, 107-124.
2. Henderson, N.D.; Turri, M.G.; DeFries, J.C.; Flint, J. (2004). QTL analysis of multiple behavioral measures of anxiety in mice. *Behav.Genet.*, **34**, 267-293.
3. Kas, M.J. and Van Ree, J.M. (2004). Dissecting complex behaviours in the post-genomic era. *Trends Neurosci.*, **27**, 366-369.
4. Nadeau, J.H.; Singer, J.B.; Matin, A.; Lander, E.S. (2000). Analysing complex genetic traits with chromosome substitution strains. *Nat.Genet.*, **24**, 221-225.
5. Noldus, L.P.; Spink, A.J.; Tegelenbosch, R.A. (2001). EthoVision: a versatile video tracking system for automation of behavioral experiments. *Behav.Res.Methods Instrum.Comput.*, **33**, 398-414.
6. Singer, J.B.; Hill, A.E.; Burrage, L.C.; Olszens, K.R.; Song, J.; Justice, M.; O'Brien, W.E.; Conti, D.V.; Witte, J.S.; Lander, E.S.; Nadeau, J.H. (2004). Genetic dissection of complex traits with chromosome substitution strains of mice. *Science*, **304**, 445-448.
7. Singer, J.B.; Hill, A.E.; Nadeau, J.H.; Lander, E.S. (2005). Mapping quantitative trait Loci for anxiety in chromosome substitution strains of mice. *Genetics*, **169**, 855-862.
8. Turri, M.G.; Datta, S.R.; DeFries, J.; Henderson, N.D.; Flint, J. (2001). QTL analysis identifies multiple behavioral dimensions in ethological tests of anxiety in laboratory mice. *Curr.Biol.*, **11**, 725-734.
9. Visser, L.; Bos, R.; Spruijt, B.M. (2005). Automated home cage observations as a tool to measure the effects of wheel running on cage floor locomotion. *Behav.Brain Res.*, **160**, 382-388.

A complex assessment of animal behavior in a learning paradigm

L.Yu. Rizhova, O.G. Semionova, V.G. Shaliapina
Laboratory of Neuroendocrinology, Pavlov Institute of Physiology of Russian Academy of Sciences, Sankt-Petersburg, Russia.

Abstract
Since numerous cognitive and non-cognitive functions affect performance in a learning task, it is always difficult to determine what process was changed by the treatment, and improved or impaired learning. Tests should be designed so that animals with individual peculiarities, such as orientation for positive or negative reinforcement, high or low locomotion activity, different brain asymmetry profiles, etc., have equivalent conditions for learning. We adapted the T-maze test accordingly. The animals learned to select a pre-determined alley of the maze. They could find food at the end of this alley, and an exit through which they could escape from the maze and enter a dark transport cage. We segregated animals with strong/weak lateral bias from the experimental groups and compared their learning abilities. The same procedure could be followed for animals showing interest or no interest in food, having high/low locomotion activity, and rearing etc. Thus we reveal the processes affected by the treatment, and which also influenced learning.

Keywords
Learning, T-maze, lateral bias, food motivation.

1 Introduction
Evaluation of the effects of various treatments on animal learning is a popular pursuit in modern neuroscience. Since numerous cognitive and non-cognitive functions affect performance in a learning task, it is always difficult to determine what process was changed by the treatment, and improved or impaired learning. We share the view that measuring multiple behavioral parameters in a single testing paradigm can often be a valid way to answer this question, and tests should be designed so that the effects of confounding factors such as lateral bias, anxiety, positive or negative reinforcement, and high or low test animal motor activity are minimized [1,3,4.5]. Description of the behavioral characteristics of the test animals might be made under basal conditions as well as during the cognitive challenges [6]. We adapted the T-maze test for the assessment of learning in rats accordingly. We show here how this method was used to compare the rate of learning in animals subjected to different treatments. We deliberately chose treatment alternatives that had implications for the methodic aspects of the research: i.e. the injection of a control solution into either the right or left lateral brain ventricle.

2 Methods

2.1 Animals
Male Wistar rats (200 – 230 g) were used. They were housed in groups of 6 – 7 per cage (40 x 60 x 20 cm) at a constant temperature (22 ± 1°C), and were given unrestricted access to food and water. One group of the animals was kept intact (N=24), the others were treated as follows: a permanent cannula was implanted in the animals above the right (RV, N=34) or left (LV, N=33) brain ventricle one week before the experiments.

2.2 Test solutions and the injection procedure.
5 μl of phosphate buffered saline (PBS) containing 0.5% bovine serum albumin (BSA) was injected in the brain ventricle one hour before testing. The coordinates of the injection cannula tip were 1 mm anterioposterior, 1.4 -1.6 mm lateral from bregma, and 8 - 8.5 mm ventral to the skull surface. Cannula placement was verified after the experiments when the brains were removed and inspected visually.

2.3 T-maze test
The T-maze was made of brown opaque plastic. It consisted of a stem (4.5 cm wide and 19 cm long) and a crosspiece (two 8 cm wide arms with a total length of 140 cm). The floor of the crosspiece was divided into 14 equal squares. To avoid handling animals in the maze, the crosspiece had escape alleys (15 x 8 cm) at both ends, facing the same direction as the stem. A dark transport cage containing sawdust was attached to the end of each escape alley. Food was removed from the housing cages 20 h before the experiment. Pellets of cheese were placed in both escape alleys, and their exits to the transport cages were closed. The animal was placed in the stem and a stopwatch was started. When the animal moved from the stem into the crosspiece, the passage between these two compartments was closed, and the animal's choice to turn left or right was recorded. When the animal started moving towards the end of the alley, the stopwatch was stopped, and the elapsed time was recorded (latency). The animal was given 5 min of adaptation. The following behavioral parameters were registered during this period: frequency of rearing, locomotion (number of squares crossed with all four paws), duration of grooming, reaction to the food, and number of fecal boluses. The exit from the alley was then opened, and the animal was given one min to enter the transport cage, which also contained a pellet of cheese. Any animal that did not leave the T-maze within one min was mildly forced to do so. Animals were then allowed to rest for 40 min in the transport cage, before again being placed in the stem of the T-maze, and taught to go to the opposite alley to the one first chosen. This time, the originally chosen alley did not contain any cheese, and its exit was always closed. If the animal did not start moving from the stem into one of the alleys within one min, the operator knocked on the stem with a pen in order to force it to do so. As soon as the animal went into the correct escape alley, the door between the alley and the crosspiece was closed and the animal was given one minute to eat the cheese and enter the transport cage. The animals were given about 20 min to rest between subsequent trials. Each animal had to go through this procedure several times, until it made four correct turns in succession. The number of trials necessary to achieve this was taken as a measure of learning potential.

3 Results and discussion

Comparison of the rate of learning in the animals with intracerebroventricular application of the control solution revealed that the side of application affected learning. The number of trials the animals were subjected to before they made four correct turns in succession was significantly lower in the rats injected in the left brain ventricle (7.24±0.82) than in the rats injected in the symmetrical point of the right brain ventricle (11.18±1.08), $p<0.01$. Intact control animals did not differ significantly from the other two groups (8.92±1.15). The superiority of the LV-animals in learning might be due to changes in cognitive functions, lateral preference, food motivation, anxiety, locomotion or exploratory behavior. We therefore investigated whether the observed effect of the side of application on learning could be ascribed to any of these parameters. As an example, we show how this investigation was performed for lateral preference and food motivation below.

Animal lateralization and learning in T-maze

To evaluate whether the unilateral intracerebroventricular injection influenced lateral preference in the animals, we determined the number of left and right turns made during the first trial in all groups. We then calculated the probability of a left turn for the animals in each experimental group and used this as a measure of population lateral preference. Intact rats did not exhibit asymmetry with regard to the first choice of side in the T-maze. The probability of left turn was close to 0.5 in these animals: 0.46±0.10. Similar results have been shown by other researchers [2]. The animals subjected to the injection in the lateral brain ventricle and unilateral damage in the area of the parietal cortex demonstrated population lateral bias. The probability of left turn was 0.68±0.09 in LV-rats and 0.15±0.06 in RV-rats (χ^2-test, $p<0.037$ and $p=0.000$ compared to 0.5 for LV- and RV-animals respectively). The results indicated that this treatment changed the profile of brain asymmetry in the animals, and RV-animals had a higher degree of population lateral bias. Since a high lateral bias in an animal may prevent it from learning to make an opposite turn, we investigated whether this was the reason the RV-rats were slower to learn than LV-rats. We segregated animals with different individual lateral preferences within each experimental group and compared their rate of learning. Animals that exhibited one or more sequences of alternating choices (RLRL or LRLR), never making three identical choices in succession during individual learning trials, constituted a group with low lateral preference – "ambilateral animals". Animals that exhibited one or more sequences of three identical lateral choices in succession (RRR or LLL), never making a sequence of alternating choices in the course of learning trials, constituted a group with high lateral preference – "lateralized animals". We identified 6, 11 and 7 ambilateral animals, and 15, 12 and 7 lateralized animals in the RV, LV and intact control groups respectively. 12, 10, and 10 rats were left in the residual group in the RV, LV and intact control groups, respectively. As shown in Fig. 1-A, the lateralized animals from the RV group were significantly slower to learn than the corresponding LV animals and intact controls. The ambilateral and residual animals did not differ significantly in the groups subjected to treatment. This fact indicates that the degree of lateralization in the animals influenced their cognitive performance in this test, and the lateralized animals alone created the difference in the rate of learning between RV and LV rats.

Food motivation and learning in T-maze

Since food motivation strongly influences animal learning, in the tests based on food reinforcement we investigated whether the side of application affected interest in food. For this purpose, we compared the number of animals which ate in the T-maze during the first 5 min. in all groups. No significant difference was found: 44%, 58% and 42% of animals consumed food during the first 5 min in the RV, LV, and intact control groups, respectively. Thus all the animals were similarly motivated with regard to food consumption before the beginning of the learning procedure. During the cognitive challenge, the rats were observed to change their food consumption. It was found that the number of animals which ate in the T-maze increased significantly with the number of trials for the LV and intact animals. The slope of the regression line was 0.039 and 0.035 respectively ($p<0.001$). No such effect was observed for the RV animals. The slope of the regression line was -0.003 in this group. Although the LV and intact animals had a population tendency to increase

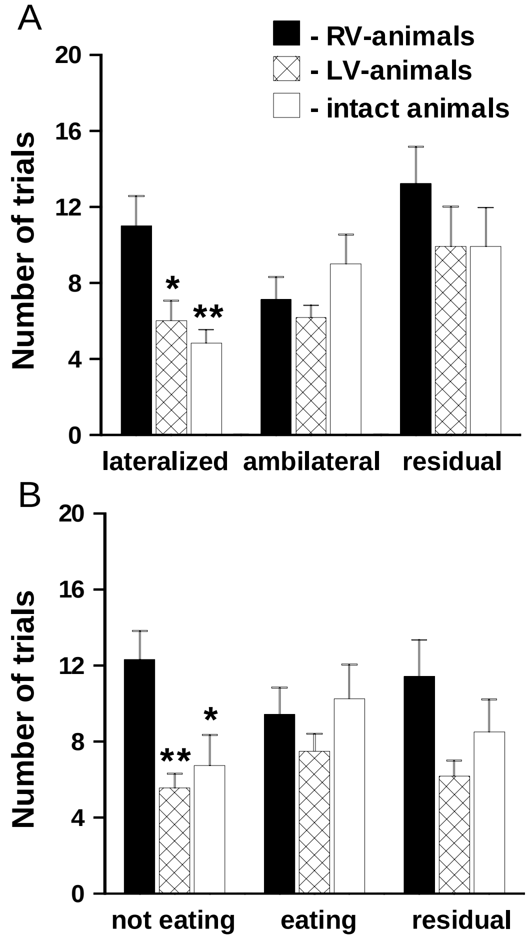


Figure 1. *Learning in T-maze. Each animal had to go through a series of trials, until it made four correct turns in succession. 5 µl vehicle solution were injected into the right (RV-animals) or left lateral brain ventricle (LV-animals) one hour before the test. Intact animals were not subjected to any operation or injection. (A) Shows learning in groups of rats with high (lateralized) and low (ambilateral) lateral preferences and in a residual group. (B) Shows learning in groups of rats eating and not eating and in a residual group. Data are given as means ± SEM. *p<0.05, ** p<0.01, when compared to RV animals, 2-tailed Student's test here and elsewhere.*

food consumption with the number of trials, some animals did not exhibit interest in food even at the end of the learning procedure. They may have use escape from the maze as reinforcement for learning. To elucidate whether the side of application affected the orientation to positive or negative reinforcement, which also influenced learning, we segregated the animals which ate and did not eat within each group and compared their rate of learning. The number of animals which ate in no less than 75% of trials was 28%, 37% and 33% in the RV, LV and intact controls respectively. The number of animals that consumed food in no more than 25% of trials was 51%, 44% and 50% in the RV, LV and intact controls respectively. 21%, 19% and 17% constituted the residual group in the RV, LV and intact controls respectively. As shown in Fig. 1-B, the effect of the side of application on learning was only observed in the animals which did not eat. The results therefore indicate that food motivation in the animals during the learning procedure depended on the side of the unilateral brain damage provoked by the injection in the lateral brain ventricles.

Overall, the results show that the unilateral injections of control solution in the lateral brain ventricles changed the profile of brain asymmetry in the animals and influenced behavioral parameters. The T-maze procedure which was used in our research to measure learning in the animals reveled that at least two non-cognitive functions, lateral preference and food motivation, were affected by the treatment in the animals, and these influenced their rate of learning. In this case, we are therefore unable to say that the treatment the animals were subjected to influenced cognitive functions, such as working memory. Many contemporary studies fail to take into account the impact of a particular treatment on non-cognitive functions and therefore run the risk of drawing erroneous conclusions about the cognitive effects of such treatments.

References:

1. Dockstader, C.L., van der Kooy, D. (2001). Mouse strain differences in opiate reward learning are explained by differences in anxiety, not reward or learning. *J Neurosci*, **21**, 9077-9081.
2. Douglas, R.J., Isaacson, R.L. (1965). Homogeneity of single-trial response tendencies and spontaneous alternation in the T-maze. *Psychol Rep*, **16**, 87-92.
3. Escorihuela, R.M., Fernandez-Teruel, A., Gil, L., Aguilar, R., Tobena, A., Driscoll, P. (1999). Inbred Roman high- and low-avoidance rats: differences in anxiety, novelty-seeking, and shuttlebox behaviors. *Physiol Behav*, **67**, 19-26.
4. Ohl, F., Roedel, A., Storch, C., Holsboer, F., Landgraf, R (2002). Cognitive performance in rats differing in their inborn anxiety. *Behav Neurosci*, **116**, 464-471.
5. Saghal, A. (1993). Practical behavioural neurosciense: problems, pitfalls, and suggestions. In: *Behavioural Neuroscience. A practical approach*, **vol 1** (Saghal A, ed), pp 1-8. Oxford: Oxford University Press.
6. Wellig, F. (1991). Roman strains as a psychogenetic model for the study of working memory: behavioural and biochemical data. Pharm*. Biochem. Behav*, **40**, 7-16.

A new approach to study preference: Applying Hernstein's matching law in a two bottles paradigm

C. Sanchis-Segura[1]; R. Spanagel[1]; F. A. Henn[2]; B. Vollmayr[2]
[1]*Department of Psychopharmacology, CIMH, Mannheim, Germany*
[2]*Biochemical Laboratory, CIMH, Mannheim, Germany*

Abstract

In preclinical research, as they cannot be directly assessed, hedonic properties of a stimulus are usually inferred by analyzing its preference over another "control" stimulus (i.e water). Here we present a method to obtain a more sophisticated index of preference. This procedure is very similar to other two-bottles based sucrose preference tests. Thus, this method also consist in an evaluation of the preference of different sucrose concentrations over water but it also includes subtle comparisons among different sucrose concentrations simultaneously available for a restricted time (in our case, 1 hour). These discrete preference comparisons are then used to estimate a general pattern in the preference allocation and its correspondence towards the relative magnitude of the reinforcement sources available. This is accomplished by applying the generalized matching law equation (Hernstein, 1970) and basically consists on the estimation of the regression line relating changes in preference according to changes in the ratio of two reward sizes simultaneously available. This test has some advantages over traditional methods: 1) It does not require deprivation. 2) It is more sensitive. 3) It is obtained across a wide range of conditions. 4) It provides a function describing the dynamic relationship between reinforcer magnitude changes and individual's behavioral allocation of preference.

Keywords

sucrose, matching law, preference, anhedonia

1 Introduction

In the present study we present a method aimed to provide a measure of the allocation of preference of orally consumed reinforcers. This procedure is similar to the classic two bottles free choice test of preference between reinforcer/ neutral stimuli (i.e. water) but it also incorporates more subtle comparisons between different magnitudes of the reinforcer.

The results obtained in this procedure are analysed by using the Generalised Matching Law equation [3]. This equation provides two indexes, corresponding to the origin and slope of the regression line defining the relationship between preference allocation and the relative magnitude of two reinforcement sources simultaneously available at the se same cost. Therefore, the origin is a measure of the "bias" for any or other alternative, whereas the slope accounts the sensitivity to the changes in the ratio of two reward sizes. We propose that the later could be a better indicator of the hedonic properties of orally derived reinforcers than the traditional measures based on their preference over water.

We had already used a similar procedure to unravel reward-based difference in mice [5] and here we apply it to rats. More specifically we compare two strain of rats, selectively bred for being prone/ resistant to helplessness (cLH and cNLH, respectively). Accordingly to the current views of this animal model of depression, sucrose rewarding properties should be reduced in cLH rats. In this regard, here we show that cLH display a mismatch between their preference allocation and the relative magnitude of two reinforcement sources. Since this preference mismatch is observed between reward sources at the same cost and simultaneously available, we discuss this finding as a possible indication of a lack of sensitivity to reward magnitude and "anhedonia". In addition, we also show that the individual score of this mismatch inversely correlates with the individual value of an *a priori* defined index of helplessness, then supporting the proposed relationship between helplessness and anhedonia..

2 Material and Methods

Eight male cLH and 7 male cNLH rats from the 52nd generation were used in the present study. The origin and selective breeding of both strains has been previously described [2, 6]. Rats were singly housed in standard laboratory conditions (21°C ± 0.5 and 50% relative humidity; 12 hr light/dark cycle; water and food *ad libitum*). The experiments were approved by the Committee on Animal Care and Use of the relevant local governmental body and performed following the German Law on the Protection of Animals.

Rats were tested for helplessness [6]. The test consisted of 15 trials in which an electric foot shock (0.8 mA, 60 s) could be stopped by pressing a bar. Trials not stopped after 20s were considered as a failure. This test was conducted when animals were 9 weeks old and to prevent any interference of the helplessness, 23 weeks after the test for helplessness, sucrose consumption was assessed using a two bottles procedure. Special caps were used in the bottles to prevent spillage [5]. Thus, completing 5 sessions (duration: 1h) per week and in a total time of 8 weeks, 14 different conditions testing sucrose preference were implemented in the rats home cage. A condition was defined as one specific comparison between 2 sucrose solutions of four possible concentrations (0, 1, 2, or 7% w/v) with a specific location of the bottles (i.e. a different placement of the same pair of solutions was considered as a different condition).

Data were analyzed by using the "generalized matching law equation". This equation states that log B1/B2=a(log r1/r2)+log c, where B represents the allocation of the behavioral responses to alternatives 1 and 2, and r represent the rate or relative reinforcing magnitude of the two alternatives, and a and c are empirically obtained parameters which illustrate the individual sensitivity to the ratio reward and bias for one or the other alternatives respectively [1]. Applying this equation to a two bottles free choice procedure is a less usual application of this equation. In this regard it should be noted that the volume consumed is part of the final reinforcer magnitude. However, although this procedure can be slightly different of the usual application in an operant setting, it has been successfully used to study preference patterns for ethanol and sucrose [4, 5]. More specifically, in this procedure, the consumed volume (V_X) of each bottle provides the index of the relative behavioral allocation, whereas the concentration of the available solution (C_X) represents the magnitude of the reinforcer. Thus, for each rat the ratio of the consumed volume from the bottle located in the left over that located in the right (V_L / V_R) as well as the concentration of the respective solutions (C_L / C_R) were calculated. When any of the terms was zero, 0.1 substituted this value. This value represents the accuracy limit of our measurement method, and by this substitution division by 0 (and posterior undefined logarithm) is avoided with minimal distortion of the subsequent calculations. In a second step, the logarithms (base e) of these ratios were plotted on arithmetic co-ordinates and, by using the method of the least squares, the best-fit of regression line was estimated. For this purpose the coefficient of determination is provided.

3 Results

As expected cLH showed a poor performance in the escape task as expressed in a significantly higher number of escape failures [t (13)=8.34, p<0.0001] than cNLH (mean $\pm$ SEM were 9.5 $\pm$ 0.68 and 1.85 $\pm$ 0.59 for cLH and cNLH respectively) despite the fact that cLH had no inescapable shock exposure.

Interestingly, the slope value, but not the bias (intercept value), index was inversely correlated with the number of failures to escape in the test for learned helplessness (r= –0.64; p<0.01, and r=-0.38; p>0.15 respectively), illustrating the inverse relationship between helplessness and sensitivity to sucrose reward**.** A similar association between helplessness and preference allocation was also observed when genotype differences were considered. Thus, the calculated individual intercepts and slopes were averaged by genotype and compared by two Student's t test for independent samples. These tests show that whereas the average of the intercepts were not different [t(13)=-1.09; p>0.29], the slopes markedly differ [t (13)=-2.92 p<0.02] between genotypes (slope mean $\pm$ SEM were 0.57 $\pm$ 0.04 and 0.71 $\pm$ 0.02 for cLH and cNLH, respectively).

From the same data, the average of ml consumed of a sucrose solution (1, 2 or 7 % w/v) and its preference over water were calculated. A two-way repeated measures ANOVA (genotype x concentration) revealed that the sucrose concentration factor [F (2,26)=164.11, p<0.0001], but not the genotype [F(1,13)= 1.50, p>0.24] or their interaction [F (2,26)= 0.51, p>0.60] yielded a significant effect over sucrose consumption**.** Thus, whereas cLH rats consumed 5.25 $\pm$ 0.72, 5.1 $\pm$ 0.84 and 16.6 $\pm$ 1.96 ml of a 1, 2 and 7 % (w/v) sucrose solution, cNLH rats consumed a little more, namely 6.31 + 0.5, 7.44 + 0.53 and 17.72 + 0.46 ml for the same sucrose concentrations. Similar results were found when preference instead of consumption was considered. Again the sucrose concentration factor [F (2,26)=61.12, p<0.0001], but not the genotype [F(1,13)= 2.80, p>0.12] or the interaction [F(2,26)= 1.23, p>0.30], resulted in a significant effect. In this case the average preference for a sucrose solution (1, 2 or 7 % w/v) over water was 79.28 $\pm$ 2.36, 68.70 $\pm$ 2.84 and 91.07 $\pm$ 0.83 for cLH and 79.89 $\pm$ 1.11, 75.12 $\pm$ 2.78 and 93.55 $\pm$ 0.67 for cNLH. Furthermore, and again in contrast to the slope values of the matching law, the number of avoidance failures was not correlated with the sucrose consumption or preference over water at any tested concentration (1, 2 or 7 % w/v).

4. Discussion

We observed reduced matching between preference and the relative magnitude of two reinforcers simultaneously available in cLH rats, as reflected in their lower slope values of the generalized matching law equation [3] as compared to those of the cNLH rats. Although *sensu strictu* these slope values reflect the sensitivity to the changes in the ratio of two reward sizes, this index could be a better index of the rewarding properties of a stimuli and, consequently, of anhedonia than traditional measures of preference [5]. Indeed it has been previously labeled as "sensitivity to reward". [4], and conversely to raw measures of sucrose consumption or preference over water, it identified a genotype difference and showed a significant correlation to the previously established helplessness criterion (individual number of failures in an escape paradigm). then confirming highly sensitive measures are needed to detect subtle differences in sensitivity to sucrose reward (i.e. hedonia) in non-deprived animals. However, as for any measure based on preference, the procedure referred here is not independent of the ability to discriminate the available reward sources. Future efforts are guaranteed to circumvent this problem and try to isolate the perceived reward value in this paradigm.

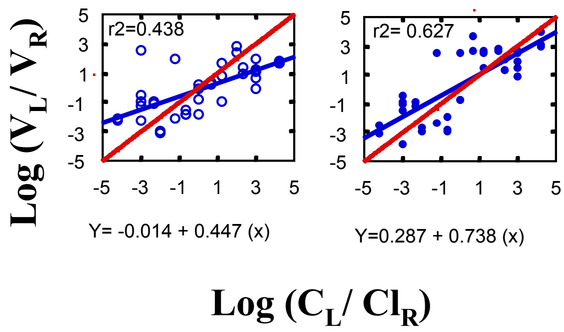


Figure 1. *Representative patterns of sucrose choice in rats that exhibit congenital helplessness (cLH; open circles) or resistance to learned helplessness (cNLH; filled circles) rats. As described in the methods section, for each rat the ratio of the consumed volume (V_L/V_R) as well as the concentration of the respective solutions (C_L/C_R) were calculated. Then, the logarithms (base e) of these ratios were plotted on arithmetic co-ordinates and, by using the method of the least squares, the best-fit of regression line was estimated. Individual parameters of the empirical regression lines (plain line) and wellness of fitting (r^2) are included. For a better comparison the theoretical perfect matching behavior is shown (dashed line).*

References

1. Anderson, K.G. & Woolverton, W.L. (2000). Concurrent variable-interval drug self-administration and the generalized matching law: a drug-class comparison. *Behavioural Pharmacology*, **114,** 13-20.
2. Henn, F. & Vollmayr, B. (2005). Stress models of depression: Forming genetically vulnerable strains. *Neuroscience and Biobehavioral reviews* (Accepted).
3. Hernstein, R.J. (1970). On the law of the effect. *Journal of Experimental Analysis of Behavior*, **13,** 243-266.
4. Martinetti, M.P., Andrzejewski, M.E., Hineline, P.N., Lewis, M.J. (2000). Ethanol consumption and the matching law: a choice analysis using a limited-access paradigm. *Experimental and Clinical Psychopharmacology*, **8,** 395-403.
5. Sanchis-Segura, C., Cline B.H., Marsicano, G. Lutz, B., Spanagel, R. (2004). Reduced sensitivity to reward in CB-1 knockout mice. *Psychopharmacology* **176,** 223-232.
6. Vollmayr, B. & Henn, F. (2001). Learned helplessness in the rat: Improvements in validity and reliability. *Brain Research Protocols*, 8, 1-7.

The application of Active Allothetic Place Avoidance (AAPA) task in the study of cognitive deficit and hyperlocomotion following MK-801 administration in rats

A. Stuchlik[1], V. Bubenikova[2], K. Vales[1]
[1] *Institute of Physiology, Academy of Sciences, Prague, Czech Republic*
[2] *Prague Psychiatric Center, Prague, Czech Republic*

Abstract

Blockade of brain NMDA receptors with MK-801 results in specific behavioral alterations, which can be ameliorated by antipsychotics. Nonetheless, the effect of antipsychotics on the cognitive deficit following MK-801 application is not elucidated yet. The present study tested whether clozapine alleviates the cognitive deficit following administration of two doses of MK-801 (0,1 a 0,3 mg/kg, s.c.).in the AAPA task. In this task, rats are required to avoid a room-frame fixed shock sector on the rotating arena, and this task requires the rats to separate spatial stimuli into coherent representations and use only the room-frame system for the navigation. Application of MK-801 increased locomotion and decreased spatial efficiency. Clozapine decreased total distance walked in the session but it was ineffective in alleviating decreased spatial efficiency following MK-801. It is summed up that in the AAPA task, clozapine is capable of decreasing hyperlocomotion but incapable of alleviating deficits in spatial cognition.

Keywords

MK-801, dizocilpine, animal model of schizophrenia, spatial cognition

1 Introduction

Application of non-competitive NMDA receptor antagonists is used as an animal model of schizophrenia with a relatively high face and predictive validities. Blocking NMDA receptors with MK-801 leads to specific behavioral changes, including hyperlocomotion, stereotypies, social deficit and impaired sensorimotor gating (measurable as deficit in the prepulse inhibition of acoustic startle reflex). These changes can be alleviated by application of classical and atypical antipsychotics. Cognitive deficit following systemic administration of MK-801 in the AAPA task was described recently [1]. However, the effect of antipsychotics on the cognitive deficit following NMDA receptor blockade is not understood yet. Aim of the present study was determine the effect of clozapine on the cognitive deficit following administration of two doses of MK-801 (0.1 and 0.3 mg/kg, s.c.).in the Active Allothetic Place Avoidance (AAPA) task.

2 Methods

2.1 Animals

Forty-eight naive male adult rats of the Wistar strain (5-month-old, weighing 350-450g) obtained from the Institute's breeding colony were used in the experiment. Animals were accommodated two per cage in 25 x 30 x 45-cm transparent plastic cages in a laboratory air-conditioned animal room with constant temperature (22^{o}C) and 12:12 light/dark cycle (with the lights on at 7:00 a.m.). Under light diethylether anesthesia, all animals were implanted with a low-impedance connector made from a hypodermic needle, which pierced the rat's skin between the shoulders. The sharp end of the needle was cut off and bent with tweezers to form a small loop, which prevented the connector from slipping out and provided anchor for an alligator clip, which was connected to a shock-delivering wire. Water and food were available *ad libitum* throughout the experiments. All manipulations were done in accord with the Law on Animal Protection of Czech Republic and with the appropriate directive of the European Communities Council (86/609/EEC).

2.2 Drugs

MK-801 (Dizocilpine maleate; [5R,10S]-[+]–5-methyl-10,11-dihydro-5H-dibenzo[a,d] cyklohepten-5,10-imine; Sigma Aldrich) was dissolved in saline (0.1 mg/ml and 0.3 mg/ml) and injected 30 min prior to behavioral training (0.1 and 0.2 mg/kg). Clozapine (Sigma; Aldrich) was dissolved in saline (5 mg/ml) acidified with 20 microl of concentrated acetic acid and injected 60 min prior to behavioral testing (5 mg/kg).

2.3 Apparatuses and behavioral procedures

The AAPA apparatus was described in detail elsewhere [2]. Briefly, it consisted of a smooth metallic circular arena (80 cm in diameter) enclosed with a 30 cm high transparent Plexiglas wall and elevated 1 m above the floor of the dimly lighted 4x5m room with the abundance of extramaze landmarks. At the beginning of each training session, a rat was placed on the rotating arena (1 rpm), where a directly imperceptible 60-degrees to-be-avoided sector (shock sector) was defined by the computer-based tracking system (iTrack, Biosignal Group Corp., USA), located in an adjacent room. The location of the shock sector could be determined exclusively by its spatial relations to distal orienting cues located in the room. The rat wore an infrared light-emitting diode (LED) fixed between its shoulders with a light latex harness, and its position was tracked every 40 ms and recorded onto a computer track file, allowing subsequent reconstruction of the track with an off-line analysis program (TrackAnalysis, Biosignal Group Corp., USA). Whenever a rat entered the shock sector for more than 0.5 s, mild electric shocks (50 Hz, 0.5 s) were delivered at intervals of 1.5 s until the rat left the shock sector for at least 0.5 s (Fig. 3). The shocks were delivered through a thin subcutaneous low-impedance nichrome wire implant on the back of the rat standing on the grounded floor. The appropriate shock current (ranging between 0.2-0.7 mA) was individualized for each rat to elicit a rapid escape reaction but to prevent freezing. Since the arena was rotating, the rat had to move actively away from the shock in the direction opposite to the arena rotation, otherwise it was passively transported to the shock sector.

2.4 Design of experiments and data analysis

Rats (N=48) were divided into six experimental groups. The animals from the group SAL (n=8) were injected with 1ml/kg of saline (60 min and 30 min prior to behavioral testing), rats from the group SAL-CLOZ (n=8) were injected with clozapine (5 mg/kg) 60 min prior to testing and saline (1ml/kg) 30 min prior to testing, animals from the groups MK01 (n=8) and MK03 (n=8) were injected with 1ml/kg of saline 60 min prior to testing and with 0.1 mg/kg and 0.3 mg/kg of MK-801 30 min prior to testing, respectively. Subjects from groups MK01-CLOZ (n=8) and MK03-CLOZ (n=8) were injected with 5mg/kg of clozapine 60 min prior to the testing and with 0.2 mg/kg and 0.3 mg/kg of MK-801 30 min prior to the testing, respectively. All animals received injections of the same volume of liquid per kg weight.

Animals were trained for four consecutive daily sessions in the AAPA task, with the shock sector located in the north of the room. Experimental sessions in AAPA lasted 20 min and each rat had one session every day, carried out during daylight hours.

The following parameters were recorded and analyzed in order to assess rats‘ behavior in the AAPA: the total distance traveled in a session (DISTANCE) measured in the arena frame (which only takes into account active locomotion) reflected the locomotor activity of rats and the number of entrances to the shock sector (ERRORS) reflected the efficiency of spatial performance in the AAPA task. Latency to the first entrance into the shock sector TIME1ST reflected between-session learning. Maximum time a rat spent in the safe part of the arena between two ERRORS in a particular session was also recorded (MAX T). It reflected the ability to remember the shock sector location and to avoid it. Data from the four days of training in the AAPA task were analyzed using a two-way ANOVA (Treatments x Sessions) with repeated measures on Sessions. Tukey’s HSD test was used when appropriate. The significance was accepted on probability level of 5%.

3 Results

No vocalizations, increased defecations or ataxia were observed during and after the injections of MK-801 and clozapine. On the contrary, visual inspection revealed hyperactivity after injection of higher dose of MK-801, which was later confirmed by measuring the total DISTANCES. Rats were able to maintain correct postural positions and their main neurological reflexes were preserved.

As concerns DISTANCE, it was increased significantly in the groups MK01 and MK03 with respect to the group SAL, whereas clozapine blocked the MK-801 induced hypermotility. Analysis of spatial performance using the parameter ENTRANCES revealed that lower dose of MK-801 was without effect on ENTRANCES, whereas higher dose significantly increased the number of ENTRANCES. Clozapine failed to block the MK-801-induced increase of ENTRANCES.

As concerns the MAXT parameter, it was significantly increased in the groups MK01, MK03, CLOZ, CLOZ-MK01 and CLOZ-MK03 with respect to the group SAL, i.e., both clozapine and MK-801 impaired spatial navigation efficiency analyzed with this parameter.

4 Discussion

The results demonstrate that application of MK-801 increases locomotor activity and decreases spatial efficiency, measured as the increased number of shock sector entrances and decreased maximum time avoided. Administration of clozapine decreased total path traveled in the session but was ineffective in alleviating decreased spatial efficiency following MK-801.It is concluded that in the AAPA task, clozapine is capable of decreasing hyperlocomotion but incapable of alleviating deficits in spatial cognition.

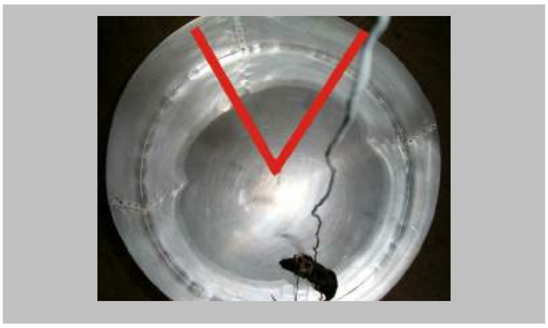


Figure 1. *Overhead photograph of the experimental AAPA arena. The rat walked on the featureless rotating arena and avoided an unmarked sector (denoted by a dashed line) fixed in the coordinate frame of the room. It must be pointed out that the sector is unmarked (defined only in the computer-based tracking system). The rat is connected to a shock-delivering wire, which delivers mild shocks whenever the rat enters the shock sector. Thick arrow denotes the sense of arena rotation (1 rpm).*

Supported by grants MZ ČR-NL-7684-3, GACR No . 309/03/P126 and GACR No 309/03/0715 and by MSMT CR project 1M0002375201.

References:

1. Stuchlik A., Vales, K. (2005) Systemic administration of MK-801, a non-competitive NMDA-receptor antagonist, elicits a behavioural deficit of rats in the Active Allothetic Place Avoidance (AAPA) task irrespectively of their intact spatial pretraining *Behav. Brain Research* **159**(1):163-171.
2. Cimadevilla JM, Wesierska M, Fenton AA, Bures J. (2001) Inactivating one hippocampus impairs avoidance of a stable room-defined place during dissociation of arena cues from room cues by rotation of the arena. *Proc Natl Acad Sci U S A*; **98**:3531-36.

Human threat test: a method to test anxiety related behavior in a marmoset monkey

S.A.M. van Vliet[1,2], M.J. Jongsma[1], R.A.P. Vanwersch[1] and I.H.C.H.M. Philippens[1,2]
[1]*Dept. of Diagnosis and Therapy, TNO Defence, Security and Safety, Rijswijk, The Netherlands*
[2]*Dept. of psychopharmacology, Utrecht University, Utrecht, The Netherlands*

Abstract

The human threat test is a non-human primate putative model of anxiety. It is based on findings that marmosets will exhibit anxiety-related behavior in the presence of a human observer in front of the cage. The test is applied to our marmoset colony to asses the effects of three psychoactive compounds: diazepam, modafinil and Δ^9-THC on anxiety-related behavior.

Keywords

Human threat test, Marmoset monkey, anxiety, modafinil, Δ^9-THC.

1 Introduction

Defensive responses are components of the daily life of primates. These responses are still present in captivity born animals. Based on this behavior, the human threat test (HTT) is a widely used method in the assessment of anxiety/fear related behavior in a non-human primate [3]. The test is based on findings that non-human primates will exhibit anxiety-related behavior in the presence of a human observer. Retreating to the back of the cage and display of aggressive and anxiety-related postures are the main pronounced reactions.

This study focused on the effects of three psychoactive compounds on the anxiety-related behavior of the marmoset monkey measured with the HTT. First, the sedative compound diazepam, known to decrease anxiety, was applied to validate the HTT on our marmoset colony. The second compound, modafinil (modiodal, D,1-2-[(diphenylmethyl) sulfinyl]acemide) is a vigilance stimulant used as treatment of narcolepsy. The third compound was Δ^9-tetrahydrocannabinol (Δ^9-THC), the main euphorigenic component of marijuana *(Cannabis sativa).*

2 Materials and methods

2.1 Animals

Adult male and female marmosets *(Callithrix jacchus),* aged 2-6 years with body weights between 350-550 g were obtained from BPRC, The Netherlands and Harlan, United Kingdom. The animals were kept one to a cage and housed in one animal room. The ambient temperature was regulated at 25±2 ºC and the relative humidity was always >60%. A 12-h light-dark cycle was maintained. All aspects of animal care are described in Standard Operating Procedures, which are in agreement with current guidelines of the European Community. The independent TNO committee on Animal Care and Use approved all protocols for the animal experiments.

2.2 Human threat test

The behavior was assessed in the home cage (40 x 60 x 60 cm high x wide x deep) with a hanging basket in the back of the cage, a wooden board (20 x 10 cm, 30 cm above cage floor) on the left side in the back and on the other side a perch, at the same height, positioned from the back to the front of the cage.

To assess the behavior, the observer stood approximately 30-100 cm from the cage front and made eye contact with the marmoset throughout a 2-min. test period. The observer was the only human in the room to prevent distraction. During this period the movements, behavior and position of the marmoset in the cage were recorded by video registration. Fifteen minutes before start of the test, the recording apparatus was placed in front of the cage to let the animals habituate to the new situation.

A range of parameters were obtained (according to [2] based on [5]):

- The number of characteristic postures exhibited:
  Tail posture: tail raise to present the genital region,
  Scent marking: the anal and genital area is pressed against the substrate to be marked with excretion of the present glands,
  Arched pilo: arched back posture with full body piloerection,
  Slit stare: stare with the eyes half closed in combination with tufts flattened and exposure of the teeth,
  Rearing: upright position with flexed paws,
  Twisting: head and torso movement from side to side.
- The time spent in the front of the cage,
- The number of position changes in the cage,
- The number of movements from the back of the cage to the front,
- The number of jumps from the left side of the cage to the right side or vice versa.

In addition to these parameters, any difference from normal behavior was noted.

2.3 Study design

The HTT was carried out on the time point when the drug was most effective. The effects of diazepam were tested 45 minutes after administration, Δ^9-THC was tested after 90 minutes and modafinil 120 minutes after administration. Baseline values were obtained in the same animals without treatment. No differences were found between non-treated and vehicle-treated animals (data not shown).

2.4 Drug administration

Diazepam (Roche, Switzerland) was administrated intramuscular in a dose of 0.25 mg/kg. Modafinil (Laboratoire L. Lafon, France) was given in doses of 50, 100, 150 and 225 mg/kg p.o. Before usage the grinded tablets were homogenized freshly in a 10% sugar solution in a dose volume of 1.5 ml/kg. Δ^9-THC (dissolved in ethanol (28 mg/ml), IBL, Leiden, The Netherlands). The Δ^9-THC solution was administrated simultaneously with freshly prepared syrup (Karvan Cevitam with water (1:1)) p.o. To control for the effects of the solvent, ethanol was administrated comparable to a dose of 4 mg/kg.

2.5 Statistics

The results are presented as mean ± SEM. Nonparametric statistical analysis was applied to all data. For each dose, the differences from baseline values were calculated with the Wilcoxon signed rank test (wsrt). The doses effects were analyzed with the Kruskall-Wallis test and Mann Whitney test. Since the groups did not differ on their baseline performance, the values after administration of the compound were used. Significance levels of $p<0.05$ were used.

3 Results

Diazepam (0.25 mg/kg, n=8) was used to validate the test. Administration of diazepam altered the response of the marmoset to a human threat. Diazepam increased the time spent in front and the number of characteristic body postures (wsrt, $p<0.05$). Diazepam had no influence on the parameters concerning the movement through the cage.

Modafinil at 50 (n=6), 100 (n=13), 150 (n=8) and 225 (n=6) mg/kg affected the HTT. The characteristic body postures were reduced after 50, 150 and 225 mg/kg (wsrt, $p<0.05$) and the time spent in front was increased after 225 mg/kg (wsrt, $p<0.05$). After the doses of 100, 150 mg/kg and 225 mg/kg modafinil, the number of movements were increased (wsrt, $p<0.05$), except the parameter 'movements forward' after 225 mg/kg.

Δ^9-THC at 2 (n=6), 4 (n=6) and 8 (n=7) mg/kg reduced the number of body postures, but had no effect on the other parameters of the HTT. The solvent ethanol did not interfere in these effects while no effects were found on the HTT.

Table 1. *Overview of the effects of the different dosages of diazepam, modafinil and Δ^9-THC on the parameters of the HTT. The numbers indicate the mean ± SEM of the parameter after administration as a percentage of the baseline values. *p<0.05 versus baseline values.*

	mg/kg	**Body postures**	**Time in front**
Diazepam	0.25	16 ± 6.3 %*	173 ± 41%*
Modafinil	50	17.2 ± 9.9 %*	105 ± 28 %
	100	104 ± 37 %	455 ± 245 %
	150	38.6 ± 15.7 % *	152 ± 41 %
	225	31.8 ± 8.9 % *	307 ± 128 % *
Δ9-THC	2	46.3 ± 11.4 % *	125 ± 48 %
	4	10.5 ± 4.7 % *	211 ± 85 %
	8	20 ± 12.2 % *	74 ± 18 %

4 Discussion

The results show that the three compounds have anxiolytic-like effects. Diazepam is a known anxiolytic compound and therefore used for validation. The effects of diazepam on the HTT are in concordance with previous studies. The anxiolytic-like effects of modafinil have not been reported earlier. The found effects of Δ^9-THC are similar to earlier reports, where Δ^9-THC seems to have a biphasic effect on anxiety [4].

The HTT is newly applied to our marmoset colony. Application of methods in a new setting needs always adaptations and will affect the outcome. This is especially important in the case of measuring behavior concerning anxiety. The anxiety level of an animal depends on factors like the habituation to the environment and to the observer, the intrinsic anxiety level and the handling of the animals by the experimenters. The overall anxiety level of our marmoset colony is low compared to baseline measurements described in other articles [2,3]. Our colony is highly habituated to their environment and human presence. Another reason can be the animal friendly method of handling. The animals are trained to jump into a handling box, whereas in other institutes the animals are caught with thick gloves inducing the anxiety.

The overall low anxiety level is reflected in the outcome of the test. The increased 'time spent in front of the cage' was only to be found significant after 0.25 mg/kg diazepam and after the highest dose of modafinil. This is striking, because this parameter is very sensitive to anxiolytic effects and is the most frequently affected behavior in other studies [1]. Since modafinil and Δ^9-THC reduced the body postures, the compounds seem to be anxiolytic and were expected to increase the time spent in front. Probably, the parameter 'time spent in front' is most susceptible to habituation. In normal situation our animals already spent 32 ± 3% of their time in front of the cage.

Also, no selection of animals on the level of anxious behavior was carried out beforehand in order to reduce the inter-individual variability. Prior screening of the animals is fairly common in applications of the HTT [3]. Our expectation is that putative anxiolytic effects of modafinil and Δ^9-THC would be more distinct after pre-selection and less habituated animals.

Is this test system also able to measure anxiogenic effects? The effects of anxiogenic compounds with the HTT were tested by another research group [2], but they found no changes in the response to the human threat. Probably inclusion of more body postures, as carried out in our study, will add more sensitivity. In our approach body postures as rearing and twisting were included. Together with slit stare, these body postures can be categorized as postures expressing fear-like behavior. Whereas the other used body postures like tail posture, arched pilo and scent marking can be categorized as impressing behavior. The increase of fear postures and decrease of impress postures would be an indication of anxiogenic behavior.

In conclusion, this study indicates that the HTT is a sensitive test to measure anxiety-related behavior and can be adapted to other marmoset colonies.

References:

1. Barros, M.; Tomaz, C. (2002) Non-human primate models for investigating fear and anxiety. *Neurosci Biobehav Rev*, **26**, 187-201
2. Carey, G.J.; Costall, B. (1992) Behavioural effects of anxiogenic agents in the common marmoset. *Pharmacol Biochem Behav,* **42,** 143-153
3. Costall, B.; Domeney A.M. (1988) Zacopride: anxiolytic profile in rodent and primate models of anxiety. *J Pharm Pharmacol* , **40**, 302-305
4. Chaperon, F.; Thiebot M-H. (1999) Behavioral effects of cannabinoid agents in animals. *Crit Rev Neurobiol,* **13**, 243-281
5. Stevenson, M.F.; Poole, T.B. (1976) An ethogram of the common marmoset *(Calithrix jacchus jacchus)*: general behavioural repertoire. *Anim Behav,* **24**, 428-451

The effect of practice modalities on bi-lateral transfer of a motor skill

Benjuya, N and Wegman, E.
Department of Physical Education, Kaye College of Education, Beer-Sheba, Israel.

Abstract

The ability to transfer a motor skill bilaterally was examined both in adults and children. Three groups of randomly selected children were trained to learn a gross motor skill which required total body involvement, each according to distinct 3 training modalities; constant, mixed and variable. Over 4 weeks the groups were allowed to perform 120 trials. Both temporal and surface electromyography (EMG)s from the abdominal and thigh muscles were recorded performing on dominant side at the beginning (pre), end of practice period (post), and on non-dominant side (transfer). Although each group improved both in the total time to execute the skill and inter group EMG variability over the training sessions on the dominant side, the variable group regardless of the type of skill and age, showed superior transfer to the non-dominant side. The results suggest that asymmetrical transfer in gross motor skill performance which requires both lateralities is best via variable practice.

Keywords

General motor program, motor learning, practice, skill acquisition, bilateral transfer

1 Introduction

In many sport activities, bilateral skill development is an important aspect of the training process. A novice, when given a choice, will emphasize the limb/side with which he or she feels the most comfortable. Despite the existing controversies in skill training and rehabilitation techniques, the majority of transfer studies have found that most transfer occurs from the preferred to the nonpreferred limb (asymmetric transfer) automatically.

Both amounts of practice and variability in practice experiences (involving variations of the skill being learned) are essential conditions in developing a coordinated sequence of actions (motor programs)[1-3]. A program is called a generalized motor program (GMP) when called upon successfully in a variety of performance situations, regardless if it is a closed or open skill and can be executed by different muscles (e.g. either limb). This feature of GMP-effector independence- can be enhanced by practice variability [1, 4]. Recently, this notion was questioned when studies showed that movement pattern and force characteristics can become integrated when the trainee adopts the unique characteristics of the effector over the practice sessions [5]. In another study which tested transfer of motions with different emphasis, timing or the scaling of force, an asymmetry in transfer was found for force control, with significant transfer only in the preferred-to-nonpreferred direction. Transfer of timing occurred similarly in both directions [6]. These results indicate timing as a powerful component for bilateral transfer, while force control showed to be more dependent on practice with the specific muscular system. Thus, the demands of the task might be a crucial factor in determining the effects of practice on the cost associated with changing the muscle groups.

Schmidt's Schema theory is relatively specific about practice conditions that enhance GMP learning. The execution of the GMP is via a recall schema for movements that share mainly similar kinematic profiles that upon its completion is checked by the recognition schema for corrections and updating the future recall schema. Schmidt and Lee [3] emphasized that in order to optimize the recall schema, a set of relevant movement parameters needs to be employed in a variable practice that will prepare the trainee for unexpected experiences as in open skill learning.

Wulf and others [7-9] suggested that a stable GMP developed previously by constant (blocked) practice is a requisite for developing an effective recall schema by variable (random) practice. Shea et al [8] presented evidence that the relative timing of movement sequence-movement consistency- affect GMP learning across paradigms especially in view of contextual interference (CI) experiments in which random practice was superior to blocked practice. It is shown that different practice schedules create different CI, such as blocked or constant practice promoting low CI and, in contrast, serial, random or variable practice yielding high CI [9-10].

Since the time for instruction and practice of the weaker laterality is limited we focused here on the potential effects of random training following constant practice. Our purpose was to determine the effects of constant, variable and mixed practice schedules on absolute accuracy of performance and muscular activation levels, both measures of GMP development. We hypothesized that children, like adults, will exert themselves less and perform faster in the variable practice group than in the other practice groups during novel skill acquisition.

2 Methodology

We chose a skill that is an essential part of the curriculum in early judo training. O Goshi is the first lifting throw we teach and is considered a training throw. That is, it is part of the foundation for training and the parent to many other throws, but most students find they prefer other throws for actual self defense. From this throw the student learns the basic lift (with his legs, not his back), control of his partner (both to do the throw and to help him fall properly), and flexibility.

2.1 Subjects

36 children (28 males, 8 females; ages 8-10) were selected by the coach as unskilled in the execution of the O GOSHI. The children were dressed according to the code and practiced barefoot on top of padded surface.

2.2 Apparatus

A metronome that guided both the trainer and the trainee of the pace of the performance required.

Surface EMG signals were recorded by using silver/silver chloride monopolar surface electrodes (Medicotest N-00-S 30x22 mm, Olstykke, Denmark). The electrodes were positioned at a 3 cm center-to-center distance over the muscle bellies. The EMG was recorded continuously on a portable data logger MEGA ME 3000 (MEGA Electronics Ltd, Kuopio, Finland). The raw EMG signals were first treated by the preamplifiers located on the electrode leads and then filtered (15-500Hz, CMMR 110 dB and gain of 412) and digitized (12 bit with sampling rate of 1000 Hz).

Off-line, using the MegaWin software (verion 2.2) (MEGA Electronics Ltd, Kuopio, Finland). We rectified and integrated (iEMG) at 0.2 sec intervals. IEMG traces were used to measure the total activation and the magnitude. To allow comparisons of the data recorded for each child, the iEMG values of AB (Rectus Abdominis) and RF (Rectus Femoris) were normalized, as a percentage of the value of iEMG recorded for the loading phase of the skill calculated from the pre trials values.

2.3 Task

In O Goshi throw in order to lift the opponent, the proponent must bend at the knees to get her hips below the opponent's center of gravity. The proponent then clamps the opponent to his back using both arms to prevent him from sliding down or to the side. Finally the proponent straightens his legs to lift the opponent up on the back of the proponent's hips (Figure 1).

Figure.1. *Major phases in O Goshi throw: grab-load-throw.*

2.4 Procedure

The subjects were then divided randomly into 3 practice groups: 1) Constant practice (C); 2) Variable practice (V); and 3) Mixed practice (M). Each group practiced the new skill in its own distinct practice modality, using the dominant arm and side. The coach taught the skill as follows: a) C group- using an evenly paced rhythm (1-2-3), a constant "grab" height, and a single, fixed direction for approaching the opponent and executing the follow-through motion; b) V group- using a varied, random pace (1--2-3, 1-2--3, etc.), a high and low "grab" height, and forward/backward/sideways directions for approaching the opponent and executing the follow-through motion; c) M group- using C group conditions for the first 2 weeks of trials and the V group conditions for the second 2 weeks of trials.

Each group executed 120 practice trials over 4 weeks. Each group was tested three times: a pre-test before the practice period, a post-test after the practice period and, on the day of the post-test, a transfer test on the dominant hand of the contra-lateral limb/side.

2.5 Measured variables

a) Behavioral- time to completion of 6 throws in sequence (T); b) physiological- iEMGs of the bi-lateral Rectus Abdominis (AB) and Rectus Femoris (RF).

2.6 Statistical analysis

For statistical analysis purposes, we chose a one-way ANOVA with a rejection level of 0.05.

3 Results

3.1 Execution time

The variable training group had better transfer results in execution times than the C training group, which was better than the M training group improvement (Figure 2). All training groups improved significantly from Pre to Post in execution time.

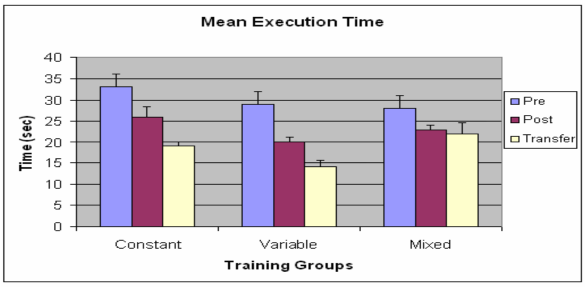


Figure 2. *Total execution times by training modalities.*

3.2 Muscle activation

Both in post and transfer tests the C group co-activated both RAs. This group also activated the RF in an alternating (reciprocal) fashion. As expected the order of RF changed from L-R-L in the post test to R-L-R in the transfer test; The V group shifted from RA co-contraction to reliance on the right RA and left RF unlike the dominant side (post) that preferred co-contraction of all four muscles. The M group linked right RA with both RFs during transfer unlike left RA and left RF linkage during post testing (Figure 3).

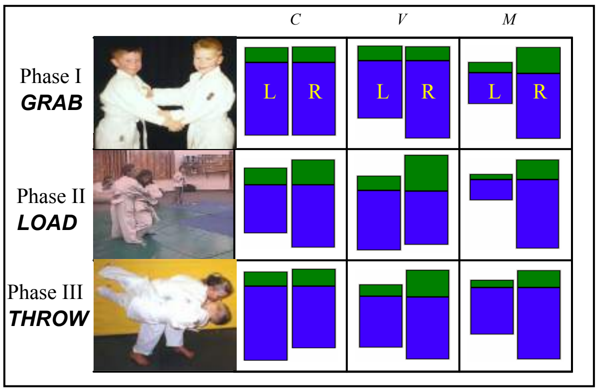


Figure 3. *Average IEMG distributions of RA and RF in relative % at different phases of skill practice by laterality (left-L, right-R) during the transfer test.*

4 Discussion

Earlier studies of fine motor skill acquisition showed preference to M group- starting with C practice followed by V practice, apparently generating a more stable GMP [11]. While acquiring the described novel skill, the V practice group achieved better transfer than the C or M group. The current results suggest that coaches of children's athletic activities should apply the V practice modality when attempting to achieve bi-lateral motor skill transfer where force control is crucial.

References

1. Schmidt, R.A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, **82**, 225-260.
2. Schmidt, R.A. (1985). The search for invariance in skilled movement behavior. *Research Quarterly for Exercise and Sport*, **56**, 188-200.
3. Schmidt, R.A.; Lee (1999). Motor control and learning: A behavioral emphasis. Champaign, IL: *Human Kinetics*.
4. .Hicks, R.E.; Gualtieri, T.C.; Schroeder, S.R. (1983). Cognitive and motor components of bilateral transfer. *Am. J. of Psychology*, **96**, 223-28.
5. Park, J.H.; Shea, C.H. (2003). Effect of practice on effector independence. *J. Motor Behavior*, **35 (1)**, 33-40.

6. Teixeira, L.A. (2000). Timing and force components in bilateral transfer of learning. *Brain Cogn*. Dec, **44(3)**, 455-69.
7. Lai, Q.; Shea, C.H.; Wulf, G.; Wright, D.L. (2000). Optimizing generalized motor program and parameter learning. *Research Quarterly for Exercise and Sport*, **71**, 10-24.
8. Shea, C.H; Lai, Q.; Wright, D.L.; Immink, M.; Black, C (2001). Consistent and variable practice conditions: effect on relative and absolute timing. *J. Motor Behavior*, **33 (2)**, 139-52.
9. Wulf, G.; Lee, T.D. (1993). Contextual interference in movements of the same class: Differential effects on program and parameter learning. *J. Motor Behavior*, **25**, 254-263.
10. Sekiya, H.; Magill, R.A.; Anderson, D.I. (1996). The contextual interference effect in parameter modifications of the same generalized motor program. *Research Quarterly for Exercise and Sport*, **67**, 59-68.
11. Benjuya, N.; Wegman, E. (2002). The effect of practice modality on bilateral motor transfer. *Proceedings of the 18th Int. Soc. of Electrophysiology and Kinesiology*, Vienna, Austria.

Actigraphy as an Objective Outcome Measure

M. B. Holm, J. C. Rogers
Department of Occupational Therapy, University of Pittsburgh, Pittsburgh, PA, USA

Abstract
For persons undergoing medical or rehabilitation interventions, in addition to data on functional performance of everyday activities, we have added actigraphy as an objective outcome measure. Depending on the study, statistical and graphic data have been collected using the Mini-Mitter Actiwatch®-64 (MMAW), the Mini-Mitter Actical® (MMAC), or the BodyMedia SenseWear® Pro (SWP). For stroke survivors, the MMAW yields data that document differences in use between the affected and unaffected side during the performance of functional activities. For older adults with late-life depression, the MMAC yields data that allow group comparisons of a randomized medication clinical trial. For artificial heart or heart transplant recipients the SWP yields data that delineate the relationship between energy expenditure and performance of functional activities. For persons with fibromyalgia, the SWP yields data that clarify the relationship between pain medication use and nocturnal activity.

Keywords
Actigraphy, function, activities of daily living, stroke, heart transplant

Mini-Mitter Actiwatch®-64

Stroke Survivors
Objectively measuring the functional movement of the affected extremity following stroke is challenging. We used the MMAW to capture differences between use of the affected and unaffected upper extremities of 220 stroke survivors at 3, 6, 9, and 12 month follow-ups. At each follow-up home assessment, stroke survivors wore an MMAW "watch" on each wrist while participating in a performance-based functional assessment of everyday activities, lasting 1-2 hours, depending on the person's level of impairment. Prior to the assessment, the following information is entered for each subject, and uploaded to the MMAW: age, gender, height, weight. The MMAW contains a 2-D accelerometer and yields "activity counts." Data are collected in predefined epochs, such as activity counts/per minute.. Data generated include actigraphs for visual analysis as well as ASCII format files with activity counts/per epoch for data analysis.

Graphic data from a 71 year old female with an ischemic stroke to the left middle cerebral artery indicates that at 3 months post-stroke movement was restricted on her affected dominant right side (and also on the unaffected side- not shown). At 9 months post-stroke data indicated that upper extremity movement was more symmetrical and that movement was greater on her dominant right side (see Figure 1).

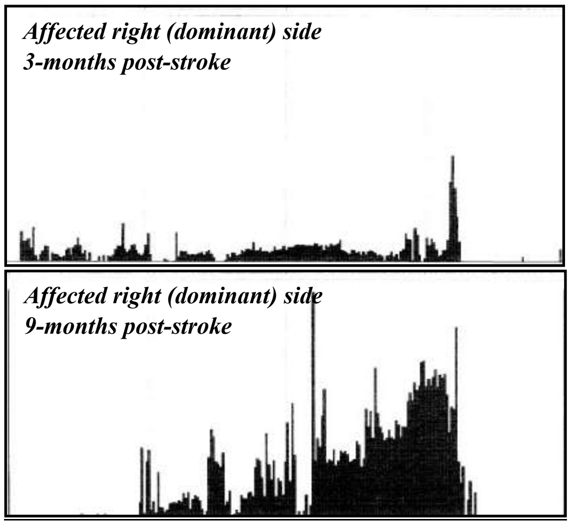


Figure 1. Movement on the affected side at 3 and 9 months post stroke.

Mini-Mitter Actical®

Late-Life Depression Patients
Sentinel markers of late-life depression can include both motor hypoactivity and hyperactivity, with older adults often restricting everyday activities during the day and increasing activity at night because of insomnia. Subjects in a 2-year randomized medication clinical trial for late life depression participate in a performance-based functional outcome assessment at baseline and at the end of Years 1 and 2. Ten days prior to each assessment, subjects are given an MMAC "watch" to wear for 7 days (see Figures 2 and 3). Prior to providing the MMAC, the following information is entered for each subject, and uploaded to the MMAG: age, gender, height, weight. The MMAC contains a multi-directional accelerometer. Data are collected in predefined epochs, such as activity counts/per minute. Data include activity counts, energy expenditure, and Metabolic Equivalents (METs) levels, and light/dark activity ratios Data generated include actigraphs for visual analysis as well as ASCII format files with activity counts/per epoch for data analysis.

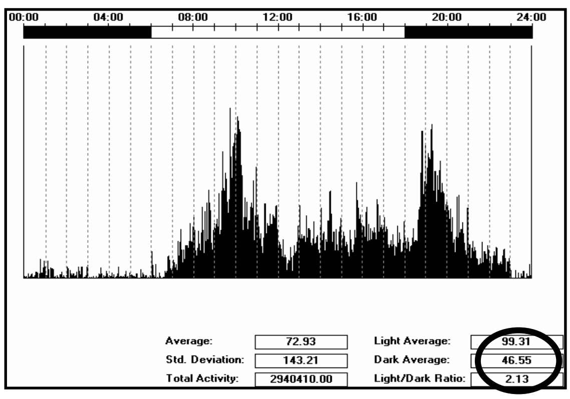


Figure 2. 24-hour "activity" levels for an age-matched non-depressed control subject.

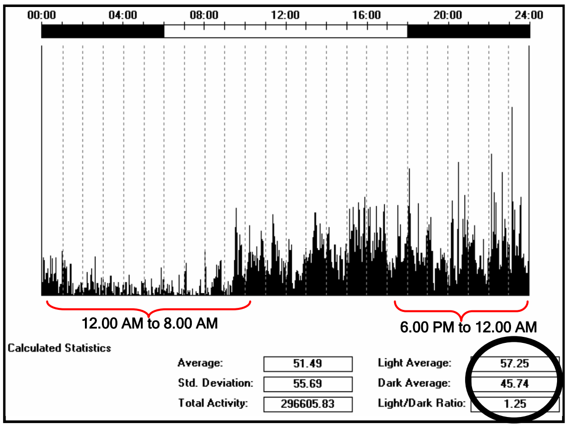


Figure 3. 24-hour "activity" levels for an 89 year old female subject with depression.

The difference between the depressed and control subjects' light/dark ratios is highlighted. The ratio should be approximately 2:1, as for the control subject (2.13:1), however, the insomnia of the depressed subject yielded a 1.25:1 ratio, indicating the lack of rest at night. This objective data was helpful for adjusting her medication, because the subject told her psychiatrist that she was sleeping well at night.

BodyMedia SenseWear® Pro

The SWP is worn on the upper arm over the triceps and does not interfere with intensive care monitoring devices. Prior to the assessment, the following information is entered for each subject, and uploaded to the SWP: age, gender, height, weight. The SWP contains a 2-D accelerometer. Data are collected 32X/minute and peak performance for each minute is recorded on multiple channels, including: accelerometer, heat flux, galvanic skin response, skin temperature, and near body temperature. Files are downloaded into .swd files which are then converted to .xls files for data analysis. Graphic data matching the .swd files are also generated for visual analysis.

Heart Transplant and Artificial Heart Patients

Patients who are wait-listed for a heart transplant often need an artificial heart as a "bridge" to transplantation. "Activity" levels were measured for 24 hours at each health status change prior to transplant, as well as 1, 3, 6 and 12 months after a status change. The SWP is worn on the upper arm over the triceps and does not interfere with intensive care monitoring devices. Data from a 29 year old male are shown in Figures 4 and 5. Energy expenditure data can be calculated from the minute-by-minute data, and the graphics provide a quick overview of status change.

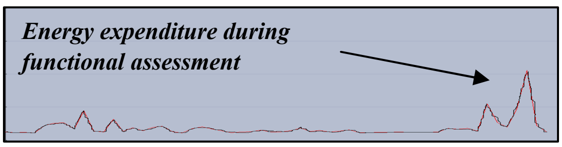


Figure 4. 24-hour "activity" level for a patient who had been on an artificial heart for 1 month.

Because those on the heart-transplant wait list are often severely debilitated, their energy levels are often below that needed for everyday activities, including dressing and toilet transfers. The SWP provided objective measures of energy expenditure over 24 hours, including the time spent performing tasks included in the performance-based functional assessment. While on the artificial heart, the subject only expended energy during the functional assessment (arrow, Figure 4) compared to post-transplant, where energy expenditure spans the 24 hour data collection period, culminating in the functional assessment (arrow, Figure 5).

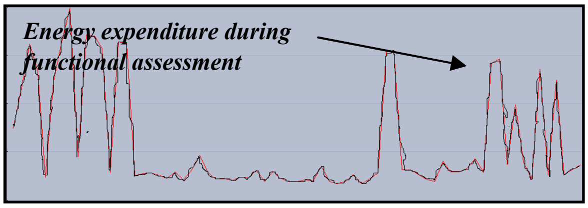


Figure 5. 24-hour "activity" levels for the same patient 1 month after receiving his heart transplant.

Fibromyalgia Patients

The SWP is currently being used as a means of providing subjects in a randomized clinical trial with data about their overall activity levels, which they can then use to monitor the relationships between activity, pain and fatigue. In addition to recording activity levels, subjects push a button on the SWP each time they take a medication for pain. Our findings indicated that subjects exhibited significantly more physical activity in the hour before taking a medication for pain at night than they did for the hour after taking the medication, or for the same time on the next night that they did not take a medication (see Figure 6).

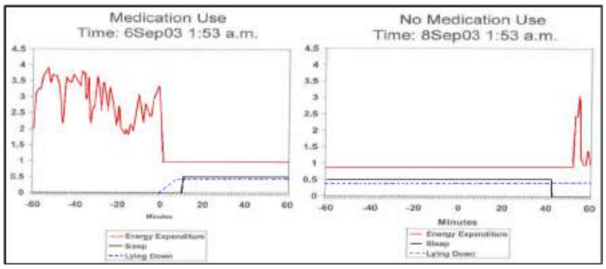


Figure 6. Comparison of activity level 1 hour before and after 0 hour (time medication was taken) on two nights.

Summary

Actigraphy can be a useful objective measure of overall daily activity for patients following medical and rehabilitation interventions. When used during interventions, it is a sensitive measure of progress. When used as a pre-post outcome measure, it can provide data sensitive to even small changes.

A new approach to gait scoring in broilers

N. Karacay
Animal Science Department, Ondokuz Mayis University, Samsun, Turkey

Abstract

Leg weakness problem manifested as reduced walking ability is regarded as one of the most important welfare problems encountered in the modern broiler industry. There are different methods with various modifications in determining the prevalence of leg weakness in live birds. The most common method which can be used for this purpose is “Gait Scoring”. Broilers grow fast and are lazy animals due to their higher body weights. They do not want to walk and spend their spare times sitting down. For this reason, it is difficult to compel them to walk during the gait scoring.

We thought that it is necessary to motivate them for walking by using a powerful reinforcer and this can be accomplished by starving them for a certain period prior to gait scoring. For this purpose, we deprived the broilers (totally 200 broilers) of feed for 10 hours prior to walking ability test. We conducted a runway test in which a group of broilers having their feeds at the feeders were used as reinforcer. Runway consists of a corridor of 300 cm (height: 70 cm; width: 100 cm) with sides covered with a cloth to prevent birds from seeing the outside. Each bird was weighed first and then put on the beginning of the runway corridor. The walking ability scores of the birds while they walking towards the birds used as reinforcer by 2 observers using a scoring of 4 points (0-3 points) were determined. We saw that the birds deprived of feed for 10 hours have walked towards the reinforcer group in an effort. We observed that the motivations of females were lower than those of males.

In this descriptive paper, this new approach and also the first results from this approach will be introduced.

Keywords

Broiler, leg weakness, gait scoring.

1 Introduction

Selection for rapid growth in broilers has also resulted in a tendency for these birds to develop leg problems. In some flocks, as many as 90% of broilers show some degree of lameness by slaughter age [6]. Leg problems are regarded as one of the most important welfare problems encountered in the modern broiler industry [4,10]. These problems have also economic results [7,10].

There are many attempts to determine the prevalence and severity of leg weakness in live birds. The most common method which can be used for this purpose is “Gait Scoring System” developed by Kestin and co-workers [6]. This system has been validated in a number of studies [7]. Recently, a new method known as Latency to Lie (LTL) was developed [9] and some modified versions of these two methods were used in assessing leg weakness [1,5].

Broilers grow fast and are lazy animals due to their higher body weights. They don’t want to walk and spend their spare times sitting down and it is difficult to compel them to walk during the gait scoring in the home pen.

We thought that it is necessary to motivate them for walking by using a powerful reinforcer and this can be accomplished by starving them for a certain period prior to gait scoring. Thus, the main aim of this study was to determine the effect of prior-to-test feed deprivation on motivation of broilers for walking to reach feed.

2 Materials and methods

2.1 Animals and husbandry

A total of 200 broilers (100 males, 100 females) were used in this study. As a material of an another study, they were grown under simulated commercial conditions in a windowed house with floor pens covering with wood shavings. Birds were fed *ad libitum* with standard broiler diets to market age (49 days). All diets met The National Research Council (NRC) requirements for poultry [8]. A 24 hour continuous lighting was provided over the entire growing period.

2.2 Equipment and carrying-out of walking ability test

To assess the walking abilities of broilers, we used a Y-shaped runway consisting of a corridor of 300 cm (height: 70 cm; width: 100 cm) (Figure 1). The frame of the apparatus was constructed using wooden material and sides covered with a cloth to prevent birds from seeing the outside.

We deprived broilers of feed for 10 hours [2] prior to test. For this purpose, 100 male and 100 female broilers were randomly selected from 10 different pens (10 from each sex in each pen) and were placed into 4 separate pens. The stocking density was approximately 8 birds/m^2 in these pens. At the night before, the feeders were removed from pens in two-hour intervals to ensure feed deprivation.

Before starting gait score measurements, the birds were carried with the crates to the runway (5 animals in each crate). The runway apparatus was placed adjacent to floor pens. Each bird was weighed first and then put on the beginning of the runway corridor. A group of 20 broilers having their feeds at the feeders were used as reinforcer to motivate birds of which the gait scores will be evaluated for walking to reach feed. The walking ability scores of the birds while they walking towards the birds used as reinforcer by 2 observers using a scoring of 4 points were determined [10]. The gait score scale is: 0: The bird walked normally with no detectable abnormality; 1: The bird was able to walk but had an obvious gait defect; 2: The bird had severe gait defect and walked only when driven; 3: The bird did not walk. Additionally, as a simple measure of motivation for walking, we determined whether or not the birds passed through the runway at the first try, if not, how much repetition were needed (repetition score).

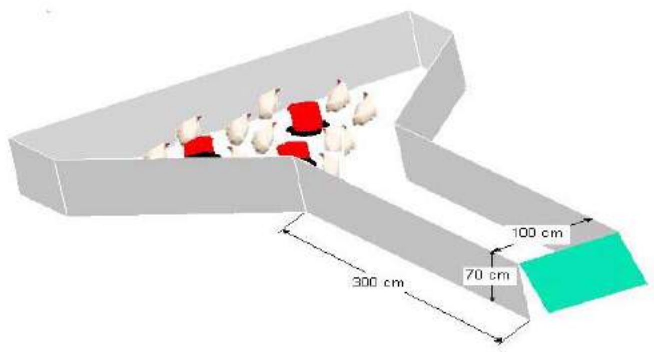


Figure 1. *The test equipment used in study.*

2.3 Statistical analysis

The differences with respect to live weights and gait scores between sexes were analysed using t-test. For the repetition scores a non-parametric alternative test (Mann-Whitney Test) was used. The all statistical analyses were performed using SPSS software package (SPSS 9.0).

3 Results

The body weights and the gait scores in each of the sex categories are presented in Table 1. As would be expected, there are important differences between sexes with respect to body weight ($p<0.001$). Similarly, the walking ability of birds was also affected by sex significantly ($p<0.001$). The males had poorer walking ability than the females (gait scores 1.21 vs. 0.75). These results were consistent with the findings of a previous study in which the same scale of measurement was used [10].

Table 1. *The effects of sex on the live weights and the gait scores of broilers.*

Traits	Males	Females	Signif. (t-test)
Live weight, g ±SE[1]	2745.13 ±32.83	2346.16 ±28.05	p<0.001
Gait score ±SE	1.21 ±7.69E-02	0.75 ±6.57E-02	p<0.001

[1] SE: Standard Error of Mean

The results of repetition score measurements indicated that the starvation motivates effectively the broilers of each sex for walking to reach feed, but the motivation of the male broilers was higher than that of females. Almost all males passed through the corridor at the first try while for a part of females several repeats were necessary (1.01 vs. 1.21 times, Mann-Whitney Test, $p<0.001$). Similarly, in a previous study [3] it has been found that the male broilers walked faster to obtain a food reward than females.

4 Conclusions

Based on the results of this study, it can be concluded that this new approach can contribute to make the gait scoring system more effective. That is because of;

- The equipment is inexpensive to purchase, easy to establish and to transport,
- The runway test is easy to carry out (on-farm conditions) and less time-consuming,
- Starvation motivates effectively the broilers of each sex for walking to reach feed and thus, the gait scores of birds can be determined more clearly and accurately.

References:

1. Berg, C., Sanotra, G.G. (2003). Can a modified Latency-to-Lie test be used to validate gait-scoring results in commercial broiler flocks. *Animal Welfare*, **12**, 655-659.
2. Bilgili, S.F. (2002). Slaughter quality as influenced by feed withdrawal. *World's Poultry Science Journal*, **58**, 123-130.
3. Bokkers, E.A.M., Koene, P. (2002). Sex and type of feed effects on motivation and ability to walk for a food reward in fast growing broilers. *Applied Animal Behaviour Science,* **79**, 247-261.
4. Danbury, T.C., Weeks, C.A., Chambers, J.P., Watermanpearson, A.E., Kestin, S.C. (2000). Self-selection of the analgesic drug carprofen by lame broiler chickens. *Veterinary Record*, **146**, 307–308, 309–311.
5. Garner, J.P., Falcone, C., Wakenell, P., Martin, M., Mench, J.A. (2002). Reliability and validity of a modified gait scoring system and its use in assessing tibial dyschondroplasia in broilers. *British Poultry Science,* **43**, 355–363.
6. Kestin, S.C., Knowles, T.G., Tinch, A.E., Gregory, N.G. (1992). Prevalence of leg weakness in broiler chickens and its relationship with genotype. *Veterinary Record*, **131**, 190–194.
7. Kestin, S.C., Su, G., Sørensen, P. (1999). Different commercial broiler crosses have different susceptibilities to leg weakness. *Poultry Science*, **78**, 1085–1090.
8. National Research Council (1994). Nutrient requirements of poultry. *9th rev. ed.*, National Academy Press, Washington D.C.
9. Weeks, C. A., Knowles, T. G., Gordon, R. G., Kerr, A. E., Peyton, S. T., Tilbrook, N. T. (2002). New method for objectively assessing lameness in broiler chickens. *Veterinary Record,* **151**, 762-764.
10. Yalcin, S., Settar, P., Dicle, O. (1998). Influence of dietary protein and sex on walking ability and bone parameters of broilers. *British Poultry Science*, **39**, 251–256.

Real world intelligent monitoring of prosthesis and footwear (REAL-PROF)

F.C. Schasfoort [1], M. de Niet [1], J.B.J. Bussmann [1], F. Thorsteinsson [2], C. Jarvis [3], C.J. Nester [4]
[1] *Department of Rehabilitation Medicine, ErasmusMC, University Medical Center, Rotterdam, the Netherlands*
[2] *OSSUR HF, Grjothalsi 5, Reykjavik, Iceland*
[3] *REMPLOY, Remploy Health Care Division,, Leeds, United Kingdom*
[4] *Centre for Rehabilitation and Human Performance Research, University of Salford, Salford, United Kingdom*

Abstract

Real-Prof is an EU funded project which aims to develop technologies for monitoring lower limb amputees and patients with orthotic footwear. The main target is to identify tissue deterioration in an amputee’s stump or in a patient’s foot before it progresses too far. The project delivers pre-commercial prototype systems that can monitor patients in real-time everyday life. Sensors are integrated into the shoe or prosthesis socket and data communicated to a data storage and communication unit worn on the waist. The data communication unit transmits data to remote clinicians via secure wireless communication. Advanced data processing tools are developed to enrich and interrogate data in order to provide clinical decision support. This study presents both Real-Prof systems and describes the results of a user evaluation from the perspectives of actual users: clinicians and patients. User evaluations are critical for further development and ‘domestication’ of innovative systems such as Real-Prof.

Keywords

Real-world monitoring, prosthesis, footwear, user evaluation

Introduction

The principal objective of the EU funded Real-Prof project was to perform the necessary research, development and validation for (a prototype of) an advanced intelligent personal health system integrated with prostheses and footwear. These systems had to be able to collect, interpret and visualize previously unavailable data from prostheses and footwear, and present these data to users (clinicians and designers of prostheses/footwear) through custom interfaces. Real-Prof had to meet two key requirements: 1) new micro-scale sensors mounted in or on the prosthesis or footwear (further referred to as hardware), and 2) novel data interpretation and representation tools (further referred to as software). The eventual systems had to provide intelligent decision support to enable early illness detection and timely and targeted allocation of health care resources. So, significant research challenges had and have to be addressed before all this is combined into a ‘real-world’ intelligent personal health system for users of therapeutic footwear and prostheses.

Footwear-system

The current hardware of the pre-commercial prototype consists of 8 tri-axial force sensors (vertical and shear forces) for 8 areas under the foot (University of Kent, UK) and an Xsens MT9 sensor that measures acceleration and angular velocity. The force sensors are integrated in an insole and the Xsens sensor is mounted on the shoe. A Sensor Interface Unit is also mounted on the shoe. The SIU provides signal conditioning, analogue to digital conversion, and serial communication to a Communication & Data Logger Unit (CDLU, Tadiran Spectralink, Israel), which is worn on the subjects' belt. The data are stored at 100 Hz and can be transmitted to an FTP server (maximal and minimal value per second). The footwear-system software development is still ongoing.

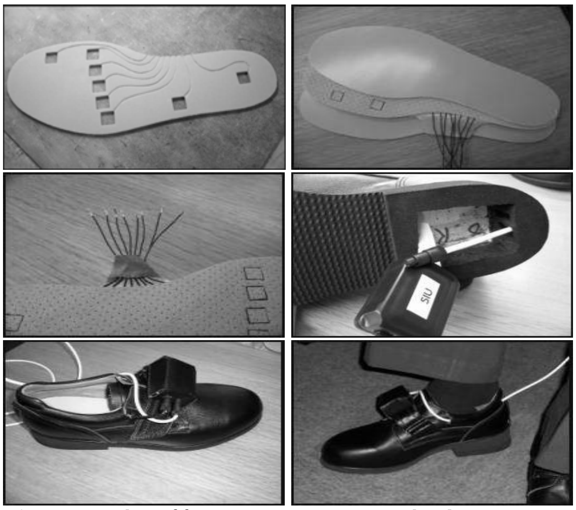


Figure 1. *Real-Prof footwear system prototype hardware*

Prosthesis-system

The prosthesis-system hardware is adopted from another EU funded project; the MAPS project (Monitoring Amputee Progress with Sensor Socket, IST-2000-27519). The prosthesis-system consists of three +/- 1 g accelerometers (one uni-axial and one bi-axial sensor, Analog Devices, Breda, the Netherlands) and three (prototype) pressure strips (RSScan International, Olen, Belgium). The bi-axial accelerometer is integrated in a Sensor Interface Unit (SIU), which is attached to the lateral side of the prosthesis with the accelerometers’ sensitive axes in the longitudinal and sagital direction. The third accelerometer is attached laterally on the upper leg just above the upper crest of the prosthesis with its sensitive axis pointing from dorsal to ventral. The pressure strips were placed laterally, ventrally and dorsally inside a double-layered silicon liner in the prosthesis, crossing important anatomical landmarks (e.g. fibular head, tuberositas tibiae). All data is collected through the SIU, which provides signal conditioning, analogue to digital conversion, and serial communication to the waist-worn Communication & Data Logger Unit (CDLU). Data is stored at 100 Hz on a MMC card and can be transmitted to an FTP server (maximal and minimal value per second).

As for the prosthesis system software, the synchronized pressure and acceleration data on the MMC can be exported as ASCII files and used in Excel, MatLab etc. through special software (RSScan International, Olen, Belgium). This software also allows the pressure strips to be divided into a maximum of seventeen units that can be considered separately in relation to anatomical landmarks.

Figure 2. *Real-Prof prosthesis-system prototype hardware*

From the accelerometer data, it is possible to detect whether a subject is sitting, standing or walking. Also, an automated accelerometry-based system for estimating initial contact (IC) and terminal contact (TC) timing information from walking patterns has been developed and validated [1].

Highly innovative projects, such as Real-Prof, face several challenges and are accompanied by uncertainties. Good usability of innovations generally optimizes health care. It is therefore important for the EU, the project partners and other stakeholders to determine whether actual end-users of the prosthesis and footwear systems consider the identified potential projects benefits indeed as actual benefits. Moreover, it is essential to identify gaps between the current status of the prototype systems and routine clinical use of the eventual Real-Prof systems by clinicians and patients. In other words, the aim of this study was to determine the needs of the footwear and prosthesis monitoring systems to get full user acceptance, i.e. to get domesticated in the long run. The fact that the Real-Prof systems were still in their developmental phases did not hamper the evaluation because the creative capacity and critical consideration of actual end-users in all phases of technological innovations are valuable to shape technological developments.

Method

User needs and hardware and software requirements depend on the aim of a monitoring system and its (clinical) application. In research and clinic, there are several application options to use such systems, e.g. to analyze and/or prevent footwear and prosthesis complaints / problems, or to determine the effects of adjustments to footwear or prostheses. It was therefore decided to perform the user evaluation from the perspective of 2 main applications: short measurements in a (standardized) laboratory environment and long-term measurements in a subjects' home environment (with an integrated real-time warning system).

Subjects included were actual end-users of the systems, i.e. clinicians (physician, CPO, physical therapist, etc.) and patients. We performed interviews and group discussions with subjects familiar with the systems, as well as patients and clinicians not familiar with Real-Prof systems. These 'inexperienced subjects' were included because the number of subjects that is actually familiar with the system currently is small, and secondly, the latter group may be biased and therefore less 'open minded'. The sessions focused on requirements of the two key elements of the project for application in a laboratory and/or home environment: the hardware (new sensors, other components of the systems, configuration and measured signals) and the software (data processing tools, output parameters, and data presentation).

Results

Because the user evaluation for the footwear system is currently ongoing, the result section will only cover the prosthesis system. We interviewed and discussed with clinicians from a variety of professions, ranging from physician, technician, physical therapist, human movement scientist, certified prosthetist & orthotist to instrument maker. Some patients had amputation of traumatic origin and others had amputation caused by vascular problems. These latter patients are usually older, their physical capacity is often reduced, and the healing of their wounds on the stump may be complicated. Such differences clearly had consequences for the requirements of the Real-Prof prosthesis system.

The objectively collected data was considered very useful by clinicians and was seen as a potentially important addition in the diagnosis, evaluation and treatment in clinical practice and research. All clinicians agreed on the importance of measuring pressure, although it would be optimal if shear forces of the stump could also be measured. Although the current hardware components are not optimal, clinicians as well as patients would use the Real-Prof prototype system in short measurements. However, only on the condition that the objective information has added value information to current practice. Disadvantages were 1) too many wires, 2) SIU and DLU too big and heavy and 3) too much time needed to apply the system. Besides this, the pressure strips are not yet reliable and valid enough. For long-term measurement in the subjects' home environment, these issues become even more problematic. The current set-up was unacceptable for use in a home environment since it would interfere daily life activities (and data would not be representative). To partly overcome these problems, applying the system should be easier and preferably the sensors and other hardware components should be integrated into the liner, socket and/or artificial lower leg.

As for software requirements, clinicians would like to quickly get the output parameters of interest after measurements. The majority would like to see 3D-pictures of the stump during the day with high pressure marked with colors. Furthermore, they would like a small report (e.g. one A4) including main parameters like the amount and type of activity (mainly walking and standing), and mean and peak pressure on specific anatomical points. It was considered useful for clinicians that they can chose some specific parameters from a larger set of output parameters for each patient individually since each patient-stump-prosthesis combination is unique.

Conclusion

For the Real-Prof prosthesis system, additional work has to be done, both for applications in a standardized laboratory environment and in the subjects´ home environment. The amount of work necessary for valid measurements in a laboratory is less, however. The software needs far more work to meet the end-user requirements than the hardware. To facilitate domestication it is suggested to take the information and feedback of end-users of the system seriously. Another important issue for future projects is to already gather detailed end-user information on software requirements in or even before development of the hardware.

Supported by the EU Information Society Technologies under Framework 5 (IST/2001-38429)

References:

1. Selles R.W., Formanoy M.A.G., Bussmann J.B.J., Janssens, P.J., Stam H.J., 2005. Automated Estimation of Initial and Terminal Contact Timing using Accelerometers; Development and Validation in Transtibial Amputees and Controls. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **13(1)**, 81-88.

Using the Actiwatch® activity monitor for measurement of physical activity and its association with anemia

R Juyal [1,3], C.P. Khokhar[3], U. Dhingra[2], R.M. Angulo-Kinzler[5], P. Verma[1], A. Sarkar[1], P. Dhingra[1], M. Osmany[1], V.P. Menon[1], M. Black[4], R.E. Black [2] and S. Sazawal[1,2]

[1]*Center for Micronutrient Research, Annamalai University, India*
[2] *Johns Hopkins Bloomberg School of Public Health, Baltimore, USA*
[3]*Departments of Psychology, Gurukul Kangri University, Haridwar, India*
[4]*University of Maryland School of Medicine, Baltimore, Maryland, USA*
[5] *Division of Kinesiology, Center for Human Motor Research, University of Michigan, USA*

Abstract

Iron deficiency affects the physical activity in adults,but data in children are limited. We used two Actiwatch (arm & leg) (Mini-Mitter Inc.) for recording 24-hour activity with one-minute epochs and evaluated association between iron deficiency and physical activity in 126 children aged 13-69 months. Hemoglobin (Hb) was estimated using coulter and children were classified as severe anemics Hb<=7.0g/dl, moderate anemics Hb 7.1-9.0g/dl, mild anemics 9.1-10.9g/dl and non anemics Hb>=11.0g/dl. Combined (leg + arm) scores were used for analysis. Multivariate analysis corrected for age and socio economic status revealed physical activity scores among severe anemics (β=-0.39, $p<0.001$) and moderately anemics (β=-0.16, $p=0.08$) were significantly lower than non anemics. Actiwatch was successfully used and accepted by children to measure 24 hours physical activity. Children with severe and moderate anemia are less active which could affect their development.

Keywords

Actiwatch, Preschool children, Iron Deficiency Anemia

Introduction

Physical activity is hypothesized to have beneficial effect on development. Associations between physical activity and Iron deficiency anemia have been reported[1, 2]. These studies evaluated either infant or school aged children and had small sample size. No data are available for physical activity with different levels of anemia in preschool children. With the advent of accelerometer devices measurement of physical activity is their application and acceptability in developing countries field setting was not known. This study was conducted to evaluated the acceptability of this methods among young children for 24 hour and to evaluate the association of physical activity with iron deficiency anemia among the children aged 12-69 months.

Methods

Study was conducted in a peri urban locality of New Delhi, India as an add-on to ongoing clinical trial. Trained nurses in the presence of project physician took Blood sample and hemoglobin (Hb) was estimated using coulter.

Actiwatch

Actiwatch is an activity monitor, designed for long term monitoring of gross motor activity in human subjects. It contains an accelerometer that is capable of sensing any motion with a minimal resultant force of 0.01g/rad/s. The actigraph counted each detected acceleration, digitizing and storing in memory the total number of accelerations per interval chosen. The weight and dimensions of this device are minimal (27×26×9 mm), weight (17g) and durable casing. Epoch length: The period of time actiwatch will accumulate activity counts before saving the sample and setting the counter to zero.In this study 1 min epoch length was set in all the actiwatches.

Procedure

Prior appointments from the parents were taken before recording the activity of the child to make sure he is not suffering from any illness. Activity recording were postponed if the child is not well. Trained scientists, after giving the instruction to the caregiver loaded the necessary identification information of the child in actiwatch, which was then tied on arm, and leg of the child for 24 hours. Activity measurement of the child was performed within the 7 days of blood sample.

Figure 1. *Observer tying the actiwatch to the arm of child.*

Statistical Analysis Plan

For univarate analysis mean activity scores in 3 anemic categories were compared to non anemic category using t-test.For multivariate analysis regression models were fitted using activity score as dependent variable and 3 dummy variables,as independent variable for anemic categories, age and socioeconomic status.

Results

The mean difference of Arm, Leg and Total activity scores among severe anemic children were lower as compared to non-anemic children (Table 1); among moderately anemic children Leg and Total showed reduced activity (Table 2), while this trend was not observed in mild anemic children (Table 3). Multivariate analysis adjusted for age and socio economic status revealed physical activity scores among severely anemic and moderately anemic children were significantly lower than non anemic children (Table 4).

Table 1. *Comparison of Activity scores for severe anemic. non-anemic children.*

	Hb<=7.0(n=16)	Mean diff (95 %CI)	P value
Total	8656.92 ±4349.14	-14217.31 (-20137.15, -8297.47)	0.00
Leg	2435.08 ±4418.32	-10384.91 (-15506.33, -5263.49)	0.00
Arm	6221.84 ±3712.09	-3832.40 (-7475.992, -188.88)	0.04

Table 2. *Comparison of Activity scores of moderate vs. non-anemic children.*

	Hb<=9.1-10.9 (n=27)	Mean diff (95 %CI)	P value
Total	20082.68 ±10591.57	-2791.55 (-8101.83, 2518.73)	0.30
Leg	11908.53 ±10837.26	-911.47 (-5748.30,3925.36)	0.71
Arm	8174.15 ±6803.87	-1880.08 (-5148.22,1388.05)	0.26

Table 3. *Comparison of Activity scores of mild anemic vs. non-anemic children.*

	Hb<=7.1-9.0 (n=27)	Mean diff (95 %CI)	P value
Tota l	18272.19 ±10237.63	-4602.04 (-9608.41,404.34)	0.07
Leg	11567.08 ±9651.94	-1252.92 (-5687.71,3181.88)	0.58
Arm	6705.11 ±5153.13	-3349.12 (-6221.78,-476.46)	0.02

Table 4. *Association of Total Physical activity with age, SES and different levels of anemia.*

Independent variables	Coefficient (Beta)	R^2	P value
Age	0.10		0.25
SES	-0.17	0.20	0.05
Hb<=7.0	-0.39		0.00
Hb 7.1-9.0	-0.16		0.08
Hb 9.1-10.9	-0.09		0.31

Discussion

Actiwatch was successfully used and accepted by children to measure physical activity. This study concluded that anemia in children is associated with a reduction in physical activity of children. These results are consistent with the studies reported by Angulo-Kinzler et al[1,3], McVeigh et al[4] and studies reporting lowered motor activity in iron deficient children[5,6]. Further studies are needed to explore causality of this relationship on larger sample.

References:

1. Anglo-Kinzler R.M, Peirano P, Lin E, Algarin C, Garrido M, Lozoff B.(2002) Twenty four hour motor activity in human infants with and without iron deficiency anemia. Early human development: 1-17.
2. Harahap H, Jahari AB, Husaini M, Saco-Pollitt C, Pollitt E.(2000) Effects of an energy and micronutrient supplement on iron deficiencyanemia, physical activity, and motor and mental development in undernourished children in Indonesia. Eur J Clin Nutr;54:S114-9
3. Angulo-Kinzler RM, Peirano P, Lin E, Garrido M, Lozoff B.(2002 Feb) Spontaneous motor activity in human infants with iron-deficiency anemia. Early Hum Dev.;66(2):67-79.
4. L A McVeigh, S A Norris and T de Wet.(2004) The relationship between socio-economic status and physical activity patterns in South African children.Acta Paediatr .93:982-988
5. Akman M, Cebeci D, Okur V, Angin H, Abali O, Akman AC.(2004 Oct) The effects of iron deficiency on infants' developmental test performance. Acta Paediatr.;93(10):1391-6.
6. Walter T.(1989 Sep) Infancy: mental and motor development. Am J Clin Nutr.;50(3 Suppl):655-61; discussion 661-6.

Measuring intracerebral release of vasopressin during the performance of aggressive behavior using microdialysis

D.I. Beiderbeck, I.D. Neumann, A.H. Veenema
Department of Behavioral Neuroendocrinology, Institute of Zoology, University of Regensburg, Regensburg, Germany

Abstract
Aggressive and violent behaviors are a public health problem. However, animal models studying aggressive behaviors are limited. Moreover, knowledge about the dynamic release patterns of neuropeptides and neurotransmitters during the display of aggression is lacking. Therefore, we have used intracerebral microdialysis to monitor the *in vivo* secretory activity of neurons within specific brain regions in male Wistar rats before, during and after the display of aggression. One day before and two days after stereotaxic implantation of the microdialysis probe, rats were tested for aggression using the resident-intruder test. There was no difference in the level of aggression between the two test days, indicating that the surgery had no negative effects on aggressive behavior. The display of aggression was accompanied by a significant increase in the release of the neuropeptide vasopressin within the lateral septum. Thus, intracerebral microdialysis can be used to measure neuropeptide release patterns during the performance of aggressive behavior.

Keywords
Aggression, vasopressin, microdialysis, resident-intruder test, septum

1 Introduction
The prevalence of aggression and violence in human society increases the need of understanding neurobiological mechanisms underlying aggressive disorders. Studies in humans and rodents indicated involvement of vasopressin (AVP) in the regulation of aggression (1-3). However, these correlations are based on measurements in postmortem brains and CSF samples, and do not reflect the physiological activity of the AVP system. To reveal the functional significance of AVP in aggression, it is important to gain knowledge about the dynamic release patterns of AVP within relevant brain regions during the performance of aggression. Therefore, we have used intracerebral microdialysis to monitor the *in vivo* secretory activity of AVP neurons within the lateral septum during the display of aggression in adult male Wistar rats.

2 Methods

2.1 Animals
Male Wistar rats (14-16 weeks of age) were kept under standard laboratory conditions (12:12 L/D cycle, lights off at 01:00 PM; 21°C; 60% humidity; food and water *ad lib.*). Ten days prior surgery, each male was housed together with a female in polycarbon observation cages (40 × 24 × 35 cm) to stimulate territorial behavior.

2.2 Resident-intruder test
The experiments were performed during the early dark phase between 01:00 and 04:00 PM. Male rats were tested twice for aggression (i.e. one day before and two days after stereotaxic surgery). The female was removed thirty minutes before the first resident-intruder (RI) test, and was returned after the test until surgery. At the time of testing, a smaller unfamiliar male Wistar rat was introduced into the home cage of the resident male for a period of ten minutes. The tests were videotaped for subsequent behavioral scoring. The following parameters were scored (using Eventlog version 1.0, October 1986, R. Hedersen): attack latency, aggressive behavior (lateral threat, offensive upright, keep down, clinch: Fig. 1), social behavior (investigating opponent, anogenital sniffing, mount), exploration and self grooming.

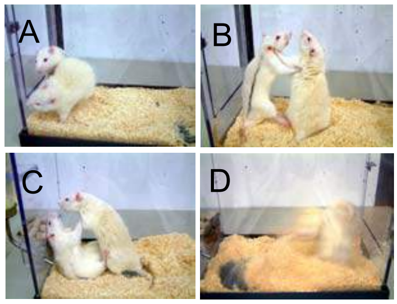


Figure 1. *Elements of aggressive behavior during the resident-intruder test (intruder marked with black lines): (A) lateral threat, (B) offensive upright, (C) keep down, (D) clinch.*

2.3 Surgery
Rats were stereotaxically implanted with a U-shaped microdialysis probe (molecular cutoff of 18 kDa, self-made probes: 4) aimed at the lateral septum (implantation coordinates (5): 0.2 mm rostral to bregma, 2.5 mm lateral to midline, 6.0 mm below the surface of the skull with an angle of 20° to avoid damage to the sagittal sinus) under isoflurane anesthesia (6). After surgery, rats were singly housed in the observation cages. One day before microdialysis, the rats were carefully accustomed to the microdialysis procedure in order to reduce non-specific stress responses during the experiment.

2.4 Microdialysis procedure
Microdialysis is based on the principle that substances in the local extracellular fluid will diffuse from a higher to a lower concentration, i.e. into the dialysing medium through a semi-permeable membrane (4, 6). As the blood–brain barrier is intact two to three days after probe implantation, this is the time window recommended for microdialysis studies (4, 6). Thus, two days after surgery, the microdialysis probes were perfused with sterile Ringer's solution (pH = 7.4) for two hours. Then, five consecutive 30-min microdialysates were collected. Following two basal dialysate collections, male rats were tested for aggression, using the RI-test, during ongoing microdialysis (see video). Thereafter, two further collections were obtained. All dialysates were instantly frozen on dry ice and stored at -20°C until AVP quantification by radioimmunoassay (7). At the end of the experiments, rats were decapitated under CO_2 anesthesia, brains were removed, frozen in ice-cold isopentane, and

stored at -20°C. 40µm Cryostat brain slices were stained with cresylviolet to verify correct placement of the probe. Only rats with correct placement were included in the result section.

3 Results

3.1 Influence of surgery on aggressive behavior

One day before and two days after the implantation of the microdialysis probe, male rats were tested for aggressive behavior. There was no difference in the attack latency nor in the aggression level between the two test days (Fig 2). This indicates that surgery had no negative effects on aggressive behavior.

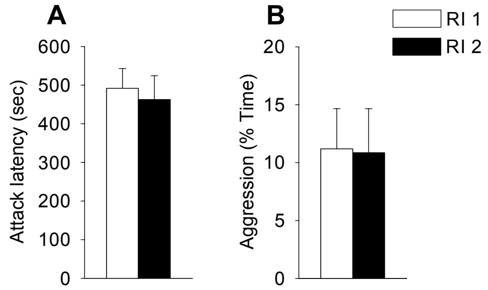


Fig. 2. (A) Attack latency and (B) aggressive behavior of adult male rats (n=11) during two RI tests, one day before (RI 1) and two days after (RI 2) surgery. No negative effects of surgery could be found on the display of aggression. Mean + SEM, ANOVA.

3.2 Septal AVP release during aggression

Based on the behavioral scoring during the second RI test, male rats were divided in aggressive (one or more attacks) and non-aggressive (no attacks) rats. Basal AVP release within the lateral septum was not different between aggressive (3.2 ± 0.5 pg/dialysate) and non-aggressive (5.3 ± 1.5 pg/dialysate) rats. In contrast, exposure to the RI-test (third dialysate sample) induced a significant increase in AVP release in aggressive rats, whereas non-aggressive rats did not show a change in AVP release (Fig 3).

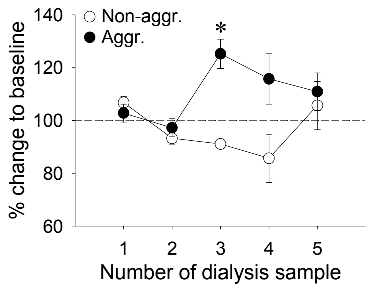


*Fig. 3. AVP release patterns in 30-min dialysates collected consecutively within the lateral septum of aggressive (n=7) and non-aggressive (n=4) male rats before (sample 1, 2), during (sample 3), and after (sample 4, 5) exposure to the RI-test. Mean ± SEM. * $p < 0.05$ vs. sample 1, 2, and vs. non-aggressive rats, ANOVA followed by Bonferroni post-hoc test.*

4 Discussion

We aim at understanding neurobiological mechanisms underlying aggressive behavior. So far, information about dynamic neuropeptide or neurotransmitter release patterns within specific brain regions during the performance of aggressive behavior is lacking. We have used intracerebral microdialysis to monitor the *in vivo* secretory activity of AVP neurons within the lateral septum of male Wistar rats before, during, and after the display of aggression. Microdialysis provides a direct approach to monitor changes of local extracellular concentrations of neuropeptides (4, 6). Dialysate samples were collected in defined time intervals. Therefore, the first two samples collected before the RI-sample could serve as internal control, making it possible to normalize for interindividual differences in basal AVP release.

Although microdialysis is an invasive technique and can cause a varying degree of tissue damage, histological examination of probe lesions suggests that the damage is limited (8). In addition, we showed that the two-day-recovery-period after surgery had no negative effects on the display of aggression, as the level of aggressive behavior did not differ before and after surgery.

We further analyzed the extracellular AVP concentrations within the lateral septum before, during and after exposure to the RI-test. Male rats that attacked the intruder showed a significant increase in AVP release, while this was not seen in non-attacking male rats. This might indicate a role for AVP in the regulation of aggression.

The extracellular AVP concentrations in the two samples collected after the RI-sample were not significantly different anymore from basal AVP concentrations. This indicates a reliable time-resolution of the microdialysis approach.

In conclusion, application of intracerebral microdialysis promises to increase our understanding of the involvement of neuropeptides as neurotransmitters/neuromodulators in the regulation of aggression.

Supported by the VolkswagenStiftung (IDN) and the Bayerische Forschungsstiftung (AV)

References:

1. Coccaro, E.F.; Kavoussi R.J.; Hauger, R.L.; Cooper, T.B.; Ferris, C.F. (1998). Cerebrospinal fluid vasopressin levels: correlates with aggression and serotonin function in personality-disordered subjects. *Arch Gen Psychiatry*, **55**, 708-714.
2. Ferris, C. (1992). The role of vasopressin in aggressive and dominant/subordinate behaviors. *Ann N Y Acad Sci*, **652**, 212-226.
3. Koolhaas, J.M; Everts, H.; de Ruiter, A.J.; de Boer S.F.; Bohus, B. (1998). Coping with stress in rats and mice: differential peptidergic modulation of the amygdala-lateral septum complex. *Prog Brain Res*, **119**, 437-448.
4. Horn, T.F.W.; Engelmann, M. (2001). In vivo microdialysis for nonapeptides in rat brain – a practical guide. *Methods*, **23**, 41-53.
5. Paxinos, G.; Watson, C. (1998). The rat brain in stereotaxic coordinates. 4th Edition. *Acadamic Press*, San Diego, USA.
6. Landgraf, R.; Neumann, I.; Holsboer, F.; Pittman, Q.J. (1995). Interleuking-1 beta stimulates both central and peripheral release of vasopressin and oxytocin in the rat. *Eur J Neurosci*, **7**, 592-598.
7. Landgraf, R.; Neumann, I.D. (2005). Vasopressin and oxytocin release within the brain: a dynamic concept of multiple and variable modes of neuropeptide communication. *Front Neuroendocrinol*, **25**, 150-176.
8. Gerin, C.; Privat, A. (1996). Evaluation of the function of microdialysis probes permanently implanted into the rat CNS and coupled to an on-line HPLC system of analysis. *J Neurosci Methods*, **66**, 81-92.

Local and global information in a dynamic model of male howler monkey interactions

G. A. Lopardo[1], J. Veà Baró[2]
[1] *Agents Research Lab, University of Girona, Girona, Spain*
[2] *Primate Research Center and Department of Psychiatry and Clinic Psychobiology, University of Barcelona, Barcelona, Spain*

Abstract

This paper describes a dynamic model of male-male interactions in howler monkeys (*Alouatta palliata*, Primates) inhabiting a patchy environment. Howler monkeys are folivorous primates living in groups with low aggressive behaviour. Solitary males approach resident males to attempt a take-over. The male–male interactions are described and analyzed. The entropy of the interactions was measured with two indices, the index of Shannon's entropy and the Fisher's information. Both measures of entropy are compared to each other and with a theoretical model. The theoretical model is constructed on the basis of the Zipf-Mandelbrot model. The behavior of the interactions is studied analyzing the local and global components of the information. From this analysis the most suitable use for each type (local and global) of information is presented. In addition a possible predictive scenario for the analysis of the interactions between males is proposed.

Keywords

Howler monkey, Information theory, dynamic model, Fisher information, Shannon entropy, Zipf-Mandelbrot model.

1 Introduction

Howler monkeys, genus *Alouatta*, is a relatively well-studied group of the Neotropical Family Cebidae, order Primates. *Alouatta* is a simple genus with six species. The six species currently recognized are distributed from southern Mexico to northern Argentina [11]. *Alouatta palliata*, the mantled howler, is distributed from southern Mexico, southern Guatemala, and south through the Central America to the Pacific coast of Colombia and Ecuador. The mantled howler is seriously threatened primarily because of forest destruction within its small geographic distribution [11].

Howlers are sexually dimorphic in body size with males larger than females (adult female weight is 84% of adult male weight). An outstanding anatomical peculiarity of the genus *Alouatta* is the possession of an enlarged hyoid bone, what acts as a resonator and amplifier when howler monkeys produce their characteristics long calls. In *A. palliata* the hyoid volume is the smallest among howlers [9]. Howlers give these loud calls most often in the morning, but another troop or extratroop animals can stimulate roaring. Since hyoids are sexually dimorphic in size, the age and sex of the howlers are somewhat revealed by their roars. Thus, it is possible that roaring may provide information about the composition of the troop. The presence of a relatively large number of males, as revealed by the vocalizations, may often deter extratroop males from attempting to invade. Male incursions have been reported for mantled howlers [2]. These male-male interactions are largely related to the distribution of food resources sought at a particular point in time [3,5,7], as well as by differences in social organization imposed by the structure of the habitat (e.g. patchiness, size of the patches) and small extragroup parties or individuals. In this arboreal habitat extragroup males are a common feature of social structure [8]. For mantled howlers, as would be predicted for a polygynous, sexually dimorphic species, male-male competition, severe male invasions and associated infanticides have been reported [2]. Therefore it seems likely that a male individual's choice of group might be influenced by the quality of environmental resources, especially in cases of habitat destruction. In these patchy habitats, howlers are usually found living with reduced resources, but even in such disturbed environments they spend time in male-male interactions.

2 Dynamic model

In the forest of southern Mexico, the mantled howlers live in groups and spend most of their time in the canopy. They live, eat and sleep in the canopy. The canopy, for the most part of this model, is so patchy that resources are limited to a reduced number of options. Male-male interactions, male invasions and takeovers arise from the complexity of social structure and habitat patchiness. This is the central assumption for the model.

Let us lay out a horizontal plane, which runs through the canopy of all the trees, and is divided into cells of unit width and height. The symbol X measures the distance from the origin, and Y measures the distance perpendicular to X. The extragroup male is considered a point in this X-Y plane and moves in a straight line for the visual sight distance in the canopy reaching the resident male. We assume that a male cannot travel exactly in the direction it desires to go because of lack of suitable branches.

The behavioral part of the model includes the male-male interactions. The problem here is to calculate the intensity of male-male interactions. Let us consider each male-male interaction as a "conflict point" (x_n). In our model we seek a measure of disorder whose variation derives the phenomenon. Fisher information (I) is a measure of the state of disorder of a system or phenomenon. This information arises as a measure of the expected disorder in a measurement [4].

For our purposes, it is useful to work with the discrete form of Fisher information (I):

$$I = \Delta x^{-1} \sum_n \frac{`[p(x_{n+1}) - (p(x_n)]^2}{p(x_n)} \quad (1)$$

A quantity related to I is the Shannon entropy H, called Shannon information [10]. This has the discrete form

$$H = -\Delta x \sum_n p(x_n) \ln p(x_n) \quad (2)$$

Like I, H is a functional of an underlying probability density function (PDF) $p(x)$. The analytic properties of the two information measures are quite different. Thus, whereas H is a global measure of smoothness in $p(x)$, I is a local measure. Hence, when extremized through variation of $p(x)$, Fisher's form gives a differential equation while Shannon's always gives directly the same form of solution, an exponential function. Graphically, this means that if the curve $p(x_n)$ undergoes a rearrangement of its points (x_n, $p(x_n)$), although the shape of the curve will drastically change the value of H remains constant. H is then to be a global measure of the behavior of $p(x_n)$. By comparison, if the curve $p(x_n)$ undergoes a rearrangement of points x_n as in Fisher information I, discontinuities in $p(x_n)$ will now occur. Since I is sensitive to local rearrangement of points, it is said to have a property of locality. Thus, H is a global measure, while I is a local measure, of the behavior of the curve $p(x_n)$.

In this work a theoretical model (based on Zipf and Mandelbrot´s developments) is introduced to compare the information measurements [6,12]. It may be noted that several ecological studies have considered the Zipf-Mandelbrot distribution to be an appropriate form for the rank-abundance relationship [1]. The Zipf-Mandelbrot distribution is given by:

$$p_r = N_1 (r + \beta)^{-1} \quad (3)$$

where p_r is the relative frequency of a particular point of rank r when ranked in decreasing order of relative abundance, β and γ are model parameters and N_1 is a normalising constant:

$$N_1 = \frac{1}{\sum_{i=1}^{n} (i + \beta)^{-\gamma}} \quad (4)$$

where n is the total number of points.

A measure of this kind should ideally be nonparametric and statistically accurate. It should be applicable to any conditions independent of point abundance distribution (x_n), and should have small bias and sampling variance in samples of moderate size.

3 Conclusions

We suggest that measures based on the Fisher information have a number of desirable properties. It is stated that consideration should also be taken of the Shannon index due to its central role in information theory and statistical physics. This means that the improved sensitivity properties of I and H make them the preferable measurement set. As well as providing a useful practical tool for ethology, behavioral ecology, the additive properties of H as well as I have implications for the theoretical study of the social behavior. This may have a more general utility with respect to other analytical or multianalytical properties of dynamic models such has been presented in this study.

Supported by UAM Programme Mobility Research Grant and UdG Dynamical Physical Agents and Ecosystems Project.

References:

1. Aoki, I. (1995). Diversity and rank abundance relationship concerning biotic compartments. *Ecological Monographs*, **82**, 21-26.
2. Clarke, M.R. (1983). Infant-killing and infant disappearance following male takeovers in a group of free-ranging howling monkeys (*Alouatta palliata*) in Costa Rica. *American Journal of Primatology*, **5**, 241-247.
3. Estrada, A. (1984). Resource use by howler monkeys (*Alouatta palliata*) in the rain forest of Los Tuxtlas, Veracruz, Mexico. *International Journal of Primatology*, **5**, 105-131.
4. Fisher, R.A. (1959). *Statistical methods and scientific inference*. 2nd ed. London, Oliver and Boyd.
5. Glander, K.E. (1978). Howling monkey feeding behavior and plant secondary compounds: A study of strategies. *The ecology of arboreal folivores*. G.C. Montgomery, Washington DC, Smithsonian Institution Press.
6. Mandelbrot, B.B. (1977). *The fractal theory of nature*. New York, W.H. Freeman and Co.
7. Milton, K. (1980). *The foraging strategy of howler monkeys*. New York, Columbia University Press.
8. Scott, N.J.; Malmgren, L.A.; Glandre, K.E. (1978). Grouping behavior and sex ratio in mantled howler monkeys. *Recent advances in primatology*, **1**, D.J. Chivers, J. Herbert (eds), NY, Academic Press.
9. Sekulic, R. (1981). The significance of howling in the red howler monkey (*Alouatta seniculus*). PhD diss., College Park, University of Maryland.
10. Shannon, C.E.; Weaver, W. (1949). *The mathematical theory of communication*. Urbana, University of Illinois Press.
11. Wolfheim, J.H. (1983). *Primates of the world: Distribution, abundance, and conservation*. Seattle, Washington University Press.
12. Zipf, G.K. (1965). *Human behavior and principle of least effort*. New York, Addison-Wesley.

An automated test of social interaction using trios of mice

M. Pratte
P3M (Functional Genomics, Pathology and Behavior), CNRS, Marseille, France

Abstract

An experiment was designed to assess the relative level of social interaction during encounters involving trios of juvenile inbred mice. This design was used to compare social behavior toward a familiar versus an unfamiliar individual in trios consisting of two familiar cage mate males plus a third unfamiliar male from another strain. The two strains used (C57/BL6 and 129S2/Sv) are frequently employed in transgenic research.

The automation of the spatial positioning was obtained by using a video-tracking program, the EthoVision version 2.3 system. In addition social behaviors were manually scored. Correlations between manual and automated parameters showed that spatial parameters provided by the EthoVision system correctly fit the level of social interaction between mice.

Notable differences were shown between the two strains C57/BL6 and 129S2/Sv: the total duration and frequency of bouts during which two individuals were close to each other were good indicators of the discrimination of social novelty.

Keywords

mice, inbred strains, social behavior, automated analysis.

1 Introduction

The social interaction test is used widely to measure anxiety-like behavior [1]. The test use a natural form of behavior as the dependant measure, namely the time spent by pairs of males rodents in social interaction (e. g. sniffing, following or grooming the partners). The test quantifies the level of social behavior between pairs of rats and it is usually based on "manual" analysis of behavior. An automated version of the social interaction test on rats using a video-tracking system was proposed [3]. Yet, the score in different pairs depends on the individual baseline level of social activity that is not estimated in this procedure.

In this paper I describe a procedure designed to evaluate the relative level of social interaction in a social system involving trios of juvenile inbred mice [2]. The procedure was applied to trios consisting of two familiar cage mate males plus an unfamiliar individual (intruder) from another strain. This design maximized the social interactions by increasing the differences between familiar and intruder males. The two strains used (C57/BL6 and 129S2/Sv) are frequently used in transgenic research. There are marked strain differences on measures of anxiety (with the 129S2/Sv mice being typically more anxious). In contrast, however, the strains do not differ on measures of cognitive performance, including habituation to a novel context.

2 Material and Methods

2.1 Animals

Eight-weeks old mice were provided by IFFA-CREDO company (St Germain sur l'Arbesle, France). Males C57BL/6Jico (C57, n = 18) and 129S2/SvPasIco (129Sv, n = 18) were housed by group of three same-strain males per cage (30x18x18 cm), in a 12 h/12 h light/dark cycle with light onset at 07:00. The behavioral testing was carried out when the mice were between 13 and 18 weeks old.

2.2 Experimental procedures

The test was performed in an open field consisting in a clear plastic box, 38 cm long, 30 cm wide, and 15 cm high. Two UV neon tubes (15 W, 254 nm) fixed 70 cm above the open field lit the apparatus (70 lux at floor level). The arena was filmed with a color CCVT Panasonic camera mounted on a stand, 1 m above the open field. The movements of the mice were analyzed on line at a rate of 5 im/sec using a commercial video tracking system (EthoVision Color-pro version 2.3, Noldus Information Technology, The Netherlands).

A colored tag fixed on the fur, between shoulders, about two hours before observation, marked each mouse. The tags were made of thick paper (10x15 mm) colored by a mixture of fluorescent pigments (Fiesta daylight fluorescent colors, Swada, London) and speed-dry nail varnish. The fluorescent pigments showed up as sharp colors under the UV light and clearly stood against the dark mice fur, thus ensuring that only the markers were detected. To avoid any reflection that could disturb the tracking, a black drawing paper covered the floor of the open field.

"Mixed-strain" trios were formed of 2 males from one strain (familiar cage mates) and 1 male from the other strain (intruder). Six encounters of 2 C57 males and 1 129Sv male, and 6 encounters of 2 129Sv males with 1 C57 male were observed for 10 min. At the beginning of the test, each mouse was placed in a different corner of the open field. Neither food nor water was provided during the observation period. After each encounter the arena was wiped with a damp tissue (Surfanios) and the black paper covering the floor was changed. All encounters started between 11:00 and 15:00.

2.3 Automated analysis

For each 10-min observation period the following parameters were analyzed:

<u>Distance Moved</u>: total distance traveled (cm). The minimal distance moved was set at 2 cm.

<u>Movement</u>: state during which the running velocity of the mice exceeds 3 cm/sec (Start velocity threshold). The state becomes "off" when the velocity drops below 1.5 cm/sec (Stop velocity threshold). Parameters were frequency (number of occurrences) and total duration (sec).

<u>Mean Distance Between Two Mice</u> (cm).

<u>Proximity</u>: State during which a mouse is within 10 cm (approximately the body length) from another mouse. Parameters were frequency and total duration (sec).

Moving To: Total duration of relative movement of a mouse towards another mouse (sec). The maximum interaction distance was set to 15 cm.

2.4 Manual rating

Some behaviors were scored with the Manual event recorder included in EthoVision version 2.3, by pressing keys during data acquisition.
Grooming: cleaning the fur or scratching (total duration).
Rearing: raised on the hind legs sniffing into the air (frequency).
AlloGrooming: cleaning the fur of another mouse (total duration).
Body Sniff: sniffing the body or the face of another mouse (frequency).
Genital Sniff: sniffing the genital region of another mouse (frequency).
For the last three parameters, the actor and the receiver were identified

3 Results

3.1 Comparison of the manual and automated data acquisition methods

The comparison between the two data-acquisition methods determined how much the parameters from automated analysis fitted manual records. Duration of Grooming and Duration of AlloGrooming did not correlate well with any of the automated analysis parameters. However, the AlloGrooming behavior occurred infrequently and therefore did not usefully describe the social interaction. This parameter was not included any more in the study. Frequency of Rearing highly correlated with Duration of Movement ($r = 0.86$, $P < 0.001$, Pearson's correlation coefficient). Frequency of Body Sniff and Genital Sniff highly correlated with Duration of Moving To ($r = 0.67$ and $r = 0.69$ respectively, $P < 0.001$).

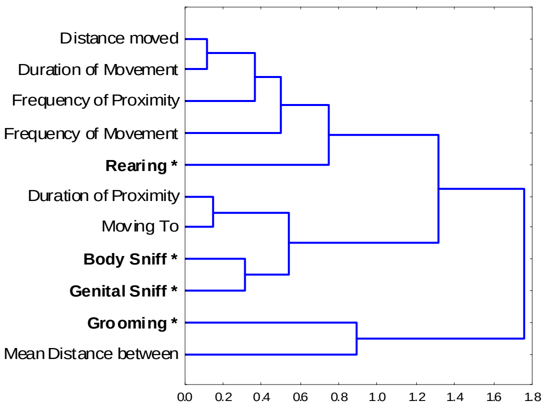


Figure 1. *Hierarchical cluster analysis of the parameters from the manual and automated data-acquisition method, based on the behavior of 18 C57 male mice and 18 129Sv male mice observed in mixed-strain trios for 10 min. The cluster analysis used Pearson's correlation coefficients and the complete linkage method. (* = manual rating parameters).*

A hierarchical cluster analysis (Figure 1) showed that the parameters from manual and automated analysis could be divided into 3 main groups: (1) Distance Moved, Duration and Frequency of Movement, Frequency of Proximity, Frequency of Rearing formed one group that represents the explorative behaviors. (2) Duration of Proximity, Moving To, Frequency of Body and Genital Sniffs quantify the active social behaviors. (3) Grooming and Mean Distance Between Two Mice represent inactive and resting periods.

3.2 Relations in mixed-strain trios

Within each mixed-strain trios, the social behavior of a male toward the familiar male was compared with the behavior toward the intruder male. The nonparametric Wilcoxon's matched pairs test for dependent samples controlled the variation due to individual differences in subjects. In mixed-strain trios regrouping two C57 males and one 129Sv male, C57 males ($n = 12$) were less frequently in proximity of the intruder than of the other C57 male (89.4 ± 29.7, and 132.7 ± 24.1, respectively; $P < 0.01$), but they stayed longer close to the intruder (144.9 ± 40.7, and 88.6 ± 18.5, respectively; $P < 0.01$). C57 males sniffed more frequently the intruder than their cage mate (Body Sniff: 7.7 ± 2.7 and 1.1 ± 0.9 respectively, $P < 0.01$) (Genital Sniff: 1.8 ± 1.5 and 0.3 ± 0.7, $P < 0.01$). In mixed-strain trios composed of two 129Sv males and one C57 male, 129Sv males ($n = 12$) approached more frequently the intruder than the familiar male (98.5 ± 37.6 and 68.5 ± 43.2 respectively, $P < 0.01$) but did not stay longer close to the intruder (157.8 ± 58.9 and 183.3 ± 47.5, N.S.). They sniffed more frequently the body and genital region of the C57 male than of the familiar male (Body Sniff: 8.6 ± 6.3 and 5.0 ± 1.8 respectively $P < 0.01$) (Genital Sniff: 5.4 ± 4.5 and 1.0 ± 1.4 respectively, $P < 0.01$).

4 Conclusion

Correlations between manual and automated parameters showed that spatial parameters provided by the EthoVision system correctly assessed the level of social behavior between males of different mouse strains. They revealed marked differences between the two strains C57 and 129Sv in how males reacted to an unfamiliar individual. Particularly the total duration and frequency of bouts during which two individuals were close to each other were good indicators of the discrimination of social novelty. When males encountered two other males, one familiar and of the same strain, and the other unfamiliar and of a different strain (intruder), the social approach differed according to the actor' strain. C57 males approached less frequently a 129Sv intruder but stayed close to it longer. Conversely, 129Sv males approached more often a C57 intruder, even if they did not remain close to it longer. In both strains body and genital sniffs toward the intruder increased. However, the experimental procedure did not allow to separate the features "other strain" from "unfamiliar individual". This will be tested in a "same-strain trios" study where the unfamiliar male forming the third individual of the trio would be of the same strain as the two other mice.

References:

1. File, S.E.; Seth, P. (2003). A review of 25 years of the social interaction test. *European Journal of Pharmacology*, **463**, 35-53.
2. Hughes, A.L. (1989). Interaction between strains in the social relations of inbred mice. *Behavior genetics*, **19**, 685-70.
3. Sams-Dodd, F. (1995). Automation of the social interaction test by a video-tracking system: behavioural effects of repeated phencyclidine treatment. *Journal of Neuroscience Methods*, **59**, 157-167.

Factor analysis to understand the inner structure of mouse agonistic behavior

O.Yu.Vekovischeva [1] , T. Aitta-aho[1], E.V.Verbitskaya[2], E.R.Korpi [1]
[1]*Institute of Biomedicine/Pharmacology, University of Helsinki, Helsinki, Finland*
[2]*Laboratory of Biomedical Statistics, Pavlov State Medical University, St-.Petersburg, Russia*

Abstract
Factor analysis was applied to understand the behavioral structure of agonistic behavior of Turku aggressive (TA) and Turku non-aggressive (TNA) mouse lines selected for high and low levels of isolation-induced male aggression. The monitoring of discrete behavioral elements to exclude subjective interpretation leads to a matrix structure of the behavior where links between the elements and their meanings remain sometimes problematic. Also, the interpretation of rare elements is difficult. The statistical comparison of resident-intruder interaction data by one-way ANOVA identified the differences between the lines in agonistic behavior in quantitative sense. The other possible differences between behavioral structures based on the opposite coping strategies were not investigated. However, Factor analysis identified some behavioral patterns and supported an ambivalent character of behavioral manifestations which are depending on the context. The analysis supported the hypothesis about different structure, not only quantity, of aggressive manifestations in TA and TNA mice.

Keywords
Aggression, mice, statistic, factor analysis

1 Introduction
A subject observation in social or environmental context to acquire the sequence of its discrete behavioral elements is wide-spread in experimental psychopharmacology (7, 10). If the aim of the observation is to minimize the motivational interpretation of the behavioral elements and to record simple behavioral demonstrations, the task of analysis is to understand meaning of the elements for the behavioral structure and to combine them in motivational categories that may suffer from subjective interpretation. Sometimes the usual statistical analysis of the behavioral alterations by drugs or other environmental factors does not clearly identify how the elements and artificially identified behavioral categories were changed. Often, it excludes elements that are observed rarely and cannot be expounded clear. Animal aggressive patterns very well relate to human psychiatric disorders and are widely involved in pharmacological testing (2, 4, 6), and, therefore the understanding their mechanisms appears important. We wanted here to test a Factor analysis that allows to group behavioral elements based on their space vectors and indicates causal relationship between them, on social behavior of Turku Aggressive (TA) with acute form of aggression towards an intruder and Turku Non-aggressive (TNA) mice displaying feebly marked agonistic behavior.

2 Method
TA and TNA mice of the 75th generation by male selection from a Swiss albino (referred to as SW) foundation stock in Turku, Finland (3, 5, 8) were used in this study. The social behavior of nine animals per line at age of 3 months was monitoring. Group-housed SW males of the same age were used as passive standard opponents.

2.1 Behavioral testing
Nine-minute encounters with unknown SW opponent in a neutral round glass arena (18.5 cm in diameter and 11 cm high) without sawdust were video recorded at 5-15 minutes after the saline injections. The arena was cleaned after each encounter. The video records were observed subsequently using a computer-assisted data acquisition system (Ethograph, 2.06, Ritec, St.Petersburg, Russia) (7, 11) and monitored for the frequency and duration of 44 discrete behavioral elements.

2.2 Statistical analysis
To compare structures of TA and TNA mouse social behavior Factor analysis was used with Equamax rotation method (SPSS 12.0.1). This is a combination of the Varimax method, which simplifies the factors, and the Quartimax method, which simplifies the variables. In many cases, rotated solution is more straightforward, allowing easy interpretation. The number of variables that load highly on a factor and the number of factors needed to explain a variable are minimized. Initially, free factor separation was analyzed suppressing absolute values of the elements less than 0.5. The order of the Factors rated its significance. Although Factor analysis is usually calculated on large number sampling points, we analyzed the drug-independent behavior (saline group, n=9) taking into account both the relative frequencies of the elements and their total durations together.

3 Results and discussion
It was shown before (5) and supported by present study that consummate aggression (fighting, biting, boxing) in TA mice take about 9-15% of the social behavior on neutral territory while TNA agonistic patterns take about 3-7%. The TNA agonistic behaviors included defense by ambivalent vertical postures, freezing or evasions from the contacts and a few short-term expressions of aggression, such as biting and fighting, occurring in one or two subjects per group. It happens usually in response to partner's provocation.
Factor analysis identified 2 and 3 main Factors in TA and TNA mice, respectively (see Table 1). The backbone of TA mouse behavior appeared to be emotional expression of the aggression. This might be interpreted as "antisocial" aggression (1) and be associated with high levels of impulsivity (9). Absolute aversion to any contacts with a partner by kicking, evasion and freezing was identified. At the same time, a high level of correlation between these factor and "passive contact", when the approach is initiated by the partner, mean interrelation between the behaviors: TA males avoided touch immediately SW approached opponent.

***Table 1.** The content of main behavioral factors deciding the behavior of TA and TNA mice.*

Factor analysis performed for data on	N	Main Factors (based on Rotation Sums of Squared Loadings)	% of variance that the factor explains	Content of the Factor (those behavioral elements are presented that have the correlation of >0.5 between them and the rotated Factor) (N – number of animals, RF- relative frequency, TD – total duration)
TA mice	9	1	9.4	Passive contact with the partner (RF&TD) - Kicking (RF&TD) - Evasion (RF&TD) - Sitting (TD) - Freezing (TD)
		2	6.5	Biting (RF) - Rushing to the attack (RF&TD) - Tremor (RF&TD)
TNA mice	9	1	12.2	Tail rattling (RF&TD) - Threat (RF&TD) - Climbing under (RF&TD) - Tossing the partner (RF&TD)
		2	9.5	Lying with sniffing (RF&TD) - Palpebral closure (RF&TD) - Approach (RF)
		3	8.6	Bite (RF&TD) - Sniffing of the partner's body (RF&TD) - Sniffing of anogenital partner's area (RF)

On the contrary, the biting of TNA mice can be related to previous exploration of the partner. It might be suggested that a provocation by the partner only was the reason of the bite. Behavioral structure of TNA males contains ambivalent aggressive elements such as tail rattling and threat, which are closely related to climbing under and tossing the partner. The last mentioned elements might be classified also as ambivalent aggressive elements. If the partner does not provoke any aggression, the contacts (approach) from the direction of TNA males were accompanied by passive exploration of the environment (lying with sniffing) and palpebral closure that means a comfort behavior in presence of the partner.

4. Conclusions

Factor analysis confirmed statistically that TA and TNA mice demonstrated opposite coping strategies despite of presence of an aggressive pattern in both of them. The behavioral structure of expressed aggressions was different suggesting distinct regulative mechanisms.

The authors are indebted to Professor Kenneth Sandnabba and his associates for the provision of mice.

References:

1. De Waal, F.B. (2004). Evolutionary ethics, aggression, and violence: lessons from primate research. *J Law Med Ethics*, **32**, 18-23.
2. Ferrari, P.F., P. Palanza, S. Parmigiani, R.J. Rodgers (1998). Interindividual variability in Swiss male mice: relationship between social factors, aggression, and anxiety. *Physiol Behav*, **63**, p. 821-7.
3. Lagerspetz, K.M.J. (1964). Studies on the aggressive behaviour of mice. *Ann.Acad.Sci.Fenn*, **B**, p. 133 pp.
4. 4. Miczek, K.A., S.C. Maxson, E.W. Fish, S. Faccidomo (2001). Aggressive behavioral phenotypes in mice. *Behav Brain Res*, **125**, 167-81.
5. 5. Nyberg, J.M., O. Vekovischeva, K. Sandnabba (2003). Anxiety profiles of mice selectively bred for intermale aggression. *Behav Genet*, **33**, 503-511.
6. Podhorna, J.,M. Krsiak (2000). Behavioural effects of a benzodiazepine receptor partial agonist, Ro 19-8022, in the social conflict test in mice. *Behav Pharmacol*, **11**, 143-51.
7. Poshivalov, V.P., I.A. Suchotina, E.V. Verbitzkaya (1988). Ethological pharmacology of anxiogenic and anxiolytic drugs. *Ann Ist Super Sanita*, **24**, 453-9.
8. Sandnabba, N.K. (1996). Selective breeding for isolation-induced intermale aggression in mice: associated responses and environmental influences. *Behav Genet*, **26**, 477-88.
9. Swann, A.C. (2003). Neuroreceptor mechanisms of aggression and its treatment. *J Clin Psychiatry*, **64 Suppl 4**, 26-35.
10. Troisi, A. (2005). The concept of alternative strategies and its relevance to psychiatry and clinical psychology. *Neurosci Biobehav Rev*, **29**, 159-68.
11. Vekovischeva, O.Y., T. Aitta-Aho, O. Echenko, A. Kankaanpaa, T. Seppala, A. Honkanen, R. Sprengel, E.R. Korpi (2004). Reduced aggression in AMPA-type glutamate receptor GluR-A subunit-deficient mice. *Genes Brain Behav*, **3**, 253-65.

Combined use of the Observer and Ethovision in the rat Social Interaction test

B. Biemans, R. Willems, H. Duytschaever, S. Lopes, and D. Ashton

CNS Discovery Research, Johnson & Johnson Pharmaceutical Research & Development, a division of Janssen Pharmaceutica NV, Beerse, Belgium.

Abstract

In the study of animal behavior, a trade-off always exists between speed (automation) and precision (detail) of the analysis. In this paper, we compare manual scoring in The Observer with automated analysis by Ethovision, as well as automated tracking of analogue versus digital images. It became clear that Ethovision can only providing reliable information on measures directly related to tracking patterns, and that The Observer is necessary for detailed behavioral analysis. Furthermore, a discrepancy was found in analyzing digital and analogue files for traveled distance. We await further development of the software.

Keywords

Social interaction, manual scoring, automation, digital files, behavior

1 Introduction

Manual scoring is a time consuming business, and automation of the analysis of behavior is thus desirable. On the other hand, in a test like the social interaction test, it is of great importance to get as detailed information as possible about an animal's behavior. The social interaction (SI) test [1] is used to study deficits in social behavior. In the past, we used Ethovision (EV) alone to score behavior and track patterns, but we needed to trace back individual events in time, so recently we started using The Observer (OBS). To obtain more detailed information about social interaction, an attempt was made to tease apart different behaviors belonging to SI category. For this, the event-logging keyboard (ELK) was switched from standard mode to the press-hold down mode. By combining the two programs, we could compare standard measurements of social behavior and exploration available in EV (proximity, rearing) to our manual scoring in OBS. With the upsurge of digital files, we need to identify disparities between analogue versus digital tracking for comparing new with old data. For this we evaluated distance traveled calculated by EV of tracked live images versus digital images of the same sessions. Test runs revealed a discrepancy of 5-10%.

2 Methods

In the SI test, two unfamiliar male rats are placed together in an arena, and videotaped for 5 minutes. Social and non-social behaviors are scored. Rats are painted yellow and blue with Jerome Russell "punky colour" gel (Figure 1).

The experimental setup

A Panasonic color CCD camera is suspended vertically 1.5 m over a black circular arena (Figure 1). The camera is connected to two separated PCs. On the first pc, analogue images are converted via a frame grabber (Picolo mounted on PCI-board) and analyzed online with Ethovision Color-Pro (v3.0.15, Noldus IT, The Netherlands). On the second pc, analogue video frames are digitized with Canopus' Mediacruise AD converter (v2.24.000, UK) and imported into OBS (v5.0.25, Noldus IT, The Netherlands). All programs run under Windows 2000 Professional edition. Light intensity in the room is approximately 20 Lux.

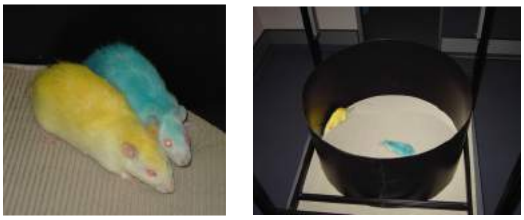


***Figure 1**. Dyed rats and experimental setup*

Ethovision

In the arena two objects are defined on the basis of color. EV quantifies the following parameters: Proximity, Total distance, Time in central zone (inner 33% of arena).

Observer

The following behaviors were scored manually for each rat, using the ELK in the standard mode: Rearing, Self-grooming, Inactive, Social Interaction (SI), Exploration (default). Behaviors are mutually exclusive, and Explore is chosen as default, since it is the most prevalent of all behaviors. With this setting a behavior is active until another behavior is pressed.

2.1 Scoring sub-behaviors of SI category

To obtain a more detailed analysis of the group of social behaviors we define as "SI", we decided to divide this category into 5 "sub-behaviors". These are described in Table 1. To validate this new method of scoring, 32 digital files (previously scored with old method) were re-scored only for the SI sub-behaviors. With these settings a behavior is active as long as the respective key is pressed down. Time in SI scored in ELK standard mode ($StMd^{ELK}$) was compared to cumulated time in sub-behaviors of SI, scored with ELK in press and hold down mode (PHD^{ELK}).

Manually scored behaviors	
Event-Logging Keyboard	
$StMd^{ELK}$	*PHD^{ELK}*
Exploration (default)	
	Body investigation
	Genital investigation
Social Interaction	Groom partner
	Close following
	Play/ play fighting
Self grooming	
Inactive	
Rearing	

***Table 1.** Behaviors scored manually in OBS with common method (left column), and more detailed method (right column)*

2.2 Manual versus automatic scoring of behavioral parameters

Since we want to automate our setups to the greatest extent possible, we compared several parameters available in EV that are identical to, or resemble behaviors scored manually in OBS. These parameters were rearing and proximity. Rearing frequency can be calculated in EV based on body surface area, and proximity should be similar to manually scored SI. For rearing, the threshold for decline in surface area was set to 25 or 50 %, in combination with averaging intervals of 5 or 40 samples, with a fixed sample rate of 12.5/s. Proximity was defined as no more than 20.5 cm distance between the two objects in the arena. Proximity was compared to individual SI scores.

2.3 Digital versus analogue images

For future reference, we were interested to know if EV's calculations of parameters such as traveled distance are identical whether digital files or live images are used. 24 Tracks were recorded with high or low resolution of the frame grabber, and these video images were digitized with compression method MPEG1 or MPEG2. The arena was calibrated with an A4-paper in the arena, and images of this calibration procedure were digitized. This image was used for arena calibration for tracking of digital files.

2.4 Statistics

Linear regressions were performed to determine relations between parameters. Significance level was set at $p<0.05$.

3 Results

3.1 Scoring of SI sub-behaviors using PHD^{ELK}

With some training, our experienced observer was able to master scoring all sub-behaviors of the SI category. There was a strong correlation between cSI and SI (Figure 2). The deviation of cumulated sub-behaviors (cSI) compared to the original SI scores obtained with the standard mode of OBS had a normal distribution around zero. The average deviation was calculated as follows: $|cSI\text{-}SI| / SI*100\%$, and was on average 7.3 ± 1.1 %. No relation was found between cSI and absolute deviation i.e. there was no increase in error with higher cSI scores.

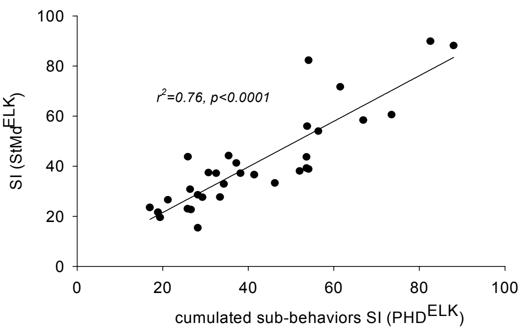


Figure 2. Correlation between sub-behaviors of SI and previously scored SI with the two different ELK modes. N=32.

3.2 Manual versus automatic scoring of behavioral parameters

In none of our tests, a relation was found between manual scoring of rearing frequency and rearing identified by EV. Different averaging intervals and percentages of original surface area were tested on tracks of rats from different treatment groups (data not shown), but none of them resulted in even a trend of a significant correlation. Proximity correlated only weakly with manually scored SI (R^2=0.38, p=0.1) (data not shown).

3.3 Discrepancy digital versus analogue images

EV calculated total distance from live images and from the digitized images at different frame grabber resolutions and compression methods (Table 2). An absolute deviation of about 5% in total distance arose from this analysis. This deviation was not always in the same direction (+ or –, as well as + *and* –). With the frame grabber set on medium resolution compared with digitized files with MPEG1 compression, the deviation went up to 10-20%, depending on whether an input filter threshold was used for small apparent movements caused by noise of the system.

		Digital			
Live resolution		**MPEG2**	**Δ**	**MPEG1**	**Δ**
High	**Filter 0**	4.5 ± 2.4	<	5.5 ± 3.2	<>
	Filter 0.5	4.7 ± 3.0	>	4.0 ± 4.0	<>
Medium	**Filter 0**	nm		11.9 ± 6.6	<>
	Filter 0.5	nm		18.2 ± 7.9	>

Table 2. *Average percentage absolute deviation ± sd of digital from analogue files for each rat, in total distance. Direction of change is also indicated. Filter= input filter. nm= not measured. Shaded cells reflect files with dissimilar resolution.*

Discussion

Analysis of behavior in the SI test is time-consuming, since it has to be done manually. Both OBS and EV are necessary to suit our needs for the SI test. Some behaviors can be best analyzed by EV, where others cannot. Alongside an increased demand for automation of setups remains the need for as detailed studying of behavior as possible. With some practice, we were able to reliably shift between the two ELK modes gaining more detailed yet comparable results. The measure of social interaction available in EV (proximity) is not accurate enough to replace manual scoring of either method. Not only are detailed behaviors not distinguished, but the weak correlation between SI and proximity underscores that EV cannot discern between coincidental proximity (passing by) and true social contact. The upcoming head-tail distinction in EV will be a step forward to interpreting close body contact between animals. For now, its use for behavioral analysis remains limited to measures related to distance and time spent in zones. Rearing frequency as identified by EV is not a reliable parameter, at least not under our experimental conditions. It did not relate to manual scoring, and could not be enhanced by changing settings. In order to keep the number of behavior keys in OBS manageable, it would be very useful if such parameters were available. With modern techniques, a shift from analogue to digital processing of data is taking place. The discrepancy in the tracking of EV between digitized images versus the online images remains to be scrutinized. Possibly it is due to small resolution differences between the frame grabber and digital files. This is however not supported by the fact that files with a larger incongruity in resolution did not have larger deviations. From the results it appears that the deviation is greatest with low resolution per se. Up to now, using both EV and OBS remains the best solution for the SI test, as the programs each offer unique features. We embrace any new developments in the software such as head-tail distinctions, to enhance automatic tracking and eventually replace manual scoring of behavior.

References:

1. File, S.E.; Hyde, J.R. (1979). A test of anxiety that distinguishes between the actions of benzodiazepines and those of other minor tranquilisers and of stimulants. *Pharmacol Biochem Behav*, **11**, 65-69.

Scale for the assessment of infant developmental competencies. A new instrument for early screening in Autism Spectrum Disorders

Apicella, F.; Maestro, S.; Cesari, A.; Grassi, C.; Manfredi, A.; Muratori, P.; Petrozzi, A.; Polidori, L.; Muratori, F.
I.R.C.C.S. "Fondazione Stella Maris" – Università di Pisa - Italy

Abstract
The Infant Behavioral Scale (IBS), a grid for early screening in Autism Spectrum Disorder (ASD), will be introduced. Methodological issues in using IBS through computer-based coding system as a useful tool for detection of social behavior in the first 18 months of life will be discussed.

Keywords
Infant Behavioral Scale; Autism Spectrum Disorder; Retrospective video-analysis; Social behaviors; Computer-based coding system.

1 Introduction
Autism Spectrum Disorder is tipically diagnosed between 2 and 4 years and presents with qualitative impairments in social interaction, communication and repetitive patterns of behavior (1). Despite the overwhelming evidence of precocious neurodevelopmental substrates, clinical onset of autism remains an unresolved question. Factors interfering identification of autism onset in the first years of life regard (a) the lack of a biological marker, (b) the heterogeneity of the disorder and its developmental pathways, (c) the lack of specific criteria appropriate to infancy (2). However, many parents report concerns about development and the existence of symptoms when the child is around 12-18 months of age.
Despite pragmatic constraints, retrospective analysis of Home Movies (HM), taken by parents prior to the time of diagnosis of Autism Spectrum Disorder (ASD), has strong ecological validity (2). HM represent an excellent option for providing direct observation of the earliest features associated with autism and the only instrument to study and to understand the actual course of onset in infants with autism (3).
Previous researches made through HM technique have identified several potential indicator of risk for an early ASD. These indicators concern eye contact, affective states, attention and communication (3, 4, 5).

2 Objectives
To figure out how ASD become evident during the first year of life in children with autism is a priority in the forefront of the scientific community (2); analysis of how infants interact with their caregiver can give insight into the social deficit characteristic of autism.
No available interaction and social behaviors coding scale exists for this specific purpose: hence the need for developing a tool for an early detection of anomalies in social behaviors and in interactions between infants and their parents.
The aim of this paper is twofold: (a) to introduce the Infant Behavioral Scale (IBS), an instrument for codify the social interaction between infants and their parents in the first 18 months of age; (b) to address methodological considerations on the implementation of a computer-based coding program (The Observer 4.0®) (6).

3 Organization of the scale and glossary.
On the basis of the instruments utilized in previous researches, a scale for the detection of social behavior in the first 18 months of life were introduced to study the early interaction between infants and their caregivers and to analyze the distribution of the child's spontaneous attention to social and non social stimuli.
The IBS was specifically optimized to code behaviors as they appear in HM.
The IBS consists of 22 items describing the behavioral components of interactions between the infant and his/her caregivers. The items are divided in 4 sub-scale:
1) *The Grid of Basic Behaviors* is composed by 6 items. This sub-scale consists of a repertoire of typical motor activity that the infant uses to know, to explore and to interact with the environment:
1.1 Looking (a. at people; b. at object; c. around): The child directs his/her eyes towards an object, or a human face, or simply looks around. The coders must be sure about the child's intention of looking.
1.2 Orienting (a. towards people; b. towards object; c. to name): the child assumes a spontaneous gaze direction towards a source of new sensory stimulation (auditory, visual or tactile) coming from a person or from an object or towards someone when the person calls his/her name.
1.3 Gaze following (a. movement of a person; b. gaze of another person; c. movement of an object): the child shifts his/her gaze to follow the gaze or the trajectory of another person, or object.
1.4 Smiling (a. smile I; b. smile II; c. smile III; d. at people; e. at object; f. responsive; g. soliciting): the child intentionally smiles at human/human-like face or at object.
1.5 Seeking conctat (a. with person; b. with object): the child employs spontaneous and intentional movements to reach contact with a person (part of the body, face) or with an object.
1.6 Explorative activity (a. with person; b. with object): the child touches something (person or object) to find out what it feels like. The exploring activity may be done by hands, by mouth or by other sensory-motor actions (shaking, throwing, etc.).
2) *The Grid of Complex Behaviors* is composed by 11 items. This sub-scale consists of a number of complex behaviors that the child display within the interactions with the caregivers. The items represent inter-subjective abilities that arise in the first 18 months of life:
2.1 Enjoying (a. with other person; b. with an object): the child finds pleasure and satisfaction experiencing a physical and/or visual contact with a person or with an object.

2.2 Sintony: the child shows a syntonic response to the other's mood; he/she shows signs of congruous expressions to affective environmental solicitations.
2.3 Anticipation of other's intention: the child makes anticipatory movements predicting the other's action.
2.4 Communicative gestures: the child displays use of social gestures.
2.5 Referential gaze: the child shifts his/her gaze towards the caregiver to look for consultation in a specific situation.
2.6 Soliciting: the child displays a verbal, vocal, or tactile action to attract the partner's attention or to elicit an other kind of response.
2.7 Accepting invitation: the child's behavior is attuned to the other person's solicitation within 3 seconds from the start of stimulation.
2.8 Offering himself to another person: the child offers parts of his/her body to the other person.
2.9 Imitation: the child repeats, after a short delay, another person's action
2.10 Pointing (a. of comprehension; b. of requesting; c. declarative): the child shifts his/her gaze towards the direction indicated by another person or indicates in order to obtain an object or to share an experience with another person.
2.11 Maintaining social engagement: the child takes up an active role within a two-way interaction in order to keep the other person involved. The child interacts, vocalises and maintains turn taking.
3) *Grid of Vocalizations* is composed by 3 items. The items of this sub-scale regard the vocal activity of the infants:
3.1 Simple vocalization: the child produces sounds (vocalizing, cooing or babbling) towards people or objects.
3.2 Meaningful vocalization: the child intentionally produces sounds (onomatopoeic or meaningful words) with a stable semantic meaning.
3.3 Crying: the child starts crying after a specific/non specific event.
4) *Grid of Caregiver's Behaviors* is composed by a number of items categorized in two general area. This sub-scale consists of behaviors that the caregiver do within a sequence of interaction for stimulate the child or for regulate his/her affective states.
4.1 Regulation (a. up; b. down): the caregiver modulates the child's arousal and mood. The caregiver may act to either excite (up-regulation) or calm-down (down-regulation) the child.
4.2 Stimulation: within a sequence of interaction, the caregiver stimulates the child by
a. Requesting attention; b. Vocalizing; c. Naming him/her; d. Gesturing; e. Touching him/her; f. Showing him object.

4 Computer-based coding system

Previous retrospective researches have largely used pencil and paper methods of coding (2, 3, 4, 5). The current study attempt to explores the utility of computerized application for coding. The Observer 4.0® (6), a software package designed for the collection, management and analysis of observational data, was configured for the application of the IBS. The four sub-scale was set as behavioral class and each target behavior is assigned a unique key. Each sub-item (i.e.: looking/at people; looking/at object, etc.) was set as a modifier.
The items are divided in two type: state (1.1; 1.6; 2.1; 2.2; 2.11; 3.3; 4.1) and event (1.2; 1.3; 1.4; 1.5; 2.3; 2.4; 2.5; 2.6; 2.7; 2.8; 2.9; 2.10; 3.1; 3.2; 4.2).
In each sub-scale, a passive state-behavior named "null" was added to address time intervals between two active state-behavior.

4.1 Coders training and inter-rater reliability

Four raters were trained to use the computer-based coding system by viewing and coding several sample video-segments, to familiarize with type and quality of the videos and with the meaning of the items, to acquire competence in coding procedures and to learn to correctly identify the items.
Inter-rater reliability was calculated by The Observer 4.0.®
Raters were required to achieve a satisfactory agreement (Cohen's Kappa >/= 0,7) with an expert clinicians in six matched sequences (two for each range of age: 0-6 months, 6-12 months, 12-18 months).
The tolerance window regarding time discrepancy between items coded was set at 1 second for event-behaviors and at 3 seconds for state-behavior. The same item recorded outside of this window was reported as a coding error and was considered a disagreement.

5 Coding procedures

HM are usually edited by an expert psychologist, that select all the segments where the infant is visible and involved in interactions with their caregivers.
Each segments is coded by a trained coder and assigned a unique data-file. The sub-scales are administered separately to allow the coders to focalize attention on a restricted number of items.

6 Sample coded segments

Two coded segments regarding sequences of interaction between 2 children and their caregivers will be graphically presented to demonstrate the feasibility for using computer-based coding technologies and to show the links among the 4 sub-scales.

References:

1. American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorder* (4th ed.). Washington, DC: Authors.
2. Baranek, G.T.; Barnett, CR; Adams, E.M.; Wolcott, N.A.; Watson, L.R.; Crais, E.R. (2005). Object play in infants with autism: methodological issues in retrospective video-analysis. *The American Journal of Occupational Therapy*. 59(1):20-30.
3. Maestro, S.; Muratori, F.; Cesari, A.; Cavallaro, M.C.; Paziente, A.; Pecini, C.; Grassi, C.; Manfredi, A.; Sommario, C. (2005) Course of autism signs in the first year of life. *Psychopathology*. 38:26-31
4. Maestro, S.; Muratori, F.; Cavallaro, M.C.; et al. (2002). Attentional skills during the first 6 months of age in Autism Spectrum Disorder. *J Am Acad Child Adolesc Psychiatry*. 41:1239-1245
5. Maestro, S.; Muratori, F.; Barbieri, F.; et al. (2001). Early Behavioral Development in Autistic Children: The First 2 Years of Life through Home Movies. *Psychopathology*. 34(3):147-52.
6. Noldus Information Technology. (2002). *The Observer®. Reference Manual 4.1*. Wageningen.

Association of social referencing with developmental scores of preschool children

U. Dhingra[2], P. Verma[1], P. Dhingra[1], A. Sarkar[1], M. Osmany[1], R. Juyal[1], R.E Black[2], V.P. Menon[1], J. Kumar[1], A. Dutta[2], G.S Hiremath[2], B. Lozoff[3], M. Black[4], S. Sazawal[1,2]

[1] Center for Micronutrient Research, Annamalai University ,Delhi, India, [2] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore USA; [3] Center for Human Growth and Development, University of Michigan, Michigan, USA, [4] .Department of Pediatrics, University of Maryland, Maryland, USA.

Abstract

Though theoretical basis for the role of social referencing in normal child development exists, but objective evidence evaluating this linkage is lacking. In a clinical trial, we had social referencing and development data available on 1240 preschool children. To evaluate this linkage, we designed CMROBS using The Observer® 5.0 software installed in a laptop to assess the social referencing and Bayley Scales of Infant Development (BSID-II) to assess the mental and motor development of children. Child's social referencing skills and parent's behavior were coded during the observation session. On the same day, development was assessed providing the outcomes as psychomotor developmental index (PDI) and mental developmental index (MDI). Regression models were fitted to evaluate the association of social referencing skills with developmental scores. Among the social referencing skills, only child verbalizing with mother had an association with the overall mental and motor development.

Keywords

Social referencing, Child development, preschoolers, The Observer 5.0, BSID-II, CMROBS.

1. Introduction

Social referencing is the process of communication between the parent and the child where the infant uses expressions from parent to develop understanding of the event and regulates his/her behavior [1]. This is a measure of socio- emotional development of children that illustrates the infant's ability to use expressions and actions of mother in conditions of uncertainty and novel situations. Social referencing skills develop at around 7 months of age and then progress rapidly, starting with sensitivity to emotions and leading to an understanding of those emotions. Their evaluation provides a reflection of child's responsiveness to emotional expressions and social stimuli. Contemporary views suggest that early emotional development is a part of and leads to subsequent overall development of the child. Emotional responses have been said to define temperament, which in turn can be regarded as an individual's characteristic set of thresholds and latencies for emotional/ social responses. Studies have suggested that emotions and social interaction in early child development may have a role in two central adaptive/ organizational functions of motivation and communication. Emotions motivate the child either to approach or to withdraw from a situation, to maintain or to terminate stimulation and communication serves to guide the child's social functioning and exchanges [2].

Although theoretical evidence suggests that social referencing and social interaction are essential for normal child development, empirical data evaluating this association are lacking. We examined association of social referencing with developmental scores of preschool children**.**

2. Methodology

The study was conducted among 1240 children aged 1-3 years in Sangam Vihar, a low to middle income peri urban population in Delhi, India.

2.1 Social Referencing Assessments

Social referencing assessments were undertaken by developing a social referencing project CMROBS using The Observer® 5.0 (Noldus Information Technology), a direct observation event-time recorder software installed in a laptop [3]. Configuration file was developed to observe and code social referencing during the observation session. The social referencing project was configured by using case sensitive social referencing key codes, open ended duration and continuous mode of recording observations. Age, gender and child's identification no were defined as independent variables. Multiple components within ten behavioral classes namely, Social Referencing to Mother; Contact with Caregiver, Proximity to Mother; Mother's Affect; Child's Affect; Latency to first touch; Contact with and Distance to Stimulus toy were configured to record child's social referencing skills and mother's behavior. The project was configured in a way that only one behavior from a particular behavior class could be active at any given point of time. The screen had two columns pertaining to subject and action which ensured subject wise recording of behavior in a continuous fashion against time. The software allowed simultaneous recording of different behaviors from different behavior classes. Momentary frequencies were recorded as events and those with duration were recorded as states.

The assessments were done as direct continuous structured behavioral observations for 11 minutes in a playroom clinic setting. Child and mother were placed on a 16 squares of equal size floor mat, with toys. The mother was instructed to help the child only when the child asks for it. The session started with a free play to orient the child with the setting. Familiarization toys for first 5 minutes and 3 different stimulus toys for 2 minutes each were given to the child for exploration. Four Familiarization toys (whistling toad, 4 red cubes, stacking rings, toy car) and three stimulus toys (Piggy tailed elephant, bunny on a slide, helicopter) ambiguous to the child were selected for the social referencing observation. An examiner observed and coded the social referencing observations sitting in the same room. The initial behavioral states of the child were entered at the beginning and behaviors were recorded using appropriate key codes.The data files in the form of ODF were automatically stored in The Observer® and can be used for analysis. An editing program was available which allowed the examiner to note a mistake during the observation session and view and modify that particular code after the session. The details of procedure of social referencing assessment in this study have been discussed elsewhere.

2.2 Developmental Assessments

Global assessments of cognitive and motor development were conducted using Bayley Scales of Infant Development- II [4]. The Mental Scale is designed to assess sensory perceptual acuities, discriminative abilities, habituation, language and pre writing skills which are central to children's cognitive development. The Motor Scale is designed to assess muscle tone, dynamic and static balance and perceptual- motor development. The Mental Developmental Index and Psychomotor Development Index were calculated using standard procedures.

Reliability evaluations were conducted at the start and regular intervals during the study to achieve and maintain an inter-observer reliability of more than 80%.

2.4 Statistical Analyses

Preliminary analysis for calculating the frequency and duration of social referencing behaviors was done in The Observer® 5.0 software. The analyses files were exported from The Observer® 5.0 into SPSS 12.0 for further analyses. To examine whether social referencing skills were associated with development, we conducted multiple linear regression analysis using PDI score and MDI score as dependent variables and five social referencing skills of the child as independent variables in two separate models. To examine the effect of mother and child affect during the observation, we repeated the above models adding these as covariates.

3. Results

Among the social referencing skills, the child verbalizing with mother was significantly associated with MDI (β= 0.69, p < 0.001) as well as PDI (β = 0.66, p< 0.001) (Table 3.1 and 3.2). This association of child verbalizing mother was independently significant with MDI even when covariates (durations of child smile, mother talking, no contact with the toy and proximity to the mother) were entered in a multiple linear regression model (β= 0.56, p < 0.001). Similarly, for PDI, child verbalizing mother was independently associated even when covariates (durations of mom talk, wariness, no contact with toy, smile, and proximity to mother), had a significant association with the PDI outcome (β= 0.49, p= 0.003).

The association of child verbalizations with mother was also found to be significant with MDI (β= 0.70, p < 0.001) and PDI (β= 0.75, p < 0.001) in a separate simple linear regression model.

4. Discussion and Conclusion

The results suggest that among the social referencing skills, child looking and verbalizing with mother had an independent association with the over all developmental scores. This finding concurs with the fact that child's emotional expressions and social signals are meaningful indicators of the development and help to use contextual information to discriminate and understand an event [1]. To our knowledge no study has evaluated the association of social referencing with mental and motor development but studies have shown that emotional signals are intrinsically connected with child's social functioning and social development [2]. Studies on social referencing have indicated that in learning to discriminate and understand expressions, infants and young children rely on contextual cues which broadly include familiar persons, multimodal and amodal information. Independent significant association of MDI and PDI scores was observed with covariates like mother's response, mother and child affect during the observation session. This finding supports the theoretical suggestions that maternal response to child's signals facilitates the child's rate of habituation and consequently leads to child's cognitive development [5].

Table 3.1. *Association of social referencing skills with MDI scores. Dependent Variable: Mental Development Index (MDI)*

Independent variables	P value	B coefficient
Social Referencing Skills[a]	< 0.001	0.61
Verbal Mother	< 0.001	0.69
Bid Mother	0.39	0.41
Offer Mother	0.46	-.10
Look Mother	0.47	-.05
Show Mother	0.54	-0.17

Table 3.2. *Association of social referencing skills with PDI scores. Dependent Variable: Psychomotor Development Index (PDI)*

Independent variables	P value	B coefficient
Social Referencing Skills[a]	< 0.001	0.66
Verbal Mother	< 0.001	0.66
Bid Mother	0.41	0.44
Offer Mother	0.48	0.11
Look Mother	0.82	-.01
Show Mother	0.45	0.23

[a] *Predictors : (Constant): SR Verbal Mother, Bid Mother, Offer Mother, Look Mother, Show Mother, Verbal Mother*

[b] *Predictors : (Constant): SR Verbal Mother (for Table 3.1; 3.2)*

Our study provides an empirical evidence for association of specific social referencing skills with child development.

The study was supported by New Zealand Milk Limited. Private Bag No 12032. Auckland, New Zealand.

References:

1. Walker-Andrews, A.S. (1998). Emotions and social development: Infants' recognition of emotions in others. *Pediatrics*, **102**, 5 Suppl E, 1268- 1271.
2. Emde, R.N. (1998). Early emotional development: new modes of thinking for research and intervention. *Pediatrics*, **102**, 5 Suppl E, 1236- 1243.
3. The Observer®, Version 5 (2003). Professional System for Collection, Analysis, Presentation and Management of Observational Data, Noldus Information Technology, Wageningen, The Netherlands.
4. Bayley, N. (1993). *Manual for Bayley Scales of Infant Development* (2nd Ed). San Antonio, TX: Psychological Corporation.
5. Walden, T.A.; Knieps, L.J.; Baxter, A. (1991). Contingent provision of social referential information by parents of children with and without developmental delays. *Am. J. Ment. Retard*., **96**, 2. 177-183.

Engaged students and increased performance

O. Courteille [1] , K. Karlgren [2]
[1] *LIME, Karolinska Institute, Stockholm, Sweden*
[2] *DSV, KTH, Stockholm, Sweden*

Abstract

This study is based on a skills test based clinical assessment (OSCE) where 120 fourth year medical students, from the 4 major university hospitals in Stockholm, participated simultaneously in the same examination. The goal of the examination was to assess their skills and ability to solve clinical cases -including patient history taking- within 10mn. We observed and analysed students' reactions, engagement and performance during their interactive sessions with both a standardized patient and a computer-based simulated patient.

Keywords

Simulation, virtual patient, assessment, empathy, emotional engagement.

1 Introduction

1.1 Interactive Simulation of Patient (ISP)

ISP is a practice-based learning tool where learners can explore and solve clinical cases (1). The system tracks student's progress while they work with the computer-based simulated patient. They can start by taking the patient's history, selecting from a menu. They then can choose to:

- Perform a physical examination (using a mouse to click on an interactive patient image),
- Request laboratory or radiological tests and/for
- Diagnose the patient

To resemble the hospital rounds and to give a sense of reality, students are given a reasonable amount of freedom to perform the task, according to their intuition as they would in a real situation. ISP offers the ability to select simulated patients who differ in age, ethnicity, gender, and personality. The simulated patient consists of pre-recorded video sequences of a patient/actor. In that way the ISP feels like a realistic learning application wherein the practitioner must learn to handle each patient individually.

1.2 Standardized Patients (SP's)

SP's are real actors who play the role of patients or actual patients coached to present specific illnesses to the medical practitioner. They are performing live. Their purpose is to teach and evaluate patient assessment and interviewing skills. There are limits to how effective SP's can be for training. Given such concerns as actor training and availability, reproducibility, changing evaluation criteria and implementation cost we begun development of simulated patients as an alternative to hiring actors.

Both SP and ISP aim to help students to practice clinical reasoning skills. We stress that, although the scenarios are pre-defined, the interaction itself is unscripted. The scenario establishes initial conditions but the student's responses to the standardized or simulated patient, as well as inherent flexibility in how the SP or ISP is allowed to react, cause the conversational flow to vary from interview to interview.

2 Method

This study is based on an Objective Structured Clinical Examination (OSCE) that took place in May 2004 for all of the fourth year Swedish Medical students who belonged to the Stockholm County in Sweden. About 120 students, from the 4 major university hospitals in Stockholm, participated simultaneously in the OSCE sessions. The goal was to assess their skills and ability to perform basic surgical procedures in a very limited amount of time (5/10mn at each station). The time pressure affected their performance on different levels (decision making, visual coordination, operational accuracy, etc).

Four stations (out of sixteen) consisted of two ISP's and two SP's. The common goal for the participants at these four stations was to interview a patient and then make a diagnosis within max 10 minutes.
In addition, none of them had tried the ISP before so we offered them an assistant, sitting beside, and whose task was to introduce the system, help them to navigate and, last but not the least, handle the keyboard for asking questions to the simulated patient. It turned out that this very special and resource-consuming environment gave us a rather convenient method to monitor their respective clinical reasoning processes in a natural way. The presence of the assistant made them to think aloud spontaneously and to verbalize what they were doing and why they were doing it while they were working with the ISP. Most of the students thought this point helped them to feel calmer and avoid feeling uncomfortable or frightened.
They were observed while they were working with the simulated and standardized patients (about half of the students were video-filmed). Notes were taken on their general actions executed and on their impressions/reactions while working with different parts of the ISP. Additional data (log files from the ISP sessions) was automatically created and collected during the computer interaction.

The aims of this study were:

- To compare the flow experiences and the realism of this life-like situation between the simulated patient and the standardized patient.
- To observe the affective impact of the simulated patient on the student's confidence and ability to solve clinical problems. A cross-comparison between the 4 different groups (hospitals) has been done.
- To collect authentic emotions evoked on students for building a database.

2.1 Collected data

An evaluation and usability questionnaire was handed to the participants after the ISP sessions and they were asked to fill it out and send it back to our department. Although this data collection was on a voluntary basis we obtained a 65% answer frequency. The questions concerned both SP and ISP and were, among others, focused on student's perception and engagement as well as stimulation to solve problems and pertinence of the ISP and SP stations for examination.

2.2 Video-analyses

A coding scheme has been applied for observing and assessing the different behaviours of students interacting in front of the computer (speech, posture, gesture, gaze, think aloud, etc). The interactions student/patient and student/assistant have been analyzed by means of categorization of the task-solving components (*e.g*. action, argumentation, management, explanation, out-of task interaction, task-oriented interaction, etc). Student's facial expressions have been coded as well as voice intonation and velocity (stress factor). Information gathered in the log files has provided us additional data about performance, interaction time, confidence scale, motivation and suggested diagnosis.

3 Discussion

The originality of this experiment -a unique pilot study so far- has provided us interesting data about the correlation between stress and the power of video-based simulated patients to affect and move students (empathy).

Validity: One key issue in the observation is that the performance of the standardized patient is not regular but varies (human factors). The actor doesn't react and behave the same way and the same manner at each session. The simulated patient however offers a more reliable measurement tool in terms of provoking reactions and conveying emotions.

Preliminary results show that the time pressure due to the examination settings affected the students' performance on different levels (decision making, visual coordination, operational accuracy, etc).

The video-analyses have been performed by means of the Observer tool (from Noldus) and validated by the Cohen's kappa (by means of 2 observers). The findings from the video observations will be presented and discussed during the poster presentation.

Note: Unfortunately the logistic around these observations didn't allow us to use magnetic resonance imaging (MRI) to study how the Med students brains responded to the simulated patient and what regions of the brain were stimulated (reward centers). We will try to design and conduct another controlled study for these purposes.

References:

1. Bergin, R.; Youngblood, P.; Ayers, M.; Boberg, J.; Bolander, K.; Courteille, O.; Dev, P.; Hindbeck, H.; Stringer, J.; Thalme, A.; Fors, U.G.H. (2003). Interactive Simulated Patient: Experiences with Collaborative e-Learning in Medicine. *Journal of Educational Computing Research.*

Puberty and sexual behavior in Caspian miniature colts

S. Dordari[1], M. Rezaiee1, F. Amini[2]

[1]*Agricultural Research Center, Tehran, Iran*
[2] *Institute of Animal Science, Karadj, Iran*

Abstract

The Caspian miniature horse is an ancient breed of horse, which was thought to be distinct for many years. This breed is a charismatic numbers, has now been recognized as holding an ancestral position in the history of modern horse breeds, probably preceding that of the Arab horse. On the other hand, the Caspian horse is the only breed with odd chromosome numbers that means it may be a mixture of *Equus przewalski* and *Equus caballus*.

The objectives of this study were to (1) Determine age at puberty; (2) Characterize seminal characteristics and sexual behavior.

Puberty was studied using 7 colts of the Caspian miniature horse phenotype. The evaluation was performed during two years on a completely healthy group of colts. Data for sexual behavior represented all 7 colts at puberty. Seminal collection was attempted every 2 weeks from 50 to 140 weeks. For all collections, times to erection, mount and ejaculation and seminal characteristics were recorded. Age at puberty was defined as the first ejaculate containing 50 million spermatozoa with >10% motile.

The age at puberty was on average 24.6 months (12-34 months). Sexual behavior and seminal characteristics differed (p>0.05) between puberty and 2 years, except for time to erection, time to mount and percentage of motile spermatozoa.

In this study, the differences and similarities between this breed and the ponies were determined. Puberty, as determined by seminal characteristics, occurred between 11.5 and 14.5 months in Welsh pony stallions and between 12 and 23.6 months in quarter horse colts. As other researches show, puberty in horses varies a lot with breed. It can occur from 8-24 month and in pony breeds it is shorter than other horse breeds. The average age at puberty of *E. przewalski* is 31 months and of *E. caballus* 12 months.

The Caspian miniature horse was called Caspian pony for many years, but shows more physiological similarities to horse breeds than ponies and to *E. przewalski* than *E. caballus.*

Keywords

Caspian horse, puberty, stallion, sexual behavior, colt.

1 Introduction

There has been no systematic study of puberty in the testicular develop and age at puberty. Classic data on testicular development in horses were interpreted (Nishikawa 1985) as evidence that puberty occurs between 17 and >22 months in Anglo-Norman stallions, and Glass neck (1987) noted a great difference in testicular development among warm blood colts of similar age, but made no conclusions regarding age at puberty.

Puberty, as determined by seminal characteristics, occurred between 11.5 and 14.5 months in Welsh pony stallions (skinner & Bowen, 1986) and between 12 and 23.6 months in Quarter Horse colts (Cornwell 1972). Knowledge of the normal progression of events before and after puberty is necessary if further research is to be conducted on puberty in the horse.

On the other hand, the Caspian miniature colt is an ancient breed of horse (not pony but pony size) which was thought to be extinct for many years. This breed has now been recognized as holding an ancestral position in the history of modern horse breeds specially of the Arab horse and as a primitive position in relation to other Oriental and Iberian horse breeds The objectives of this study were to (1) Characterize seminal characteristics and sexual behavior; (2) Determine age at puberty.

2 Materials and Methods

2.1 General

Seven pure breed registered Caspian miniature colts born between March and August 2003 were used. Colts were maintained in boxes with their dams and had access to a limited amount of fresh hay until weaned at 16-24 weeks of age. Thereafter, colts received alfa alfa hay and a measured daily ration of grain, which was formatted to meet 100% NRC requirements for weanling and growing horse. Colts were together in an outdoor paddock until approximately 65 weeks, after which they were housed in individual pens till 3 years of age.

2.2 Seminal characteristics

Puberty was defined as the age when the first ejaculate was obtained containing a minimum of 50 106 spermatozoa with 10% progressively motile (Wolf et al. 1965). Seminal collections by artificial vagina (picket et al. 1987) were attempted every 2 weeks from 50 to 140 weeks. Once puberty was achieved, 4 more samples were collected to ensure that the criteria were met consistently. Volumes of gel and gel free semen, concentration of spermatozoa /ml gel free semen, and PH were measured (Pickett et al. 1987) and total spermatozoa per ejaculate was calculated by using aspectro photometer. Percentage of progressive motile spermatozoa at 37^{o}C was determined at x200 using a microscope.

2.3 Sexual behavior

For each attempted seminal collection, colts were exposed to an estrous mare of appropriate size for a maximum of 20 minutes. Sexual behavior was assessed (clay et al. 1987) by timing the intervals to (1) Erection; (2) First mount with erection and intromission; (3) Ejaculation. Number of mounts per ejaculate was also recorded.

2.4 Statistical analyses

One way analyses of variance were used to determine age-associated differences in testicular size, sexual behavior and seminal characteristics, and to analyze potential differences in sexual behavior and seminal characteristics within and between ejaculates at 2 years of age by using Ouncan's multiple range test.

3 Results and Discussion

3.1 Age at puberty

The average age at puberty was 24.6 months and ranged 12-34 months. Age at puberty was not similar to that reported by Cornwell (1972) for spring born quarter Horse colts (18 months, range 12-23.6 months) and greater than values reported by skinner & Bowen (1986) for welsh ponies (about 11.5 to 14.5 months). One of the colts was "shy" and reluctant to serve the artificial vagina. It was suspected that he had sufficient testicular activity to produce a pubertal ejaculate before he actually did. Exclusion of this colt gave an age at puberty of 20 months. The colt was excluded from all calculations involving intervals after puberty.

3.2 Sexual behavior

Data for sexual behavior represented all 6 colts at puberty but only 3 at 2 years. Data for 3 stallions at 2 years were excluded because they had poor libido and required special management to permit seminal collection. There was no significant difference between the times before first erection and first mount at puberty and 3 years of age. However, time to ejaculation and total numbers of mounts required for ejaculation were less at 3 years of age. Presumably, the decrease in ejaculation time and number of mounts at 3 years was due to a combination of age and experience. Total time to ejaculation and total number of months per ejaculation at 3 years were similar to those for adult stallions.

3.3 Seminal characteristics

Seminal characteristics fort the pubertal ejaculate and the average for the last 10 ejaculates at 1-3 years differed ($P<0.05$). The average pubertal ejaculate contained only 2   109 spermatozoa, but 27% were progressive motile.

Studies in Quarter horses showed that the percentage of motile spermatozoa increased 13 weeks after puberty (Nauden and Amann, 1990), after which normal percentages of motile spermatozoa were attained. For Caspian horse stallions this average was 15 weeks after puberty. It is possible that these stallions require a longer post-pubertal interval for attainment of 50% progressively motile spermatozoa than Quarter horses. It could also be that this group of stallions had reached their adult quality but the subjective observations were unusually harsh.

As expected, seminal volumes and sperm numbers at 3 years were much lower than values reported for adult stallions (Gebaur *et al*., 1984). Nevertheless, total spermatozoa/ejaculate increased between puberty and 3 years old.

Mean DSO was $2x10^9$ (range 0.70 –$2.21x10^9$), which is much less than the $7.0x10^9$ reported for adult stallions (Gebauer *et al*., 1984). Total spermatozoa per ejaculate differed ($P<0.05$) among stallions, but variance associated with ejaculates within stallions was not significant ($P>0.05$). The mean pH per ejaculate differed, the mean was 7.79. The mean weight of stallions at the age of puberty was 165.32 kg. In general, the average age of puberty of *Equus przewalski* is 31 months and in *Equus caballus* 12 months. Puberty as determined by seminal characteristics has been reported between 11.5 to 14.5 months in welsh pony stallions and between 12 to 13.6 months in quarter horse stallions. As the other researches show, puberty in horses varies a lot with breed and type. It can occur from 8-24 month and in pony breeds it is shorter than other horse breeds.

Although the Caspian Miniature horse was called as Caspian pony for many years because its body shape and size was similar to pony size, it shows more physiological similarities to horse breeds than ponies and to *E. przewalski*: than *E. caballus.*

References

1. Amann, R.P. (1981) Acritical review of methods for evaluation of spermatogenesis from seminal characteristics. *J. Androl***. 2**, 37-58.
2. Berndtson, W.E.; Squires, E.L. (1983) Spermatogenesis, testicular composition and the concentration of testosterone in the equine testis as influenced by season. *Theriogenology*, **20**, 449-457.
3. Clermont Y. (1982) Kinetics of spermatogenesis in mammals. *Physiol*. *Rev*., **152**, 198-204.
4. Cornwell, J.C. (1972) Seasonal variation in stallion semen and puberty in the Quarter Horse colt. *J. Anim. Sci*., **53**, 1645-1657.
5. Kosinik, K. (1985) Characteristics of successive jets of ejaculated semen of stallion. *J. Reprod. Fert*., Suppl. **23**, 59.
6. MacDonnel, S.M.; Kenney, R.M. (1987) Sexual behavior in stallion. *J. Reprod. Fert*., Suppl. **23**, 59.
7. Skinner, J.D.; Owen, J. (1986). Puberty in the welsh stallion. *J. Reprod. Fert*., **86**, 133-135.
8. Thompson, D.L.; Pickett, B.W. (1989) Testicular measurements and reproductive characteristics in stallions. *J. Repro. Fert*., Suppl. **47**, B-17.

Bedding-in and Rooming-in on the Post-natal Ward: Breastfeeding Initiation

E. Heslop, S.J. Leech and H.L. Ball
Parent-Infant Sleep Lab, Department of Anthropology, University of Durham, Durham, UK.

Abstract

This poster presents data regarding the hypothesis tested, data collected and results obtained regarding 30 mother-infant dyads who participated in a randomized control trial exploring the effects of mother-infant bedding-in (with a standard cot-side or clip-on-crib) compared to rooming-in (with a standalone bassinet) on breastfeeding initiation during the first post-natal night on a hospital maternity ward.

The behaviors and interaction of the mothers and their infants were captured to videotape using low-light intensity cameras illuminated via infra-red light. Observation periods last for the entire night and video coding was therefore complex and prolonged. Behaviors were coded as events or states using an established multi-level behavioral taxonomy via The Observer software.

Results indicate that, compared to rooming-in, infants who spent the majority of the night bedding-in had a greater breastfeeding to attempting to breastfeed ratio and fed for longer durations on the first postnatal night whilst in a hospital maternity ward.

Keywords

Breastfeeding, Initiation, Mother, Infant, Behavior.

1 Introduction

Breastfeeding and skin to skin contact between mother and infant have been shown to have considerable short and long term health benefits [9]. The UK Government set National Health Service (NHS) Trusts a target of increasing breastfeeding initiation rates by 2% per year for 3 years [6]. Similarly, the United Nations International Children's Fund (UNICEF) Baby Friendly Initiative (BFI) is encouraging hospitals to make a commitment to 'Baby Friendly' practices by encouraging the early initiation and continued promotion of breastfeeding. In response, many UK hospitals are developing policies on mother-infant bed-sharing on the post-natal ward [7].

The link between bed-sharing and breastfeeding has clear theoretical grounding in a biological-anthropology paradigm. In considering breastfeeding and mother-infant sleep from an evolutionary perspective it is clear that infants have evolved to breastfeed frequently and to be kept in close contact with their mothers both during both day and night for at least the first year of life [4]. This has been the case for over 4.5 million years of human evolution and it is only within the last 150 years that cultural practices in industrialized Western countries have seen the introduction of solitary sleeping in infancy [3]. This trend seems to have been supported to a large degree by the development of 'formula' infant supplement in the 1860's. Breastmilk is the most appropriate nutritional source for infants as it provides all the nutrition they require for the first 6 months of life [9]. However, as breastmilk is easily digested by infants it leads to breastfed infants requiring feeding more frequently and consequently having shorter sleep bouts than formula fed infants, increasing the amount of night-time parental effort required

Bed-sharing would therefore seem an evolutionarily appropriate night time care giving behavior as it allows constant close mother infant contact to facilitate breastfeeding, whilst ameliorating maternal care giving effort. Indeed, to become 'Baby-Friendly' accredited hospitals are required to implement UNICEF's 10-steps to successful breastfeeding which state that mothers should be allowed uninterrupted skin-to-skin contact for at least half an hour following delivery, midwifery staff should encourage breastfeeding within the first hour, and advise mothers to keep their baby close to them at all times [5]. In this way mother-infant bed-sharing on the post-natal ward is encouraged, as a way of facilitating the establishment of both breastfeeding [8] and bonding [2], while helping mothers gain more post-partum rest (as they can comfort, feed, and care for their babies in bed) [1].

To date, however, little data exists to directly support the hypothesized relationship between bed-sharing and breastfeeding in the immediate post-partum period, and the effects on breastfeeding behavior and success between mother infant dyads bedding-in compared to the rooming-in on the first post-natal night in a hospital maternity ward is unknown.

This study uses behavioral observation to explore the effects on breastfeeding initiation, between 30 mother infant dyads bedding-in compared to rooming-in on the first post-natal night in a hospital maternity ward.

2 Project aims

This research aims to investigate whether sleeping arrangement on the first postnatal night affects breastfeeding behavior.

3 Objectives

To observe, document and quantify the nature and extent of mother and infant night-time feeding interactions on the first night of life on the postnatal ward in three randomly allocated sleeping conditions.

4 Hypothesis to be tested

Infants sleeping in closer proximity to the mother will show:

- greater breastfeeding: attempting to breastfeed ratio
- greater total duration of breastfeeding
- more breastfeeding lying down

than infants sleeping further away.

5 Methods

5.1 Recruitment

Volunteers were recruited pre-natally at bi-weekly 'Bloomsbury Breastfeeding Workshops'. They were invited to opt-in to the study by completing and returning the enrolment and consent forms.

Mothers who volunteered and met the primary inclusion criteria were randomly allocated to one of the three sleeping conditions via sequential sealed envelopes (bedding-in with a cot side (BI), bedding-in with a clip-on-crib (CO) or rooming-in with a standalone cot (SA)), provided with a study label to place on their hand held maternity records. An identical label was sent to maternity record at the hospital, which indicated to midwives that the mother was a study participant and that they were required to telephone a specially designated 24 hour voicemail number to inform researchers that the mother had delivered.

5.2 Inclusion criteria

Ante-natal (primary) inclusion criteria for mothers were that they were non-smokers with an intention to breastfeed. Post-natal (secondary) inclusion criteria for mothers were that they had received no opiate analgesics either IV (intra-veinously) or IM (intra-muscularly) during labor. Post-natal inclusion criteria for infants were that they were born vaginally and in good health at greater than 37 weeks gestation.

5.3 Equipment set up, overnight recording and equipment take down.

Following delivery confirmation research staff firstly allocated the mother-infant pair a single room and gave them the correct equipment for their randomly allocated sleeping condition. Secondly, a set of custom designed video equipment was erected at the foot of the mother's hospital bed to film the mother-infant dyads first post-natal night. A zero-lux camera, infra-red light source, long play 12-hour record capability VCR and a Time Code Generator allowed continuous data capture onto video-tape for the entire night with time signal overlay. The mother was told that she may stop the recording at anytime by use of the remote control and that if left running the tape would stop itself after 8. Tapes were collected after the second post-natal night.

6 Data Analysis

Data on 30 mothers selected from a larger RCT sample are presented here. Behaviors were coded as events or states using an established multi-level behavioral taxonomy using The Observer 5 software (Noldus IT). Analyses involved generating output of event frequencies and state durations per dyad (e.g. frequency of feeding attempts, duration of feeding bouts).

7 Results

Results indicate that infants assigned to the bedding-in condition did not have greater breastfeeding to attempting to breastfeed ratio nor did they have greater total duration of breastfeeding or more breastfeeding lying down than infants assigned to the clip on crib or standalone cot conditions.

However, when the breastfeeding data was analyzed via actual condition (condition that infant spent the majority of the night in over the whole observational period) results show that bedding-in infants and clip-on-crib infants did have a greater breastfeeding to attempting to breastfeed ratio than standalone cot infants (2:1 vs. 4:1). Furthermore, bedding-in infants were observed to have greater total durations of breastfeeding and a greater percentage of their breastfeeds lying down than infants who spent the majority of the night in either the clip-on-crib or standalone cot.

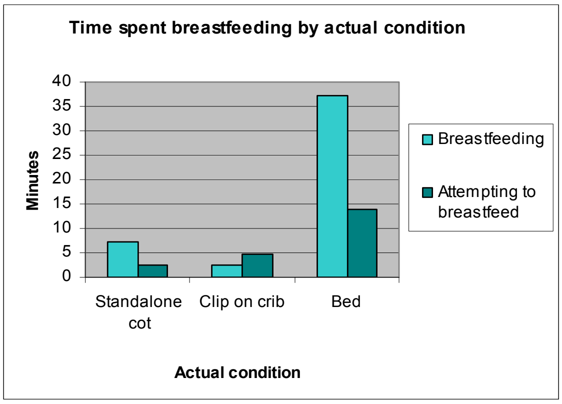


Figure 1*: Time spent breastfeeding over the whole observational period- actual condition.*

8 Conclusion

Compared to infants rooming-in, infants who spent the majority of the night bedding-in had a greater breastfeeding to attempting to breastfeed ratio and fed for longer durations on the first postnatal night whilst in a hospital maternity ward.

Supported by Babes in Arms Research Grant.

References:

1. Ashmore, S. (1997). Achieving baby friendly status in a large city hospital. *Modern Midwife*, **7**(6), 15-19.

2. De Chateau, P. and B. Wiberg (1977). Long-term effect on mother-infant behaviors of extra contact during the first hour post partum. II. A follow up at three months. *Acta Paediatrica Scandinavia*, **66**(2), 145-151.

3. Hardyment, C. (1983). *Dream Babies: Child care from Locke to Spoc*k. London, Jonathan Cape Ltd.

4. McKenna, J. J. (1990a). Evolution and Sudden Infant Death Syndrome (SIDS): Part II: Why human infants? *Human Nature*, **1**(2), 179-206.

5. UNICEF Baby Friendly Initiative. (2000). *Implementing the Baby Friendly best practice standards*. London, UNICEF UK Baby Friendly Initiative.

6. UNICEF Baby Friendly Initiative Website. (2004) www.**babyfriendly**.org.uk/

7. Ward-Platt, M. and H. L. Ball (2002). Assessing the practical considerations at the hospital. *Mothering* Sept-Oct 2002, 58-61.

8. Widstrom, A. M., V. Wahlberg, et al. (1990). Short term effects of early suckling and touch of the nipple on maternal behaviour. *Early Human Development*, **21,** 153-163.

9. World Health Organisation (WHO) (2001). The optimal duration of exclusive breastfeeding: Results of a WHO systematic review. *Note for the Press N°72 April 2001*.

Using ethnographic methods to carry out human factors research in software engineering

J.S. Karn, A.J. Cowling
Department of Computer Science, University of Sheffield, Sheffield, UK, S1 4DP

Abstract

This paper describes how ethnographic methods were used to observe and analyze student teams working on software engineering (SE) projects. The aim of this research was to uncover the effect of the interplay of different personality types as measured by a test based on the Myers-Briggs Type Indicator (MBTI) on the workings of an SE team. Using ethnographic methods allowed the researchers to record the effect of personality type on behavior towards team mates and how this related to the amount of disruption, and positive ideas brought forward from each member, also examined in detail were issues that were either dogged by disruption or that did not have sufficient discussion devoted to them and the impact they had on the outcomes of the project. Initial findings indicate that ethnographic methods are a valuable weapon to have in ones' arsenal when carrying out research into human factors of SE.

Keywords

MBTI, Empirical Software Engineering, Group Work in Software Engineering, Ethnography

1 Introduction

The goal of this paper is to describe how ethnographic methods were used to study software engineering (SE) teams for the duration of specific projects, in order to gain a greater understanding of the role of human factors over the duration of an industrial SE project. This work also had another dimension, that was to study how particular combinations of individual personalities as measured by an online version of the Myers-Briggs Type Indicator (MBTI) might either promote effective teamwork or cause clashes, and secondly how these clashes disrupt the operations of an SE team.

The importance of team work in SE has been recognized at a pedagogic level. Many undergraduate courses now have group projects as a central part of their syllabus. These projects can vary from carrying out the theoretical analysis and design of a system through to the analysis, design and implementation of a system for an external industrial client.

A computer based system is built for people and by people, it is intended to serve some human purpose and it exists in some human context. Some of the most vexing difficulties involved in SE projects are social, political or cultural, and social issues are inherent to the requirements engineering process. The work carried out by the current authors is not limited to understanding requirements engineering, but aims to encompass the entire SE project lifecycle.

So far, the authors have studied the operation of teams throughout complete projects, measuring the levels of disruption that different issues have caused to the team, and have shown how these disruptions can be related to personality clashes between individuals within the teams [1], [2]. Follow up work [3], has examined the effects of such disruptions, and in particular how they affected both the cohesiveness and the quality of the work produced.

The remainder of this paper will describe the methodology used to ascertain the aforementioned results, with a focus on explaining the evolution of the research strategy.

2 Personality Typology Used

The personality theory underlying this research is provided by Jung's work on psychological types [4], as developed by Myers-Briggs in designing the MBTI (Myers-Briggs Type Indicator) [5]. The test used in this research is not the official paper version of the MBTI, but an online test based on the MBTI developed by Human Metrics, a consortium of Israeli psychologists, who claim they have found no significant statistical differences between this test and the official version of the MBTI.

Another popular personality test is the NEO-PI model which is based on the five factor model of trait personality [6]. The NEO-PI is a paper based model and no online version was found, therefore persuading students to take the test and return the result would have been more problematic and time consuming, given that the work took place according to a strict project timetable, students would have been less likely to volunteer to take part in the research.

3 Research Methodology

Ethnographic methods were employed to observe selected student teams. In order to gain an in-depth understanding of the user culture (in this context SE teams), the authors observed, and inquired about the research subjects' normal activities throughout their specific projects.

There were several phases involved in going from the field notes stage into quantifiable data, this 'theoretical sampling' is described in detail in [1]. A pilot study was conducted over the course of a semester, during this study field notes were taken during each team meeting. All contributions from each member were noted as was any interaction with the client or manager. After giving the field notes some kind of order the next step was to classify aspects of the ordered details, such as identifying the kind of issue to which an idea, question or problem was related. As many meetings were attended a large amount of data was gathered over the course of the pilot study, during the projects the focus was on gathering as opposed to analyzing data, at this stage the authors had no clear ideas about how to classify the bewildering conglomeration of issues that were being discussed during the meetings. Analysis of the field notes on completion of the pilot study revealed several prominent themes; these themes became the headings for a nominal scale that was created to classify issues:

- PSI- Project Specific Issue
- MSI- Methodology Specific Issue
- GSEI- General Software Engineering Issue
- THF- Team Human Factors
- C- Client
- M- Manager

Another component of this field work was to observe the team formation process through the four phases of forming, storming, norming and performing. The stages affected the observation process, particularly during the pilot study, as the initial formation stages were largely ignored; more focus was placed on project specific issues as opposed to looking at the fascinating interplay of different personalities in the initial forming stage. This wasn't the case in subsequent projects. Initial forming stages tend to be dominated by THF issues as team members discuss their background, the kind of work they prefer, strengths and weaknesses etc, roles are also assigned during this period. In later stages of the formation project THF's are not as prominent as the team are concentrating more on completing project work as opposed to discussing each others background and work plans.

The process of analyzing the field notes and classifying data under the above headings was a long arduous process; but the end result was that the authors gained a rich insight into social relations, events and processes within SE teams. The derivation of ordinal scales to measure the impact of issues and the levels of disruption that they caused arose from the analysis of the fieldwork. Each issue was traced from its inception through to its consummation, and this involved analyzing the relevant field notes in order to classify each specific issue. This was in contrast to the nominal scale which was utilized whilst the fieldwork was taking place, when by following a particular issue it was reasonably straight-forward to classify it under the relevant heading. The nominal scale developed during the study and the ordinal scales described in [3], add to the reliability of the research carried out. Essentially they allow any researcher to carry out an analysis of the field notes and then to make an objective measurement of each issue.

Identifying and measuring issues and recording relevant information was a time consuming and complex exercise. In later projects it became easier to record information and to identify specific issues during the analysis of data. All of the displayed behaviour from the research subjects was viewed through the perspective of their personality type; therefore data recorded in the field notes had to be backed up by congruous information relating to the personality type of the relevant individual(s). This was another convoluted exercise as the actions of one person with a particular personality type had to be balanced against their team-mates personality type, this helped the authors to identify personality clashes and how they disrupted the workings of the team. The final aspect of this process was that of anonymizing results, and this was done in two stages. Each person was given a randomly generated number within each team, such as A1 and B2 to ensure that complete anonymity was achieved.

In classical ethnography data collection and analysis and the decision as to when to withdraw from fieldwork take place in what is referred to as 'theoretical sampling' and 'theoretical saturation', the latter referring to the time when observations no longer serve to question or modify the theories generated from earlier observations, thus rendering the theory saturated with data. For the work described in this paper, the 'theoretical saturation' was taken out of the researchers hands, as it took place when the student projects finished. This did not have an adverse effect on either the quality or the quantity of collected data.

4 Conclusions

Ethnography was employed successfully as part of a larger project looking into the effects of psychological type on SE teams. The use of this method allowed the authors to gain an insiders account into the interactions going on within the student teams. Issue could be grouped under one of the headings from the nominal scale as the fieldwork was being carried out; further analysis on completion of the fieldwork allowed one to measure each issue on an ordinal scale so as to indicate whether a specific issue upset the equilibrium or passed by without consequence.

The actual behavior of the students was measured according to Jungian psychological types; this allowed the authors to ascertain whether given combinations of personalities can work productively over the course of an SE project. It is hoped that this research will provide general pointers to others who seek to carry out ethnographic work in SE.

References:

1. Karn, J.S and Cowling, A.J, "An initial study of the effect of personality on group projects in software engineering", *Dept of Computer Science, Research Report, CS-04-01, University of Sheffield, 2004*
2. Karn, J.S and Cowling, A.J, "An initial observational study into the effects of personality type on software engineering teams", *Proceedings of EASE 2004, pp.155-65*
3. Karn, J.S and Cowling, A.J, "A study into the effect of disruptions on the performance of software engineering teams", *Dept of Computer Science, Research Report, CS-04-17, University of Sheffield, 2004*
4. Jung, C.G, *Psychological Types:* Princeton University Press, 1923
5. Myers, I.B, and Myers, P.B, *Gifts Differing Understanding Personality Type:* Davis Black, 1997
6. Costa, P, and McCrae, R, "Four ways, five factors are basic", *Personality and Individual Differences* ,vol. **13**, pp. 653-665, 1992

Continuous visual 24-hour measurement of behavior differences in Arabian Horses

F. Kuhne[1], K. Scheibe[2], R. Struwe[1]
[1]*Institute of Animal Welfare and Behavior, Veterinary Department, FU Berlin, Germany*
[2]*Institute for Zoo and Wildlife Research, Berlin, Germany*

Abstract

Continuous visual observations were carried out over 24-hours on 36 days of the year on Arabian Horses, which were kept under semi-natural conditions on pasture. The days were equally distributed over the complete year. The frequency of a particular behavior and the average duration per bout were measured for each horse.

The behavioral pattern research under semi-natural conditions means, for this study, that the environmental situations and conditions were not controlled or manipulated. Data collected under those conditions were judgeable but incapable of proof. The seasonal variations of basic behaviors were described in three-dimensional plots. The Kendall's coefficient of concordance (W) was used to determine the synchronisation of the 24-hour time budget among the horses.

This study represents an attempt to examine whether the horses exhibited significant behavioral synchrony and if there were behavior pattern differences among the horses, whether the measurements might regarding the average total time of a particular behavior rather than the frequency or average duration each.

The horses spent 13-17 h of the 24-hour period feeding and rested 5.5-8.5 h daily. The individual differences could be observed with the help of the mean grazing frequency of 7-11 times a day with an average duration of 17-34 minutes per action. Hay feeding was documented with a frequency of 16- 24 times a day, lasting 21-37 minutes each. A strong social synchronisation as often described in natural family groups was less expressed in this herd.

The observation showed that although some horses showed behavioral patterns in high degree with the average, there were a few horses with a less degree, i.e. the average of selected behavior in a small group of animals might have ignored the problem of individuality in using free available resources and the problem of the locally limited supply of particular resources.

Behavior frequency (meals or resting/day) and duration (min/meal or resting) throughout the year are useful to determine the time spent feeding and resting (min/day and year). These measurements are necessary to analyse social synchrony.

Keywords

Behavioral observations; Behavior pattern; Horses; Circadian rhythm; Seasonal rhythm

1 Introduction

The 24-hour time budget of 10 Arabian horses was investigated over one year. The study of behavior can be determined by the measurement of the frequency (bouts/day), the duration (min/bout) and the effectiveness of the behavior e.g. the meal size for foraging behavior (g/meal) (Nielsen, 1999). The total time spent on a particular behavior throughout the day or year (min/day or year) can be analysed by the frequency and the duration for each bout and can be adjusted to obtain the social synchrony. The behavior rate which reflects social stress among the individuals can clearly be assessed by the duration of behavior and its effectiveness such as the meal size. The parameters meal size and meal frequency are useful to determine the daily/yearly food intake (gram/day or year) and may reflect an individual behavioral priority, behavioral and physiological needs as well as environmental influences.

Differences were found in the behavior of the horses concerned especially the seasonal feeding pattern. A particular goal of this investigation was to suggest the role of the measurements frequency and duration which influence the evidence of synchrony of the behavior among animals.

2 Materials and Methods

2.1 Data acquisition

Visual observations were carried out over 24-hours on 36 two-day periods of the year. The days were equally distributed over the complete year. The 24-hours were divided into two periods of 12-hours on two consecutive days, so that the observations were carried out by the same person.

Total continuous recording was used to measure all specified types of behavior whenever it began or ended (Martin & Bateson, 1993). The frequency of a behavior and the average duration of action were measured for every horse.

The interval recording method was used during hours of darkness. The behavior was observable with a night vision aid every 5 minutes. This short interval was chosen because of the duration and frequency of the main behavioral pattern. A few of the horses, which had nearly similar silhouettes and were sometimes difficult to separate from each other in the darkness, were marked with white identifying bands on different parts of the body.

2.2 Data analysis

This group of horses could not change their behavior independently of each other or from the locally limited supply of particular resources, i.e. hay or water. The measured behavior was not independent from or dependent on one another in an equal degree (Lamprecht, 1999).

The Kendall's coefficient of concordance (W) which provides the simultaneous association (relatedness) between samples was used to determine the synchronisation of the 24-hour time budget among the horses. Kendall's coefficient of concordance suggests values between 0, which means a very low synchronisation and 1, which means a very high synchronisation.

The seasonal variation of basic behaviors were described in three-dimensional plots (x-axis = period length, y-axis = year, z-axis = intensity of significant period of a behavior).

The three-dimensional figure graphically represents the mean behavior pattern over 24-hours in one line, though it should be noted that the behavior duration per bout was shown by a particular color. The green and blue colors were used to show short durations up to 6 minutes and the longer durations up to 15 minutes were shown in yellow, red and purple.

3 Results

A strong social synchronisation as often described in natural family groups was less exhibited in this herd. The synchronisation of the main behavior patterns for horses, the feeding and resting behavior, was observed in a middle degree among this group of animals (Kendall's coefficient of concordance for feeding = 0.419 and for resting = 0.504, both p-values = 0.000).

3.1 Seasonal rhythm

The diurnal and seasonal rhythm of feeding was evaluated with a high variability over the hours of daylight and darkness. Figures 1-4 demonstrates the mean annual foraging pattern for two horses. Grazing was mainly a daytime activity. Hay feeding tended to occur more often at night.

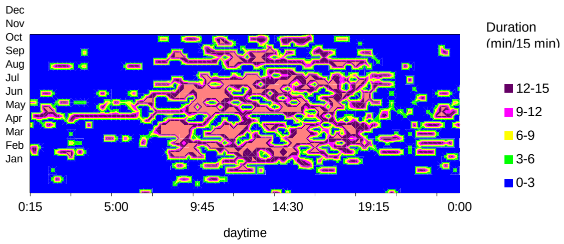


Figure 1. *Horse 8: Grazing throughout the year.*

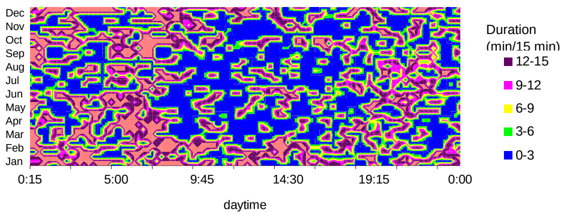


Figure 2. *Horse 8: Hay feeding throughout the year.*

3.2 Individual behavior differences

The regressions of the daily feeding time as an example for the individual differences measured with the usage of the parameters frequency and duration per bout are represented in Figure 5. The majority of the values show a clearly analyzable dependency of the two parameters. A few values and the regression line of three horses occurred with more diversity on both measured parameters. The ambivalent parameter values might be caused by the individual diversity within group-housed social animals. These diversities of parameters seem be helpful to explain the investigating issue to analyse the time spent on a particular behavior in a relevant time for an individual or a group of animals, in this case especially of daily food intake.

4 Discussion

Several behavioral data are assessable with the continuous observation over 24-hours. The time spent on a particular behavior throughout the day or year (min/day or year) analysed by the frequency and the duration for each bout can be adjusted to obtain the social synchrony, but social constraints and individual priorities of using some resources and other aspects should be considered when analysing behavior of social living animals. A few individuals can reach the same daily behavior output with less time but more intensity or higher frequency thereby compensating for a less duration per bout.

The comparison of day- and night time activities showed very clearly that the horses preferred daytime for grazing with a clear seasonal pattern. Especially in the summer months, June and July, long feeding periods nearly up to midnight were measured. The analysis of the annual and daily feeding pattern showed the preference of the horses to feeding around evening twilight rather than dawn. Nevertheless, the horses could also be observed feeding during the night, but with lower frequency and shorter duration for each bout. The diurnal pattern of all foraging behavior patterns throughout the year influenced one another e.g. grazing was preferred as a day time activity, while hay feeding took place at night. The difference of day- and night time activity of each behavior resulting in these diurnal and seasonal variabilities was possible to analyse for all behavior patterns.

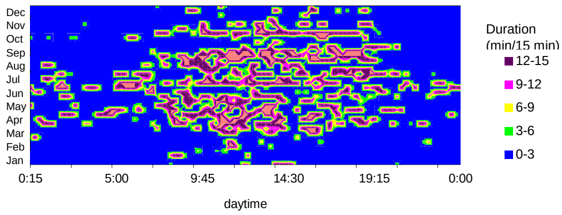


Figure 3. *Horse 10: Grazing throughout the year.*

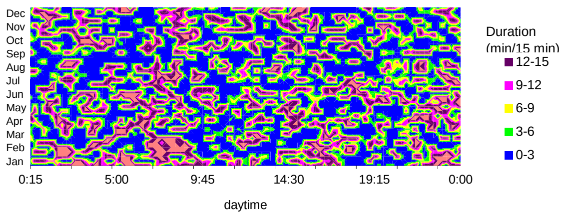


Figure 4. *Horse 10: Hay feeding throughout the year.*

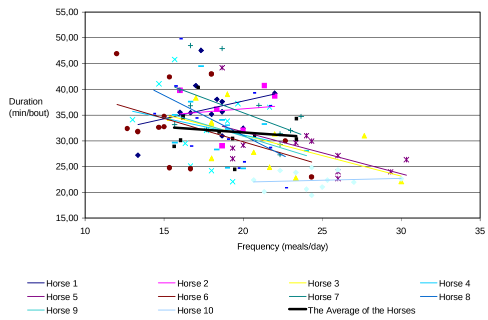


Figure 5. *Daily feeding time: an attempt to suggest the individual differences; the slope of the regression line of daily feeding time is analogous to the mean daily feeding time, but a few horses were observed with a contrary development of daily feeding time.*

4.1 Data acquisition

The results were based on data which were collected with visual behavioral recording. The data sampling and analysis were done from one person that means, for this study, a certain subjectivity is possible.

The visual observation as a method of behavioral data recording is only suitable if the following factors are fulfilled (Sambraus, 1982):

- The animals should be familiar with the person who is observing their behavior that means that every animal has made the experience, that from this person it has nothing to expect in a positive or negative way.
- If the person always stays at the same place during the observation time.
- If the presence of a person is a component of the hypotheses or of a specific question of the study.
- If the behavior parameters are dependent on the presence of a person.
- This method can also be applied for apathetic or less interested in human beings as social partner animals.

After a time of preparation each individual animal should be easily identifiable by simple characteristics.

4.2 Data analysis

The problem of research in the context of real-life situations is the external validity, which deals with interpretations and generalisations from the sample to other conditions or populations (Altmann, 1974). The combination of the two recording methods was chosen as a compromise on the real-life situation of the Arabian Horses. Altmann (1974) has stated that "… internal validity should not be purchased through complete loss of external validity." The mean duration of the main behavior pattern; gazing, hay feeding and resting, was noticed up to 30 minutes. A changing in behavior of the horses especially in the darkness could be observed with the short interval recording of 5 minutes, so that the frequencies of a behavior at day- and night time were comparable with another.

The explorative data analysis of the behavioral pattern has revealed results which have no quantitative reliability. This means for the judgement of the data that the conclusions were not arrived at from significant but from conspicuous results. The main problem of this external validity remains therefore, that the explorative evidences of conspicuous behavioral patterns are not clear at once and for all animals. The generalisation of those results on behavioral patterns should therefore always pay attention to the special conditions of the recording (Lorenz, 1996).

The predominant statistical procedures assume a normal distribution of the data. The data which were measured with the visual recording method are not normally distributed, so that for the statistical analysis nonparametrical methods were used. The disadvantage of the usage of nonparametrical methods is lesser selectivity than in parametrical methods (Lamprecht, 1999). The necessity of a higher number of random samplings was fulfilled in this study with the behavioral recording of all horses. The focal animal sampling as a well established method would not be achieved with the same external validity for this study. The behavioral patterns up to 30 minutes per bout were able to be recorded from one observer for all 10 horses.

4.3 Individual vs. group analysis

As the data analysis of the seasonal periodic of behavioral patterns came clearly into light, the combination of all data of all horses was followed by the loss of partly complete individual reactions or seasonal behavioral patterns. Nevertheless, the horses could not behave independently from each other or from the resources of their pasture; this should be taken into consideration for the interpretation of the results.

A behavior pattern can be analysed by using those parameters: the frequency, the duration per bout and the effectiveness of an action. The behavior pattern differences among these horses may not have been analysed from the total time of a particular behavior in 24-hour. The frequency or the average duration of each bout are more suitable to evaluate the behavior responsiveness to environmental difficulties for an individual which lives under social constraints.

Variations in time spent on a behavior by individual horses can be attributed to the change in various environmental situations. The feeding behavior, for example, depends to a high degree on individual social bonds (Nielsen, 1999).

The behavior of group-housed social animals depends on additional special conditions. The availability of the locally limited supply of food, social constraints and individual preferred space may have also affected the behavior of the group as a whole and likewise their synchrony of activities. Also the level of motivation in the individual animal can differ from the behavior of the group as a whole. The horses in this investigation showed only a middle degree of feeding synchronisation. This result does not correspond to other measurements made on the same species or on other social living animals, but it reflects how the living conditions greatly affect the behavior of the observed animals (Duncan & Cowtan, 1980; Mayes & Duncan, 1986; Scheibe et al., 1999). The comparability of the observations of different authors will depend on the comparability of their methods of investigation and on the living conditions of the animals they have observed.

The specific food conditions of a trial, the individual preferred food and the feeding rate (Nielsen et al., 1995) are a few of the additional factors concerning the attempt to interpret different behavioral measurements. It may be easier to regard the social group rather than the individual which adapts under the social constraints to its environment (Broom, 1986). The observation conditions were equal for all animals but the locally limited supply of same resources vs. the locally/timely limited access to special (maybe individual) preferred resources, can change the behavior pattern of some individuals more than that for the group or for less influenceable animals.

References:

1. Altmann, J. (1974). Observational study of behaviour: Sampling methods. *Behaviour*, 49(3-4), 227-267.
2. Broom, D.M. (1986). Indicators of poor welfare. *Br. vet. J.*, 142, 524-526.
3. Duncan, P., & Cowtan, P. (1980). An unusual choice of habitat helps Camargue horses to avoid blood-sucking horse-flies. *Biology of Behaviour*, 5, 55-60.
4. Lamprecht, J. (1999). Biologische Forschung - von der Planung bis zur Publikation. Fürth. Filander Verlag
5. Lorenz, R.J. (1996). Grundbegriffe der Biometrie. Stuttgart. Gustav Fischer Verlag
6. Martin, P., & Bateson, P. (1993). Measuring behaviour. Cambrige, University Press
7. Mayes, E., & Duncan, P. (1986). Temporal patterns of feeding behaviour in free-ranging horses. *Behaviour*, 96(1-2), 105-129.

8. Nielsen, B.L., Lawrence, A.B., & Whittemore, C.T. (1995). Effect of group size on feeding behaviour, social behavior, and performance of growing pigs using single-space feeders. *Livestock Production Science*, 44(1), 73-85.
9. Nielsen, B.L. (1999). On the interpretation of feeding behaviour measures and the use of feeding rate as an indicator of social constraint. *Applied Animal Behaviour Science*, 63(1), 79-91.
10. Sambraus, H.H. (1982). Der Einfluß des Menschen auf Verhalten und Leistung von Tieren, *Aktuelle Arbeiten zur Artgemäßen Tierhaltung* (pp. 122-137): KTBL-Schrift 291.
11. Scheibe, K.M., Eichhorn, K., Schleusener, T., Streich, J.W., & Heinz, C. (1999). Chronobiological analysis of animal locations - development of an automatic recording system and principles of data processing, 15th International Symposium on Biotelemetry. Junau, Alaska, USA.

A presentation of a set of new annotation tools based on the NXT API

D. Reidsma, D. Hofs, N. Jovanović
Human Media Interaction research group, Centre for Telematics and Information Technology, Enschede, the Netherlands

Abstract
The creation of large, richly annotated, multimodal corpora of human interactions is an expensive and time consuming task. Tools which are highly efficient for one particular annotation task are not necessarily so for other tasks. In some cases, the possibility to achieve a reduction of the effort needed to create annotations for a new corpus can justify the development or adaptation of a tool for that particular annotation task. At the MB2005 symposium "Annotating and Measuring Meeting Behavior" we present our view on the considerations that should drive the design of new tools geared to specific tasks. In this paper we describe several annotation tools which have already been developed using that method: a tool for markup of structured discourse entities and relations in transcriptions, a tool for labeling of continuous, non-overlapping phenomena in audio and/or video and an adaptation of the FeelTrace tool for labeling of emotion. All tools were developed on top of the Nite XML Toolkit.

Keywords
Annotation tools, large corpora, multimodal annotation, Nite XML Toolkit

1 Introduction
This paper presents a number of annotation tools that have been developed in the context of the AMI project [7]. The standoff XML data formats used for storage [2], the NXT API for using corpora stored in that format [1] and some of the design considerations underlying these tools [5] have all been discussed elsewhere. For information about such topics the reader is referred to the publications cited above. This paper simply indicates what kind of annotations can be made with these tools and how they can work together on a common corpus stored in the NXT data format.

2 General information
The tools all make use of the same corpus exchange format and API so they can operate on one shared corpus, each contributing different parts of the annotation data. A corpus in NXT format by definition is accompanied by a meta-data file describing the structure of the corpus (which layers are there, how do they relate, which label sets exist, etc). Customization of the tools to a new corpus is usually done by naming the different layers and label sets which are to be used for annotating. In fact, the tools can be adapted to a different annotation task on the same corpus by specifying different configuration settings. If more complicated changes are necessary to speed up a specific annotation process, the tools can be adapted through the extensive API available for both the underlying corpus toolkit and the tools themselves. The tools allow the user to view all video and audio signals for a recorded session, simultaneously if necessary. They contain search facilities based on the Nite Query Language [4]. Finally, all tools are part of the Nite XML Toolkit, and therefore freely available from the Sourceforge site [6].

3 Discourse labeling tools
The first set of tools that we present here are designed for *discourse labeling tasks*. One of their main features is a component designed for flexible visualization of the transcriptions in a corpus. It can be configured to find the content of the transcription from the appropriate parts of the corpus, which may be different for each corpus. The display of elements such as nonverbal vocal sounds can be customized, and the transcription view can be set to visualize annotations with an arbitrary relation to the core transcription as visual markup (e.g. inserted text and/or style changes) defined in the corpus. This transcription visualization component also implements a wide range of text-selection strategies, so the annotator can be supported in his/her task by only allowing selection of fragments which are 'valid' in terms of the layer that is being annotated. For example, a tagger might allow only selection of single words, a dialog act segmentation configuration would only allow selection of arbitrary segments of one speaker at a time, a topic labeler might allow only the selection of (several) full sentences, but of potentially more than one speaker, etc. The final feature of the transcription visualization mentioned here is the mechanism which allows different kinds of highlighting in the marked-up transcriptions, among which highlighting of search results and highlighting of elements aligned to the current time of the video player.

This module has been combined with the NXT video playing and search modules, and a number of customizable user interaction functions, to produce among other things a flexible dialog act coder with support for relational coding, a discourse entity tagger (e.g. for named entities) and an RST coder. Figure 1 shows a screenshot of one of these configurations.

Figure 1. Screenshot of a named entity tagger.

4 Continuous signal labeling tools

The second set of tools that we present here target a specific class of signal labeling tasks. There are many annotation (sub)layers that segment the timeline of a recording into non-overlapping segments with a label. If the density of these segments is not too high and the labels are based on simple observations, this task could be done at 'real-time' while the signal plays, by having the annotator press buttons or keyboard keys for each new segment, possibly with a slightly slowed-down video replay. Examples of these are focus of attention, posture annotation and emotion annotation such as annotated with FeelTrace [3]. For this class of annotation tasks we have developed a 'framework tool' which can be configured with special-purpose GUI elements data-input and visualization of certain layers aligned with a variable speed replay of video(s). Figure 2 shows this tool in a configuration used to try out the FeelTrace method for annotating emotion taken from [3].

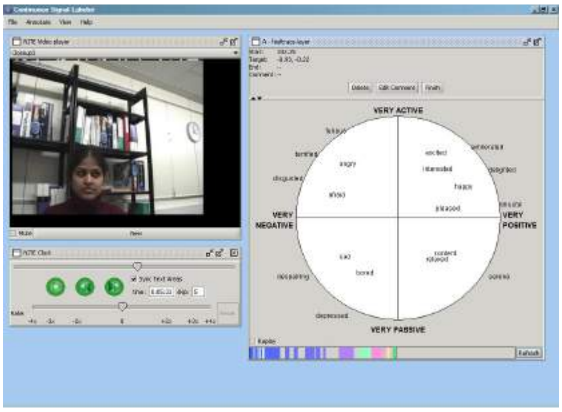


Figure 2. FeelTrace configuration of the video labeling tool.

5 Conclusions

In this paper we briefly touched upon a new set of annotation tools, divided in two classes: discourse labeling and continuous video labeling. For more information as well as for demo versions of existing configurations of these tools the reader is referred to the Nite XML Toolkit homepage, http://www.ltg.ed.ac.uk/NITE/.

This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-105).

References:

1. Carletta, J; Evert, S; Heid, U; Kilgour, J; Reidsma, D; Robertson, J; The NITE XML Toolkit. (submitted).
2. Carletta, J; Kilgour, J; O'Donnell, T; Evert, S; and Voormann, H (2003). The NITE Object Model Library for Handling Structured Linguistic Annotation on Multimodal Data Sets. *Proc. EACL Workshop on Language Technology and the Semantic Web.*
3. Cowie, R; Douglas-Cowie, E; Savvidou, S; McMahon, E; Sawey, M; Schröder, M (2000). 'FEELTRACE': An instrument for recording perceived emotion in real time, *Proc. ISCA Workshop on Speech and Emotion*. Belfast
4. Evert, S; Voormann, H (2002). NITE Query Language. *From: http://www.ltg.ed.ac.uk/NITE/*
5. Reidsma, D; Hofs, D; Jovanović, N (2005). Designing Focused and Efficient Annotation Tools. *MB2005, these proceedings.*
6. http://www.sourceforge.net/projects/nite/
7. http://www.amiproject.org/

Observational assessment of independent living skills

J. C. Rogers , M. B. Holm
Department of Occupational Therapy, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Observational performance-based assessment of daily activities in rehabilitation yields information about clients' activity abilities and disabilities and also guides intervention. Standardized tools, such as the Functional Independent Measure, include functional mobility and personal care activities. In contrast, the complexity of home management activities, such as managing medications and preparing meals, has retarded the development of observational measures. The Performance Assessment of Self-Care Skills (PASS) was developed to fill this void. The PASS is a criterion-referenced, performance-based observational tool consisting of 26 core items, categorized in 4 domains: functional mobility personal care physical instrumental activities of daily living (IADL) and cognitive IADL. For rating, each PASS item is broken down into observable subtasks and rating is based on the performance observed. Each PASS item is rated for three separate measurement parameters – independence, safety, and adequacy of activity performance. The PASS has established validity, and acceptable to excellent inter-observer and test-retest reliability.

Keywords

Function, observation, performance, activities of daily living

Background

The ability to live independently, and "age in place" requires more than dressing and feeding oneself. It is the instrumental activities of daily living such as the more complex home management tasks and the ability to manage one's finances, medications, and meal preparation that often make the difference between independent and assisted living. When competency to manage one's own affairs had to be decided by the court, judges did not accept self-report data, but rather required performance-based observational data upon which to base their decisions. This was one impetus for developing the PASS. The validity of the PASS is derived from operationalizing items from common self-report tools. Inter-observer reliability ranged from 0.80 to 0.99 for Clinic and Home versions of the PASS, and test-retest reliability with a 3-day interval ranged from 0.82 - 0.97 [1, 2].

PASS Components

The PASS measures real time performance of 26 core task situations. Each PASS item has 7 task components: (1) assessment conditions, (2) instructions, (3) subtask criteria, (4) assistance data grid, (5) safety data grid, (6) adequacy data grid, (7) summary scores grid, and (8) scoring guide. Each component will be discussed in turn.

(1) Assessment Conditions

The assessment conditions are described for standardization purposes, and an example is shown in Figure 1.

Table and
1. Seven-day medication sheet with 4 subdivisions for each day –Morning, Noon, Evening, Bedtime.
2. Two of Client's (Ct) medications:(if Ct is taking only 1 medication, substitute PASS-Home "Child-proof lid" medication containing red vitamin pills and a typed label that reads "Take 2 with breakfast and one with dinner." If Ct is not taking any prescription medications, add the non child-proof lid medication containing yellow vitamin pills and a typed label that reads "Take 1 tablet at bedtime."
3. Items 1- 2 on table in front of Ct.
4. Ct seated at the table.

Figure 1. *PASS-Home Medication management conditions.*

(2) Instructions

For research purposes, the instructions are read to the subject (see Figure 2). For clients, the instructions are sometimes modified to the person's level of understanding

"The next task involves managing medications. Please show me where you keep your medications. [Select 2 medications to use for the assessment, preferably with different time schedules]. I need to borrow these two bottles for the next task. I will put them back as soon as we are done." [Return to the table where the task items are set up].
" Please read the prescription label and find the directions for taking this medication [Hand Ct first bottle of own medication or bottle with child-proof lid and wait until Ct looks up]. If you were taking this medication today, when would you take the next pill?" [Wait for response].
"This medication organizer has the days of the week across the top [Point] and the time of the day [Point] along the side. Using the organizer, distribute the pills to be taken tomorrow and the following day according to the directions on the prescription label. Do you know what you are to do? Do you have everything that you need?" [Wait for response]
Repeat process with second medication.

Figure 2. *PASS-Home medication management instructions.*

(3) Subtask criteria

Each task consists of several critical observable subtasks. The behavior to be observed separates the task process with a double underline from the task quality criterion, which has a single underline and is followed by samples of quality criteria in parentheses (see Figure 3). The task process is also the focus of the independence rating, whereas both task process and quality are included in the overall task adequacy rating. Task safety also is reflected in performance of both the process and quality criteria.

1. <u>Reports next time</u> first medication is to be taken <u>correctly</u> (based on testing time, matches direction on label) 2. <u>Open first pill bottle</u> <u>with ease</u> (by second try) 3. <u>Distributes pills from the first pill bottle</u> into <u>correct</u> time slots <u>for the next 2 days</u> (all pills & all slots indicated; days indicated)

Figure 3. *PASS-Home medication management criteria for first medication.*

(4) Assistance data grid

For each subtask, a grid with 9 levels of assistance is provided to identify the type of assistance needed for the Client to perform the subtask according to the criteria. Levels of assistance are arranged in a hierarchy -- from least intrusive/least costly to most intrusive/most costly, as follows:

1. Verbal supportive
2. Verbal non-directive
3. Verbal directive
4. Gestures
5. Task or environment rearrangement
6. Demonstration
7. Physical guidance
8. Physical support
9. Total assist

The least intrusive assists are given first, unless there is a safety concern, and then the appropriate assist is given for the situation. All assists are operationally defined, and examples are given in the PASS manual.

(5) Safety data grid

A space is provided for each subtask for the assessor to note any aspects of task performance that are unsafe during task performance.

(6) Adequacy data grid

Spaces are also provided opposite each subtask to note inefficiencies or missing steps in the task process as well as noting when quality standards identified in the criteria are not met.

(7) Summary scores grid

Using data from the independence, safety and adequacy grids, summary scores for independence are given for each subtask, whereas for safety a single score is given for the total task, and likewise for adequacy. The scoring guide is used to assist with scoring.

(8) Scoring guide

Scoring of all 3 constructs is on an ordinal scale that ranges from 0 to 3, with higher scores indicating greater independence (see Figure 4), safety (see Figure 5, and (see Figure 6) of task performance.

3	No assists given for task initiation, continuation, or completion
2	No Level 7-9 assists given, but occasional Level 1-6 assists given
1	No Level 9 assists given; occasional Level 7 or 8 assists given, or continuous Level 1-6 assists given
0	Level 9 assists given, or continuous Level 7 or 8 assists given; or unable to initiate, continue, or complete subtask or task

Figure 4. *Scoring guide for PASS independence scale.* adequacy

3	Safe practices were observed
2	Minor risks were evident but no assistance provided
1	Risks to safety were observed and assistance given to prevent potential harm
0	Risks to safety of such severity were observed that task was stopped or taken over by therapist to prevent harm

Figure 5. *Scoring guide for PASS safety scale.*

	Quality	Process
3	Acceptable (standards met)	Subtasks performed with precision & economy of effort & action
2	Acceptable (standards met but improvement possible)	Subtasks generally performed with precision & economy of effort & action; occasional lack of efficiency, redundant or extraneous actions; no missing steps
1	Marginal (standards partially met)	Subtasks generally performed with lack of precision and/or economy of effort & action; consistent extraneous or redundant actions; steps may be missing
0	Unacceptable (standards not met)	Subtasks are consistently performed with lack of precision and/or economy of effort & action so that task progress is unattainable

Figure 6. *Scoring guide for PASS adequacy (quality and process of outcome) scale.*

Studies have shown that the three constructs measure separate aspects of task performance [3, 4, 5].

References:

1. Rogers JC, Holm MB. (1989, 1994). *Performance Assessment of Self-Care Skills (PASS-Clinic).* Unpublished test, University of Pittsburgh, Pittsburgh, PA, USA
2. Rogers JC, Holm MB. (1989, 1994). *Performance Assessment of Self-Care Skills (PASS-Home).* Unpublished test, University of Pittsburgh, Pittsburgh, PA, USA
3. Rogers JC, Holm MB. (2000). Daily living skills and habits older women with depression. *The Occupational Therapy Journal of Research,* **20** (Supplement)*,* 68S-85S.
4. Rogers JC, Holm MB, Beach S, Schulz R, & Starz T. (2001). Task independence, safety, and adequacy among nondisabled and osteoarthritis-disabled older women. *Arthritis Care & Research,* **45**, 410-418.
5. Rogers JC, Holm MB, Beach S, Schulz R, Cipriani J, Fox A, & Starz TW. (2003). Concordance of four methods of disability assessment using performance in the home as the criterion method. *Arthritis Care & Research,* **49**, 640-647.

Observational outcome measures to evaluate assistive technology use by people with dementia

G Singh[1], K Pichora-Fuller[1], E Rochon[2], JB Orange[3], & P Spadafora[4]
[1] Department of Psychology, University of Toronto, Toronto, Canada
[2] Toronto Rehabilitation Institute; Department of Speech-Language Pathology, University of Toronto, Toronto, Canada
[3] School of Communication Sciences and Disorders, University of Western Ontario, London, Canada
[4] Sheridan Elder Research Centre, Sheridan College Institute of Technology and Advanced Learning, Oakville, Canada

Abstract
This short paper describes a digital video-audio behavioral observation methodology for use in a naturalistic setting to evaluate communication rehabilitation interventions for older adults with dementia. Behavioural observation via recorded video-audio offers a number of advantages over other data collection methodologies, which can be subject to a number of biases and limitations, some of which are discussed. In this study, high quality digital audio-video recordings were collected on participants attending a respite care day program. Recording equipment was inconspicuously placed, and measurement occurred either during normal day-to-day activities or during more directed activities (e.g., playing bingo). The recordings can be used to document the occurrence of behaviors and paired behaviors of interest over extended periods of time or selected samples of interest can be downloaded for detailed analysis including lag-sequential analyses. It was found that behavioral observation can complement traditional objective measures of impairment and subjective questionnaire measures, in accordance with the World Health Organization's (WHO) *International Classification of Functioning, Disability, and Health*[1].

Keywords
Naturalistic observation; Outcome measures; Digital video-audio recordings; Communication rehabilitation interventions; Alzheimer's disease

1 Introduction
We developed new outcome measures using naturalistic observation to evaluate the effect of interventions on the communication and social interaction of people with dementia of the Alzheimer's type. The participants attended a day program in a new facility designed for the dual purposes of providing respite care and conducting research on aging. The facility includes three large activity rooms, a greenhouse, and a kitchen with a dining area. Day program rooms are equipped with built-in digital video cameras and microphones. The cameras are high quality color domes with 22x optical zoom that offer full pan, tilt, and zoom with 30 frame per second output, and rotate 360 degrees. The cameras are strategically placed (i.e., mounted on the ceiling) in the rooms to allow for very detailed close-ups of participants in target contexts (to capture focal behaviors). Alternatively, they can be used by operators to follow mobile subjects around a room with ease (to capture focal individual behaviors). The audio system is based on a Symetrix Symnet DSP system, a modular system that concurrently terminates and manages connections from 56 Shure MX-202/W microphones. The cameras and microphones are controlled by PC and connected to an American Dymanics Digital Video Recorder 160000 capable of recording 16 high resolution video channels and 1 audio channel simultaneously. The control unit can display live audio-visual signals in real-time or archive signals digitally to mass-storage computer servers capable of saving several weeks of recordings from multiple rooms. In the console, the video windows display and record all video feeds concurrently. The windows can be individually displayed at 640x480 resolution at up to 30 frames per second. The system has a 160 gigabyte hard drive for archive and retrieval of both audio and video streams. The recordings can be reviewed and selected segments stored for detailed analyses.

The recordings can be used to document the occurrence of behaviors and paired behaviors of interest over extended periods of time. Selected samples of interest can be downloaded for detailed analysis including lag-sequential analyses. First, we examined automatic recordings of daily activities to determine relevant behaviors in the repertoire of the clients with dementia participating in the regular activities of the day program. Next, we focused more closely on discourse and detailed analyses of communication behaviors during selected activities (e.g., playing bingo). We also recorded planned conversational encounters between different communication partners and individual clients. These observations were used to develop baselines against which change due to intervention can be evaluated. The first intervention was to use an FM assistive listening device in small group activities with an activity leader. The advantages of behavioral observation over existing measures and their usefulness in this application will now be discussed.

1.1 Advantages of Behavioural Observation
The World Health Organization's (WHO) *International Classification of Functioning, Disability, and Health*[1] has been a catalyst for the re-conceptualization of health care practices.[2,3] The WHO classification delineates that functioning with a health condition can be understood at the level of the body's *impairment* (e.g., hearing loss), at the level of the individual's ability to engage in *activity* (e.g., understand speech), and the level of *participation* in society (e.g., communicating as a spouse). These three levels are not necessarily highly correlated because the impact of impairment(s) on activity and participation can be modulated by personal and contextual factors, including social and physical environments that may be adverse or supportive.

In the biophysical tradition of health care, clinicians conduct objective tests to measure an individual's impairments in highly artificial but controlled standardized conditions. It is reasonable to assume that loss of hearing, vision, cognition, and language will have detrimental effects on communication and the ability to function in many everyday situations. Nevertheless, traditional tests of impairment are insufficient for estimating how a person does or could function in everyday life. Some individuals with severe impairments function well in everyday life, while others who have mild impairments experience extreme challenges. More recently, clinicians have turned to the development of questionnaires in an attempt to try to gain insight into functioning at the levels of activity and participation in everyday life. These types of subjective measures have indeed advanced our knowledge of everyday functioning. Nevertheless, clinicians remain perplexed when trying to reconcile the results of traditional clinic-based tests and the subjective reports of patients and families, and they continue to lack reliable and valid tools to evaluate everyday function in context.

The capacity to remotely record participants' focal behaviors with video and audio offers several advantages. First, it is well established that the presence of an observer can cause those being observed to alter their behaviour from that which would normally occur if the observer were not present. Although this may also be true of video-audio recording equipment, this effect is potentially minimized by the inconspicuous nature of recording equipment (e.g., video cameras covered with domes), and by the fact that video-audio recording equipment is an increasingly common feature of public spaces. A second advantage of using recorded video-audio over direct observation is that the digital recordings allow for multiple examinations of the data across time, providing the opportunity for more in-depth analysis of the focal behaviors. Further, the original data can be made available to researchers who were not present during data collection. This has the potential to increase observer reliability and validity.

Conclusion

We have developed a new naturalistic observation methodology to complement traditional objective measures of impairment and subjective questionnaire measures. Naturalistic observation of everyday behaviors will enable us to learn how to improve our techniques for assessing and treating problematic communication behaviors in relevant contexts. In accordance with the World Health Organization's (WHO) *International Classification of Functioning, Disability, and Health*[1], an observational approach enables us to understand the influence of the environment on group and personal behaviors at the levels of activity and participation. By identifying and defining observable communication and associated behaviors that can be mapped to the subjective reports of family and staff, we will develop a new ecologically valid way of objectively assessing everyday communication function and measuring treatment outcomes. Our new specialized behavioral scoring system will include measures of both focal behaviors (scoring based on target behaviors of group members) and/or focal individual behaviors (scoring based on the repertoire of behaviors of an individual in the group).

Naturalistic observation methodology can be used to help establish valid ecological measures to plan and to evaluate the outcomes of interventions in terms of how communication behaviors affect activity and participation (as defined in the World Health Organization ICF, 2001).

Supported by the Canadian Foundation for Innovation, the Toronto Rehabilitation Institute, and the Canadian Institute for Health Research.

References:

1. World Health Organization (2001). *International classification of functioning, disability and health.* Geneva, WHO.
2. Kiessling, J., Pichora-Fuller, M.K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., Davis, A.C., Erber, N.P., Hickson, L., Holmes, A., Rosenhall, U., & von Wedel, H. (2003). Candidature for and delivery of audiological services: Special needs of older people. *International Journal of Audiology, **42** (Supplement 2)*, S92-S101.
3. Worrall, L., & Hickson, L. (2003). *Communication disability in aging: Prevention to intervention.* Singular Press: San Diego.

Measuring vigilance level in demanding work setting

A.S. Smit[1], J. Benito[2], J. Troost[2] and C.H.M. Nieuwenhuis[1]
[1] *Decis Lab, Thales Research & Technology, Delft, The Netherlands*
[2] *Thales Nederland, Hengelo, The Netherlands*

Abstract
Vigilance can be measured in a number of ways, of which the electroencephalogram (EEG) is probably the most popular. However, the EEG method is not practical in a dynamic work setting for a number of reasons. This paper introduces a technique for assessing vigilance by means of easy-to-use and sensitive objective (performance) and subjective (questionnaire) behavioral measures. This technique does not include predefined or specified tasks, but a general principle of assessment. Vigilance assessment can be achieved by administering a vigilance test battery that includes a short working memory task and a short questionnaire. In situations that do not allow presentation of test batteries, both vigilance and the degree of (under-) stimulation can be assessed by performance on work-embedded tasks. This, however, requires analysis of the task set for a given job. It is discussed with what parameters a measurement technique needs to be described in order to determine exactly which measure(s) to use in a particular situation.

Keywords
Vigilance measures, alertness, work setting, task performance.

1 Introduction
Vigilance level strongly affects performance, especially on cognitively demanding tasks. Therefore, it is important to be able to determine the level of vigilance of individual people in many work settings. For example, an operator in a control room setting often needs to work in cognitively demanding situations under high mental work load. Traditionally, the level of vigilance has been determined using three different kinds of measures: physiological, performance and subjective. More stable measures can be obtained when they are simultaneously collected from different sources.
It has proven very difficult to use physiological measures as indices of vigilance level on an individual basis. Furthermore, some of them – such as the widely used EEG – are artefact-sensitive and utilize equipment that is costly, intrusive in a work setting, and requires calibration for individual differences.
It has been demonstrated [1] that vigilance can be estimated with much simpler measures such as subjective alertness scales and task performance indicators (see Figure 1).

2 Approach
Subjective vigilance questionnaires can be reliable vigilance measures, comparable with physiological data. However, it is essential that the questionnaire is multifactorial and includes all relevant factors, such as a physical factor [1].
Behavioral measures require some form of task performance indicators. The level of performance is stated in terms of reaction times and errors (misses and false alarms). It is not only important to measure average performance, but also the changes in performance over time. That is, a decrement in performance is indicative of declining vigilance of an individual [2].

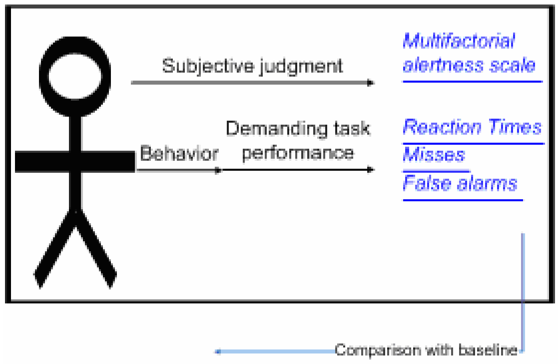


Figure 1. *Reliable and easy-to-use vigilance measures.*

Of crucial importance is the type of task used to assess vigilance. Classically, monotonous tasks have been used. Rationale for this type of task stems from the arousal theory, which states that vigilance drops due to habituation that is caused by stimulus monotony. In other words, boredom causes vigilance to drop. In contrast, cognitively demanding tasks have been used lately. These tasks are based on a resource view on vigilance; processing of information requires cognitive resources, which are limited in availability and diminish if used continuously. Particularly suitable tasks are working memory tasks [1].
Results of different studies [1,2] suggest that these two kinds of tasks address different processes. Performance on cognitively demanding tasks probably reflects vigilance in a strict sense: the availability of resources and thus the capacity for cognitive processing. If someone has performed a demanding task for a long time and processing resources are low, then performance will be seriously hampered, also on a consecutive task.
Performance on monotonous tasks reflects the level of cognitive 'underload' or the understimulation level. Indeed, if a person is left chronically understimulated, he or she may tend to fall asleep. However, after this type of understimulation a person can be activated and be 'awoken' by mere stimulation (e.g. by increasing light intensity, presenting a warning sound, advising someone to stretch his/her legs, etc). Performance will be improved after this stimulation.
Vigilance can thus be determined with a short alertness questionnaire or with a performance measure on cognitively demanding tasks. However, their combination yields more reliable vigilance indicators. Therefore, a vigilance test battery should include, at least, a short memory task and a multifactorial subjective questionnaire [1]. Cognitive understimulation can be measured with monotonous monitoring tasks, but this is harder to measure with a test battery, as it requires performance on a task that lasts at least half an hour.

However, it may be impractical or even impossible to continuously present someone with a test battery – especially for people working in chaotic environments such as in crisis situations. A different mode of measurement should then be employed.

If no test battery presentation is possible vigilance level should be derived more naturally from work environment tasks. One may think of checking incoming email or answering the mobile phone. It is important to keep in mind that for vigilance determination those tasks should be chosen which are the most cognitively demanding *and* are measurable in terms of reaction time and/or quality of response (misses and false alarms). Therefore, a one-size-fits-all measurement does not apply. Deciding which task(s) to choose for vigilance determination requires analysis of the work domain. It is the procedure or method that can be applied in all situations, but not the actual tasks themselves. In order to estimate understimulation, a task should be chosen that is monotonous and preferably requires (only) monitoring.

It will not always be appropriate or possible to combine all available measurements. In order to determine which measure or set of measures should be combined in a specific situation, every particular measure could be described in terms of the next parameters [4,5,6]:

- **Bandwidth**: the reliability and availability of the measure in assessing vigilance across different levels. A particular measure may not operate with the same accuracy in different levels of vigilance. Self-reports will be shed by those being monitored. This parameter should be specified with values for different levels of vigilance (for instance: low, medium and high).
- **Intrusion**: the degree to which the measure interferes with and disrupts the performance of tasks. An intrusive measure can degrade the performance of the tasks an individual has to carry out, and therefore, its use is undesirable in high demanding situations. However, it may be useful in low demanding situations as a way of avoiding understimulation.
- **Acceptance**: the degree of approval of the measure by the individual. The individual's opinion about the measure method determines the correctness and accuracy of the measure. Acceptance of a measure is directly related with the complexity and usefulness perceived by the individual.
- **Selectivity**: the discrimination capability of the measure for vigilance level assessment. A vigilance level measure can also be sensitive to other human states like mental workload, fatigue or stress. This conditions its use for vigilance level assessment, depending on the specific situation.
- **Expertise**: the degree to which the measure deals with proficiency aspects. In some measures, especially performance measures, the individual's experience and training may confound the estimation of the measure. Skills and expertise are known to have a considerable impact on performance [3]. Even when a person is highly vigilant, he or she may not be able to perform well when facing a totally unfamiliar situation. For example, a skilled air-traffic controller would probably not know how to fly an airplane, not even in a vigilant state. On the other hand, the performance of an exhausted person may still be acceptable , purely due to experience.

An example of a simulated descriptor for a vigilance level measure based on a questionnaire, using XML and natural language (awful, bad, reasonable, good and excellent) to appraise the parameters, could be:

```
<measureDescriptor>
<bandwidth>
        <low>excellent</low>
        <medium>good</medium>
        <high>bad</high>
</bandwidth>
<intrusion>reasonable</intrusion>
<acceptance>bad</acceptance>
<selectivity>good</selectivity>
<expertise>reasonable</expertise>
</measureDescriptor>
```

Using the right set of measurements, combined with the right description in terms of relevant parameters, vigilance can thus be assessed in a simple, non-intrusive, but sensitive manner within any type of work setting. We would like to stress that the measurement method or procedure introduced here is applicable to not only vigilance determination, but can be applied – with the use of other tasks and questionnaires – in assessing other human transitory states such as workload or stress.

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024. The ICIS project is hosted by DECIS Lab (http://www.decis.nl) , the open research partnership of Thales Nederland, the Delft University of Technology, the University of Amsterdam and the Dutch Foundation of Applied Scientific research (TNO).

This research has also been supported by a Marie Curie fellowship of the European Community program "Mutual Adaptive Human Computer Interaction" under contract number HPMI-CT-2002-00221.

References:

1. Smit, A. S. (2004). *Vigilance or availability of processing resources: A study on cognitive energetics.* PhD thesis, Radboud University Nijmegen.
2. Smit, A. S., Eling, P. A. T. M., and Coenen, A. M. L. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta Psychologica*, **115**, 35-42.
3. Parasuraman, R., and Davies, D. R. (1984). *Varieties of attention*. Academic Press, Inc, Florida.
4. O'Donnell, R.D. & Eggemeier, F.T. (1986). *Workload assessment methodology*. In K.R. Boff, L. Kaufman & J.P. Thomas (Eds.) *Handbook of perception and human performance*. New York: Wiley.
5. Wickens, C.D. (1992). *Engineering psychology and human performance*. New York: HarperCollins.
6. De Waard, D. (1996). *The measurement of drivers' mental workload*. PhD thesis, Traffic Research Centre, University of Groningen, Haren.

Observing processes reflected in and through music: a coding scheme to monitor music therapy

A. Raglio[1], D. Traficante[2,3], O. Oasi[2]
[1] *Music Therapy School "Glass Harmonica", Anni Verdi Association, Rome, Italy*
[2] *Department of Psychology, Catholic University, Milan, Italy*
[3] *CRTI – Education Technologies Research Center, Catholic University, Milan, Italy*

Abstract
We developed a coding scheme to observe the interactions between music therapist and patient during a music therapy session. We collected approximately twenty-one hours of video tape concerning active music therapy sessions, based on the sonorous-musical improvisation. Subjects were seven children (3 males; 4 females) aged 3-10 years (mean age = 5,14), diagnosed with autism disorders, and seven therapists. Our coding scheme consisted in four different behavioral classes: Verbal communication, Non-verbal communication, Countenance, Musical communication. The data collection method adopted was continuous recording, applied through *The Observer Video-Pro 4.0*. Reliability and concurrent validity indexes were computed. In general, there was a substantial agreement between the assessments of the video raters.

Keywords
Music therapy – Assessment procedure – Coding scheme – The Observer Video-Pro 4.0

1 Introduction
The aim of our work is to build an instrument to observe and to evaluate changes in the interactive behavior during music therapy session. We referred to the pychodynamic theory: in particular, the Stern's theory of affect attunement and the Benenzon's approach to music therapy.

1.1 The psychodynamic framework
According to Stern, the dance and the music are examples of the expression of the vitality affects [8] and contribute to determine the conditions of the affect attunement, on which a primitive sense of Self is built on. The sonorous element is an integral part in the protocommunicative processes, characterizing the child's development. It has natural and archaic aspects from which its symbolic potential springs. Some interesting empirical researches [1] underlined the child's competence to use the vocal rhythm in the early interactions. Trevarthen [10] showed that the child uses the sound with natural competencies, aimed to the communication and to the relationship through intersubjective exchanges. So, our hypothesis is that the sonorous element takes much part in the process of the mutually regulation of the emotions between patient and music therapist (MT) as well as between the infant and its caregiver [12]. This process has important effects on the mental organization and helps the child to modulate his/her emotions, to give sense to the world and to develop cognitive abilities. Also in the adult case it works to create transformative moments or to produce moving along in the psychoanalytic treatments [2][9].
The emotional component connected to music therapy treatments is also a main element that can influence the neuro-hormonal state and can induce a development of the human brain. Siegel [7] suggested that the states of the child's mind have a bent to the attunement with the states of parent's mind. They use non verbal communication and produce an emotional content that lead to a deep intersubjective relationship [11]. Recently the neuroscientific researches are bringing important evidence about this subject [5].

1.2 The Benenzon's music therapy approach
The Benenzon's music therapy approach [3] is based on the use of the sonorous-musical improvisation in a non-verbal context. The sonorous-musical element is conceived as the expression of the sonorous identity ("iso"), whose roots are deep and archaic: they are "essential" sounds, sonorous-musical elements used by the MT and by the patient in an expressive and relational key.
The sonorous-musical element produces a transitional area, a bridge between internal and external world, a space for the sonorous-musical exchanges, in which a shared communication can develop. The musical instrument and/or the sonorous-musical material acquire a therapeutic value, when they become mediators of the relationship ("intermediary objects"), means by which MT and patient communicate.
The aims of the music therapy intervention are pursued through a predisposition to observe and to listen to the patient, with a non-directive criterion. This kind of attitude in the MT favors the manifestation of the expressive instances of the patient, that can freely emerge in the unconscious components too.

2 Materials and methods

2.1 Indicators of attunement
We can hypothesize that the affect attunement [8], in a music therapy session, is present when:
• in the productions of MT and patient the profile of the sound intensity is analogous;
• the sonorous-musical exchanges have the same beats;
• the rhythm is followed in a synchronous and antiphonal way by the patient and the MT;
• it is possible to establish a relationship between the length of the two productions;
• there is a formal analogy in rhythmic and/or melodic framework.
The attunement, in music therapy, can emerge even in a transmodal way, that is through different sensorial channels.
From the attunement, favored by the use of the sonorous-musical element, the interpersonal relationship springs.
The indicators of the interpersonal relationship, that we considered in our work are the following [8]: the sharing of the attention, the sharing of the intentions, the sharing of the affections.

The three indicators are present even in the music therapy process: if MT and patient focus their attention on the sonorous-musical event, the sharing of attention is gradually reached. The sharing of intentions implies the voluntariness of the sound-gesture, by which the patient puts him/herself in relation with the MT. Finally, the sonorous-musical event can induce an emotional and affective involvement in the dyad, leading to the sharing of affects.
The tuning process between MT and patient seems to be the most important index of the quality of the therapeutic process. To assess this interactive modality, we built a coding scheme that can be applied to video recordings.

2.2 The coding scheme

We developed the *Music Therapy Coding Scheme* (*MTCS*) as assessment method for video recordings of music therapy sessions. The *MTCS* is a video-based, patient-therapist interaction assessment method measuring both the affective and behavioral characteristics expressed by the patient and the therapist in a dyadic interaction during a music therapy session.
We identified 4 behavioral classes: 1) *Non verbal communication*, that includes the spatial collocation in relation to the other and to the musical instruments, the actions made in relation to the other and to the musical instruments, the expression of the emotions; 2) *Countenance*, referred to looking to the other, to the musical instruments, and to the environment; 3) *Verbal communication*, in terms of presence or absence and, in case of presence, pertinent or not to the music therapy interaction; 4) *Musical communication*, referred to the quality and the shape of the musical interactions.
To code the video recordings we used *The Observer Video-Pro 4.0* software. We chose a continuous recording method.

2.3 Participants

Subjects were seven children (3 males; 4 females) aged 3-10 years (mean age = 5,14), diagnosed with autism disorders, and seven MTs.

2.4 Procedure and assessment of videotaped interactions

The patient-therapist dyads were videotaped by a fixed video camera, positioned in a corner of the room in which the music therapy session took place. We considered the 1st, the 10th and the 20th sessions for each patient.
The observers were two psychologists and two MTs. The videos were assessed independently by each of two couples including one psychologist and one music therapist. We coded the 15 min. middle part of each session.

3 Results and conclusion

In this study we adopted Cohen's kappa coefficient [4] to evaluate the similarity between the assessments of the two couple of observers. We focused our attention in particular on the items that indicate the quality of the music therapy interaction: the MT's production with attunement, the patient's production with attunement, the MT's individual production or MT's production without attunement, the patient's individual production or patient's production without attunement. The *Observer Video-Pro 4.0*, considering a tolerance window of 2 seconds, produced for these four categories k = 0,69, index of a substantial agreement, according to Landis and Koch [6].

The observation of the plots produced by the software gave us indicators useful to assess the change during the therapy. For example, figure 1 tells us that during the 1st session the patient SQ has shown very brief phases of production with attunement (*CVN sonora, Produz. sint.*) and long phases of individual production or production without attunement (*CVN sonora, Produz. individ.*). In the same plot we can observe that the therapist spent a lot of time in looking to the patient (*Espr.volto*, *Sguar. altro*), while the attention of the patient has been concentrated on the instruments (*Sguar. strum.*). Finally, the plot shows that the therapist tended to stand still, while the patient very often has been moving (*CNV, Stat, Deambulazione*).
Verbal communication was absent in this videotape.

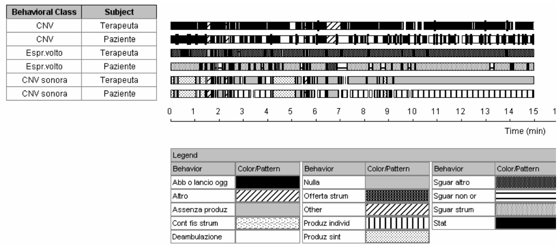


Figure 1. *Plot of the interactions between the patient SQ and the therapist in the first session.*

On the base of these results we can consider the *MTCS* a useful instrument to pick the main features of the therapist-patient interactions during a therapy session.

References:

1. Beebe, B. *et al.* (2000). Systems models in development and psychoanalysis: The case of vocal rhythm coordination and attachment. *Infant Mental Health Journal*, **21**, 99-122.
2. Beebe, B.; Lachmann, F.M. (2002). *Infant Research and Adult Treatment: Co-constructing Interactions*. The Analytic Press, Hillsdale NJ.
3. Benenzon, R.O. (1981). *Manual de musicoterapia.* Editorial Paidos Iberica, Barcelona.
4. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37-47.
5. Gallese, V. (2001). The "shared manifold" hypothesis: From mirror neurone to empathy. *Journal of Consciousness Studies*, **5**, 33-50.
6. Landis, J. R.; Koch, G. G. (1977). Measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174.
7. Siegel, D.J. (1999). *The Developing Mind,* Guilford Press, New York.
8. Stern, D. (1985). *The Interpersonal World of the infant*. Basic Book, New York.
9. Stern, D. (2004). *The Present Moment in Psychotherapy and Everyday Life*. Norton & Company Ltd, London.
10. Trevarthen, C. (2001). Autism, sympathy of motives and music therapy. *Enfance*, **1**, 86-99.
11. Trevarthen, C.; Aitken, K.J. (2001). Infant Intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry*, **1**, 3-48.
12. Tronick, E.Z. (1989). Emotions and Emotional Communication in Infants. *American Psychologist*, **44**, 112-119.

Use of BASIS Installed in Palm Device for Observation of Child's Interaction with Physical and Social Environment

A Sarkar [1], A Dutta[2] U Dhingra[2] P Dhingra [1], P Verma [1], G S Hiremath,[2] CP Khokhar[3], R Juyal[1], M Osmany [1], V P Menon [1], J Kumar[1], R. E. Black[2], S Sazawal[1,2]

[1] Center for Micronutrient Research, Annamalai University, New Delhi, India, [2] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore USA,.[3] Department of Psychology, Gurukul Kangri University, Haridwar, India.

Abstract

In this study, we assessed the behavior and social interactions of preschool children with its social and physical environment. Traditionally, the data collection is based on the manual recording of the observations of parent child interaction which is cumbersome, prone to observation errors and inconsistencies. We developed and tested 'Behavior and Social Interaction Software' (BASIS), installed on Palm Handheld device to record the observations of social interactions in natural setting for 1 hour on 296 children. The multiple screens of recording observations were child's location, broad activity of the child, child's state, and interactions of the child with socializing agent; with a set of choices for each screen. Data were downloaded and stored in a laptop everyday at the base station. In our experience, BASIS installed in the Palm Handheld device is an innovative, easy and error free method to address the complexities of measuring the child's interactions with its socializing agents in an unrestricted environment.

Keywords

Social Interaction, Preschool children, BASIS, Appforge, Palm

1 Introduction

Overall development of a child has multiple determinants and the transaction between these determinants contributes significantly to the developmental process. A critical determinant is the social matrix within which the child is developing- that is the parent child interactions. A comprehensive evaluation of parent child interaction is important to understand the quality of influence of immediate environment on the child's behavior. A structured assessment of a child's interaction with a parent in natural setting can provide important information regarding the young child's behavioral capacities to focus and sustain attention, retain and follow instructions, persist and solve problems, and make use of environmental facilitation. The parent's participation during the evaluation of social interaction of the child can provide an objective evaluation of the quality of parent child interactions [1]. Most of the studies assessing the social interaction of the child with their socializing agents have used the manual recording method of direct observation which is cumbersome, prone to observation errors and inconsistencies.

In this study, we assessed the verbal, visual and motor interactions of the child with its natural social and physical environment by developing and testing an observational automated interactive software 'Behavior and Social Interaction Software' (BASIS), and installed on Palm Handheld device to record the observations.

2. Methodology

The study was conducted among 296 children aged 1-3 years in Sangam Vihar, a low to middle income peri-urban population in Delhi, India.

2.1 Development of observation software in palm

Based on the literature, the constructs and codes of child's interaction were identified and designed in four major windows as setting, situation, state and interactions with a set of codes and subcodes for each category. BASIS was designed using Visual Basic 6.0 software integrated with AppForge Crossfire^{TM} [2]. Palm Database was used to record the data at backend. The software was compiled to a palm OS based executable file and uploaded to the palm handheld (M-125 series) device with necessary support files to run the application [3].

BASIS operates in an interactive fashion with multiple screens representing set of choices for a particular category enabling the examiner to select one, which gets recorded against time with other parameters such as setting, situation, other participating agents etc.

Identification data of scheduled children were uploaded into palm database using a custom designed software and the desired child was selected from the list of children during the assessment.

The main screen had the windows for recording location, situation and state of the child each of which was recorded independently against time. The screen also had icons for recording the child's and the socializing agents (mother, father, peer, siblings) interaction to enable continuous recording of dyadic or multiple interaction of the child with more than one socializing agents in the natural setting. The recorder needed to select the desired location option by clicking from the available behavioral classes which further represented a set of choices for that behavioral class ensuring subject wise recording of the behaviors continuously against time. The Interaction was the main window for the analyses of data whereas other windows were designed to record the supporting variables.

Alarm was set at 1 hour to help the examiner to stop coding the behaviors at the end of the session and save the recording by closing the session. The lithium ion battery of the Palm ensured 2 hours of recording in the field.

Data were downloaded and stored in a laptop everyday by performing the "HOTSYNC" operation facility provided by the Palm software at the base station located in the central office and next day's scheduled children were uploaded. Separate Desktop based software was designed to view and edit the data collected using BASIS. An in built designed software was used to convert the palm data base files to populate in access data base which could be read into SPSS data base for analysis.

2.2 Recording Observations in the field

The observer carried the palm to the child's house and conducted continuous direct observations of the child's behavior and interactions for one hour in the natural setting. Setting gave the duration of location of the child (inside or outside the house) in the observation session. Situation gave the duration of the child's broad activities in the session. State gave the general behavioral state of the child during the session. Interaction was used to record both child's and socializing agent's interaction. The main screens for recording observations in the interaction were : Ask/ Calls, Look at/ Show, Affection/ Acceptance, Play/ Activity, Verbal Aggression/ Negativity, Ignore/ Avoidance, Respond/ Talk, Care Activities, Physical Aggression, Physical Activity, Refers to Others. Each of these categories had a set of choices and the appropriate code was clicked by the observer. The Interaction opened into an Agent sub window where the particular socializing agent with whom the behavior occurred was recorded.

2.3 Outcome Measures

The behaviors for observing the caretaker-child interaction were selected and suitably modified to assess the interaction. The parent measures included warmth, aggression, neglect, nurturance, response, playfulness, and behavior involvement. The child's measures included aggression, self reliance and adequacy, independence, emotional stability, responsiveness, social behavior and quality of play.

2.4 Reliability and Training

An inter observer reliability of more than 80% was achieved and maintained for BASIS observations. A reliability program was also designed in the software to facilitate training of observers.

2.5 Analyses

The PDB files in the palm data base were exported to SPSS 12.0. The codes of a particular outcome variable were collated and the frequency and percentage of each of these outcomes measures were calculated.

Table 3.1. *Observation of Mother's behavior using BASIS*

Variables	None	Occasional	Frequent
Nurturance	50 (16.9)*	97 (32.8)	149 (50.2)
Playfulness	60 (20.3)	116 (39.2)	120 (40.5)
Positive Response	5 (1.7)	17 (5.7)	274 (92.6)
Physical Aggression	252 (85.1)	39 (13.2)	5 (1.7)
Verbal Aggression	95 (32.1)	106 (35.9)	95 (32.1)
Neglect/ Indifference	62 (20.9)	111 (37.5)	123 (41.5)

** Number (%)*

3. Results

For this paper only results of mother child interaction have been presented. The results of mother's behavior with children showed almost 50% of mothers displayed frequent nurturance and playful behavior with children and almost all mothers showed positive response to their child's initiatives. Though display of physical aggression was less but around 40 % of mothers frequently scolded and showed instances of neglect/ indifference towards their children (Table 3.1). The results of children's behavior showed around 15% of them frequently indulged in acts of physical and verbal aggression during the observation session (Table 3.2).

Table 3.2. *Observation of Child's behavior using BASIS*

Variables	None	Occasional	Frequent
Physical Aggression	160 (54.1)	91 (30.7)	45 (15.1)
Verbal Aggression	167 (56.4)	104 (35.2)	25 (8.5)
Positive Feeling	52 (17.6)	120 (40.5)	124 (41.8)
Independence	5 (1.7)	13 (4.4)	278 (93.9)
Dependence	117 (39.5)	130 (43.9)	49 (16.6)
Emotional Instability	88 (29.7)	106 (35.8)	102 (34.5)
Emotional Responsiveness	1 (0.3)	3 (1.0)	96 (1.3)

** Number (%)*

4. Conclusion

In our experience, BASIS installed in the Palm Handheld device is an innovative method to address the complexities of measuring the child's interactions with its socializing agents in an unrestricted environment. The software as well as the mobility of the Palm device enabled us to adopt a micro assessment approach to capture meaning and content of interaction. Additionally, there are inherent advantages due to the ease of recording diverse behaviors, observing more numbers of interactions, reducing data entry errors, and achieving high inter observer reliability. Without an automated software and a mobile hand held device, this level of detailed recording of structured data of mother child interaction in an analyzable form would not have been achievable. The portability of the palm device ensured that the child was never out of sight during the observation session. Though there is a slight advantage due to small screen of the palm hand held device but the manifold advantages of palm makes it a suitable and low cost option for recording intricate social interactions of child in a field setting.

The study was supported by New Zealand Milk Limited, Private Bag 12032, Auckland, New Zealand.

References:

1. Clark, R., Paulson, A., and Conlin, S. (1993). Assessment of developmental status and parent infant relationships: The Therapeutic process of evaluation. In C. Zeanah (ed.). *The Handbook of Infant Mental Health*, (pp 191-209). New York: Guilford Press.
2. AppForge Crossfire, Version 5.1 (2004). *AppForge Extends and Simplifies Microsoft® .NET Development to All Mobile and Wireless Devices*, AppForge, Inc., Georgia, USA.
3. Palm™ , Inc. (2001). California, USA.

Analysis of driver behavior at crossroads in urban areas

M. Williams
Institute of Ergonomics, Darmstadt University of Technology, Darmstadt, Germany

Abstract

The aim of the experiment was to analyze the driver behavior at two different types of crossroads (controlled either by yield sign or by traffic light) in urban areas in order to find out if there are differences between the groups of subjects and if the drivers are confronted with different cognitive and motor demands depending on the type of crossroad. The driver behavior was analyzed with the Observer ® System. The results of the study confirmed the assumption that the two type of crossroads pose high cognitive demands on the driver, while the motor demands are higher for the crossroad controlled by a yield-sign than for the crossroad controlled by a traffic light.

Keywords

Driver behavior, glance behavior, crossroad, task demands

1 Introduction

Crossroads in urban areas (especially the uncontrolled ones) are often associated with a high number of accidents. One of the reasons are the high demands (both cognitive and motor) posed to the drivers. The data from the accident statistics help us only to identify the problems, but not to develop strategies in order to improve the traffic safety. Hereto are data from in-depth studies of the driver behavior essential. The question is, if the available tools for behavioral studies are also suitable for analyzing the driver behavior in complex situations.

2 Methods

The Observer ® System for collection and analysis of observational data, Version 3.0 (Noldus Information Technology b.v., Netherlands, 1995) has been used in order to analyze observational data collected during the field experiment. The system consists of modules for data collection, evaluation and presentation [2]. In the beginning a configuration containing the relevant behavioral must be implemented. In our case the behavioral classes were: driving task, road type, glances, head movements, torso movements and the position of the hands. Every behavioral class contains several elements (e.g., the behavioral class glances contained elements like: looking to the front, looking to the left, looking to the right, looking in the internal mirror, looking in the left mirror, looking in the right mirror, looking to the dashboard, looking in an unidentified direction). The behavioral classes and their elements were the same for all subjects in order to have the possibility of making comparisons. The analysis respectively coding of the video tapes is made, using a dedicated computer, by a specially trained analyst.

2.1 Experimental Set-up

Although 36 persons participated in the experiment only the data collected from 30 subjects covered both situations and have been therefore analyzed. Three different age groups were represented: 18-25 years (6 subjects, 4 males and 2 females), 26-55 years (9 subjects, 7 males and 2 females), and over 55 years (15 subjects, 7 males and 8 females). Regarding the driving experience, 4 groups have been built: less than 100,000 km (7 subjects), between 100,001 and 500,000 km (17 subjects), between 500,001 and 1,000,000 km (3 subjects) and from 1,000,001 to 2,000,000 km.

Two driving situations, respectively two types of crossroads have been included in the test plan: stop and go at the traffic light (Figure 1) and approaching a busy crossroad, driver has to give way, (Figure 2). In case of the first situation, while approaching the crossroad the traffic light changes from green to red.

Figure 1. *Stop and go at the traffic light.*

Regarding the second situation, the subjects had to approach the crossroad, give way (see yield sign on the right side) and drive straight ahead over the crossroad.

Figure2. *Approaching a busy crossroad.*

These driving situations have been chosen because of the different cognitive and motor demands posed to the driver [1]. For example a sign-posted intersection in urban area (driver has to give way) poses high demands concerning both information processing and operation of primary controls; the driver has to notice very quickly optical differences, he/she has to judge speeds, spacing, sizes, and time limits and to make decisions under uncertainty.

The test car, upper class category and having an automatic transmission, has been equipped with three video cameras in order to collect behavioral data (e.g., glances, head movements, torso movements): one camera for the frontal road scenery, one for the rear scenery and one for the glances. By the mean of a multiplexer the three pictures were recorded on the same video tape. After the trial the images were de-multiplexed in order to allow the simultaneous view of all three images (quad-split modus). The evaluation of the tapes was made frame by frame by a special trained analyst and using the above-mentioned Observer-configuration, which was specially programmed for this purpose.

3 Results

Different groups of subjects showed a different behavior. The elderly drivers, for examples, use a different strategy than the younger drivers in order to perform certain tasks. This was obvious in the case of approaching a busy crossroad. The subjects from the 3^{rd} age group looked less frequently to the left and to the right than the other subjects. A possible explanation is that due to their longer driving experience they can estimate better and quicker than the other subjects the gaps in order to pass. In case of stop and go at the traffic light, there are not major differences between the subject groups. Due to the fact, that they had to wait until the traffic light turned green, the subjects looked mostly to the traffic light. Differences appeared for example, when the driver had to wait at the traffic light behind another vehicle, or when other vehicles were waiting behind the test vehicle.

4 Discussion

A first conclusion is that the Observer is an adequate system to study the driver behavior, although the analysis is in some cases very time-costly. The time needed for the analysis depends first of all on the number of behavioral cases and their elements. Secondly, the duration is also influenced by the changes within the driving situation itself which change the state of the elements of the behavioral classes, all these changes have to be coded by the analyst. The ratio between analyzed video-time and the coding-time was mainly 1:10. Furthermore, only a specially trained analyst can code the video-tapes according to the implemented configuration. The results of our study confirm also other findings from literature: both cognitive and motor demands are the highest for approaching crossroads in urban areas where the driver has no priority [1]. In case of stop and go at the traffic light the demands posed on the driver in terms of information processing are high, but the motor demands are rather low.

Supported by the European Union, Project Virtual (Virtual reality systems for perceived ergonomic quality testing of driving task and design), Contract No. : 1999-RD.11030.2000-2002.

References:

1. Fastenmeier, W. (1995). Die Verkehrssituationen als Analyseeinheit im Verkehrssystem. *Autofahrer und Verkehrssituationen (Mensch-Fahrzeug-Umwelt,* ***Band 38****)*, 27-69.
2. Wakula, J. (1999). Möglichkeiten und Grenzen des Videoanalysesystems "Observer" als Werkzeug bei der Arbeits- und Belastungsanalyse in Feldstudien. *44. Arbeitswissenschaftliche Kongress*, Karlsruhe, Germany.*10.-12.März1999.*

An observation and rating scheme for driving situations

M. Williams [1] , M. Rakic [2]
[1] *Institute of Ergonomics, Darmstadt University of Technology, Darmstadt, Germany*
[2] *Active Safety, DaimlerChrysler AG, Sindelfingen, Germany*

Abstract
The aim of the study, focused on driving on motorway, was first of all to develop an observation scheme for driving situations which should both give the possibility of integrating different types of data and be practicable. The observation scheme was based on the worked out definition for the driving situation. The relevant areas around the reference vehicle were identified and divided in fields, building a matrix. Parameters like speed, headway were associated to each vehicle positioned in the identified fields. The practicability has been checked on a two-lane German motorway. Subsequently, the identified driving situations were rated in terms of information sources, information processing decision-making and operating the controls. The results showed that using the developed scheme is an advantage

Keywords
Driving situation, motorway, driver behavior

1 Introduction
There are some methods already available for observation and rating of driving situations. They allow to differentiate between different aspects of a task, i.e. information processing, decision-making and operating the controls, see for example [2]. The disadvantage of these methods is that a group of situations like driving on motorway is considered as a whole i.e. there is no differentiation between the numerous situations belonging to the group. On the other hand, the dynamic character, i.e. the rapid transitions from a situation respectively component to another one is not taken in consideration. But this dynamic character is linked to high demands posed to the driver consequently also to occurrence of critical situations. This dynamic aspect is therefore relevant for traffic safety.

2 Definition of driving situations
Before starting to develop the scheme an adequate definition for the driving situation was needed. A driving situation consists of one or more components and each component can be described by the mean of specific parameters. During the whole driving situation the components and their parameters do not change. In case of driving on motorway the following components were identified: free driving, driving behind another vehicle, driving in front of another vehicle, overtaking, and being overtaken.

For example, the component "driving behind another vehicle" can be described by the following parameters:

- Headway, defined as the distance (in meters) between the reference vehicle and the proximate vehicle driving ahead in the same lane.
- Time headway, defined as the time interval (in seconds) between the reference vehicle and the proximate vehicle driving ahead in the same lane.
- Relative velocity, defined as the difference between the speed of the proximate vehicle and the speed of the reference vehicle.
- Changes in the speed of the proximate vehicle driving ahead in the same lane.
- Type of the proximate vehicle driving ahead in the same lane.
- Number of vehicles driving in the observed lane.

Lane-change maneuvers either of the reference vehicle or of other vehicles make the transition from a component to another one.

3 Observation scheme
The first step in developing the observation scheme consisted in division of the areas surrounding the reference vehicle (i.e. in front, behind, on the left side, and of the right side of the reference vehicle) in fields (Figure 1).

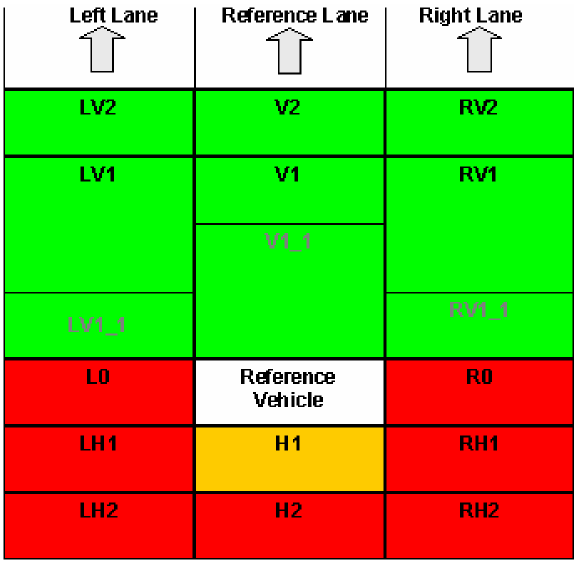


Figure 1. *Division of the surrounding areas.*

LV1, V1, and RV1 have following characteristics:

- Each field covers a distance of 100 m.
- Each of them has a temporal-defined subdivision (LV1_1, V1, respectively RV_1).

In the next step, parameters (e.g., speed, headway) are associated to each green field, respectively to the vehicles located in these fields. The data are obtained from the reference vehicle sensors, object recognition system, and frame-by-frame video analysis.

For the yellow field (directly behind the reference vehicle) only subjective data are available (the experimenter rated whether distance between the reference vehicle and the next vehicle is small or not), and for the red fields are no data available.

4 Rating scheme
The basis for the selection of relevant rating criteria was a version of the observation scheme for the FAA-rating,

which was adapted for the driving task [2]. The FAA-rating is the German version of the “Position Analysis Questionnaire (PAQ)” from McCormic, Jeanneret, Mecham [4]. The relevance of the criteria from the FAA-rating for driving on a two-lanes German motorway was checked and criteria without relevance for this driving task were removed. Furthermore criteria found to be very similar were combined.

An important theoretical underpinning for building the structure of the rating scheme was the system analytical view on a human being. Within the Man-Machine-System the human being is regarded as a system element consisting of the elements information reception, information processing and acting [1].

To establish a basis for consistent ratings, especially if the rating is done by more than one person, the definition of rating keys is necessary. Therefore the three keys importance (i), accuracy (a) and speediness (s) were defined.

Putting that all together, the following rating scheme (Table 1 and Table 2) was developed. The identified driving situations were rated according to this scheme.

Table 1. *Rating criteria regarding Information reception/ Information processing.*

source	**rating criteria**	**key**
Own vehicle	Observing instruments like the tachometer etc.	i
Other road users (RU)	Estimating the velocity of other RUs using the: ▪ windscreen ▪ mirrors	i
	Estimating the distance of other RUs using the: ▪ windscreen ▪ mirrors	i
	Observing of visual signals: ▪ optical horn ▪ brake lights ▪ blinking ▪ emergency flashers	i
	Observing disturbances on the lay-by: ▪ „traffic routing“ ▪ „break down“	i
	Observing of acoustical warning signals	i
road	Observing road signs	i

Table 2. *Rating criteria regarding Acting.*

Rating criteria	**key**
Conducting essential steering maneuvers	a, s
Conducting essential velocity changes	a, s
Blinking before lane change	i
Controlling of light switches and wipers	i

5 Practical implementation

The practicability has been checked by using data from an experiment conducted on a two-lanes German motorway.

The essential parameters regarding the observation are obtained from the reference vehicle sensors, an object recognition system and if there isn’t any information available from these sensors, e.g. if vehicles are to far away to be automatically tracked, a frame by frame video analysis is conducted. A tool, being able to display synchronously video, and both measured and manually coded data, was necessary for the video analysis. For this purpose, the SAVE Tool (v3.0), developed at the Department of Ergonomics, Technical University of Munich, was used.

6 Discussion

With the observation scheme for the classification of the space around the own vehicle a method, which considers all possible combinations within the regarded area, is given. Of course the described observation scheme is designed only for two-lanes German motorways, but an adaptation to e.g. rural roads is possible without any concerns. It is much more difficult to overview the whole tableau when all possible situations are taken in consideration.

The use of the observation scheme is very advantageous, but only if a small part of the needed parameters has to be gained trough a frame by frame video analysis. It is also evident, that this analysis is very time-consuming. But it has to be mentioned that the effort for training the analyst is very small because of the clear topology of the observation scheme.

A rating according to the described scheme allows the identification of situations with similar demands on the driver regarding Information reception/ Information processing or Acting. In this way a customized selection of situations when planning a study is possible.

Finally it has to be mentioned, that the interaction between the specific demands on the driver follows certain laws, yet, there is a lack of knowledge about these laws. Furthermore, without knowing anything about these laws no scientifically proofed all-over measure of demand can be performed, so here is a great need for further research.

References:

1. Bubb, H.; Schmidtke, H. (1993). „Systemstruktur“ and „Informationswandel durch das System“ in Ergonomie.
2. Fastenmeier, W. (1995). Autofahrer und Verkehrssituation – Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme. (Mensch-Fahrzeug-Umwelt, **Bd. 33**).
3. Landau, Kurt. (1978). Das arbeitswissenschaftliche Erhebungsverfahren zur Tätigkeitsanalyse. Dissertation, Fachbereich Maschinenbau an der Technischen Hochschule Darmstadt.
4. McCormic E.J.; Jeanneret, P.R.; Mecham, R.C. (1969). The development and background of the Position Analysis Questionnaire.

Development of the P-CAMPIS: Assessing interactions in the perioperative environment

Alison A. Caldwell-Andrews, PhD[1], Zeev N. Kain, MD[1], Ronald L. Blount, PhD[2].
[1] *Department of Anesthesiology, Yale University School of Medicine, New Haven, Connecticut, USA*
[2] *Department of Psychology, University of Georgia, Athens, Georgia, USA*

Abstract

We developed a behavioral coding system to measure child and adult interactions and communication in the perioperative environment. We established interrater reliability and preliminary validity of this system by examining videotapes of parents, medical personnel, and 45 children (2-12 years old) who underwent elective surgery and general anesthesia. The final coding system (titled the Perioperative Child Adult Medical Procedure Interaction Scale; P-CAMPIS) contains 40 codes in 4 domains. Analyses showed excellent reliability overall for verbal and nonverbal codes. Kappa values averaged 0.87 for verbal codes characterizing adult vocalizations, 0.92 for verbal codes characterizing child vocalizations, and 0.88 for nonverbal codes. Construct validity was demonstrated by finding the hypothesized associations between certain P-CAMPIS codes and children's anxiety (p=0.0001). We conclude that the P-CAMPIS is an appropriate tool for assessing child-adult behavioral interaction during the perioperative period.

Keywords

Children, parents, behavior, surgery.

1 Introduction

The perioperative environment refers to the setting and atmosphere wherein patients undergo surgery. This environment includes the various physical settings of the perioperative cycle including: admitting patients to the surgical unit, patients waiting in the preoperative holding area, bringing patients into the operating rooms, induction of anesthesia, surgical procedure, patients' recovery in the post-anesthesia care unit, subsequent discharge from the surgery unit and postoperative recovery in the hospital and/or at home. The perioperative environment also includes, however, an emotional milieu of feelings, expectations and individual appraisals of the surgery for a number of different people including the patient, doctors and nurses, as well as the many and complex interactions that occur between these above elements. Furthermore, the *pediatric* perioperative environment includes all of the above with the addition of young patients who are at varying levels of emotional and cognitive development, as well as these patient's parents and their concomitant emotions, expectations, and interactions with the both the patient and entire medical staff. Consequently, in any pediatric perioperative setting, emotions are often at the forefront of each interaction and can profoundly impact the overall surgery experience in terms of compliance with procedures, doctor-patient relationships, doctor-parent rapport, parent and patient satisfaction, and patient recovery.

A number of interventions aimed to decrease anxiety and increase compliance have been targeted to patients and their parents throughout the perioperative cycle. These interventions have most often consisted of sedative medication for the patient, preoperative preparation programs and/or the use of parental presence during induction of anesthesia (PPIA). Research into the effectiveness of PPIA continues to be a controversial issue. Although early studies suggested that children experience reduced anxiety and increased cooperation when their parents are present during induction [1, 2], later investigations indicated that routine parental presence during induction of anesthesia is not beneficial in terms of reducing children's anxiety or increasing children's compliance [3-7]. These reports, however, should be interpreted cautiously as they do not take into account *what parents actually do* during induction of anesthesia. In fact, we have been told on numerous occasions by various experienced anesthesiologists that "in their hands," PPIA is an effective practice, suggesting that variables beyond the mere presence or absence of the parent are at the forefront of effective anxiety reduction and enhancement of compliance. We have recently suggested that this field of research should shift toward an emphasis on what parents actually do during induction of anesthesia, rather than simply on their presence. Furthermore, we also strongly suggest that the perioperative behaviors of participating health care providers should also be evaluated, since these individuals also have a large potential to impact children's anxiety levels. However, an appropriate measurement instrument and technology is first necessary to make these evaluations.

Previous research into behavioral interactions during medical procedures was done by Blount and colleagues who investigated the influence of parents' and health care providers' behaviors on the coping and distress of children with chronic illness during painful medical procedures [8]. Blount developed the Child Adult Medical Procedure Interaction Scale (CAMPIS), a system of behavioral codes that categorize child and adult verbal interactions during painful medical procedures [8, 9], and consequently identified specific parental and health care provider behaviors that preceded children's distress and coping. The CAMPIS contained 35 behavioral codes including 19 codes for adults' behaviors and 16 for children's behaviors [9]. Code types on the CAMPIS include adult–to-adult vocalizations, adult–to–child, and child-to-adult vocalizations. We determined that the although the CAMPIS was the right type of instrument to use to assess parent-child-medical personnel interactions, the CAMPIS however needed extensive modification before it was appropriate for use in the perioperative environment. That is, children and parents undergoing acute stressful procedures such as surgery have very different behaviours as compare to children with chronic illness who undergo a surgical procedure. We therefore set out to modify the CAMPIS for use in the pediatric perioperative environment. The purpose of this paper is to describe this modification and the resulting instrument.

Figure 1 refers to our overall conceptual framework. Behavioral interactions are moderated by child and parent characteristics and these interactions impact the child's anxiety and subsequent postoperative recovery.

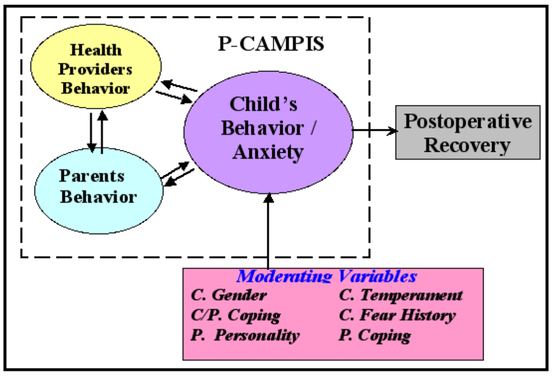


Figure 1. *Behavioral Interactions Influence Child's Anxiety and Postoperative Recovery. C=Child; P=Parent.*

2 Methods and Results

2.1 Procedures

As a first step for this investigation we assembled a multidisciplinary task force including experts in anesthesiology, pediatrics, child psychiatry, child psychology and child development. We then recruited to this study 45 children (2-12 years old) who were undergoing elective surgery and general anesthesia and whose parents were present during the induction of anesthesia. No child received sedative premedication. Exclusion criteria included children with chronic illness or developmental delay, children taking psychiatric medication, and children with parents who did not speak English. The Yale Human Investigation Committee approved this study and parents and children provided written informed consent and assent as appropriate.

Parents and children were videotaped for a period of 5 minutes as they waited in the preoperative holding area. In addition to agreeing to be videotaped, parents also completed a demographic questionnaire and allowed the research staff to rate their child's anxiety using the modified Yale Preoperative Anxiety Scale, a valid and reliable observer measure of perioperative state anxiety. In order to assure a variety of interactions, videotaping did not occur at the same time for each patient. That is, videotaping occurred for some children while the nursing staff or anesthesiologist was talking with the family, and for other children videotaping occurred when the family was merely waiting in the holding room.

The child and one parent were then brought into the operating room. The entire process of induction of anesthesia was videotaped. Children were placed on the surgical bed and were asked to breathe through an anesthesia mask that was scented with the child's choice of 'flavors' (e.g., orange, strawberry, bubble gum, etc.). Anesthesia was induced using the same technique (that is, the same anesthetic agents) for each child. Researchers again evaluated the child's state anxiety (m-YPAS) both upon entering the operating room and again upon introduction of the anesthesia mask. After anesthesia was induced, researchers escorted parents to a waiting area.

Children recruited for this study ranged in age between 2-12 years old, and the ratio of males to females was 66% to 34%. The attendant parent during induction of anesthesia was the mother in 74% of the cases.

2.2 Development of the Instrument

All verbal interactions and non-verbal interactions in these videotaped recordings were first transcribed and typed by members of the research staff. The multidisciplinary team reviewed the first 15 videotapes, noting recurring behaviors that occurred during the children's and adults' interactions. Based on our experience with previous development of behavioral instruments [10] as well as clinical experience, and upon repeated observations of these 15 particular videos, we determined which of the existing CAMPIS codes should be modified, deleted or retained.

We then began an iterative process of testing and refining the coding system. We added a number of codes representing behaviors unique to the perioperative environment and developed appropriate labels for these new codes. We also carefully considered all developmental aspects of the behaviors captured in each new and original CAMPIS code. Since children undergoing induction of anesthesia do not undergo procedural pain, we opted to delete the several pain-related codes in the CAMPIS. Likewise, some adult behaviors captured in the CAMPIS and observed during painful medical procedures were not also observed in the perioperative environment; we therefore deleted these behavior codes. We next changed multiple definitions and behavioral examples to better fit the perioperative environment.

After verbal interactions had been satisfactorily coded, the task force examined nonverbal behaviors. We first catalogued each nonverbal behavior that we could identify and noted their relative frequency across the 15 videotaped examples. We then discussed the relative importance of each nonverbal behavior in terms of its likely impact on preoperative anxiety and compliance. We also discussed the relative likelihood that these nonverbal behaviors could be reliably identified by a number of trained observers. Finally, we carefully defined each nonverbal behavior and catalogued it in a list of P-CAMPIS nonverbal codes.

We next wrote a codebook, describing each preliminary P-CAMPIS verbal and nonverbal code, and giving examples and explanations. Following the development of the preliminary codebook we used 5 additional transcriptions to explore the fit of these preliminary codes. The task force met weekly and discussed coding disagreements, modifying and clarifying codes where needed. After the task force was satisfied with these refinements, raters then coded 5 new transcripts (total transcriptions used at this point = 25) to calculate preliminary reliability of the P-CAMPIS codes. We determined interrater reliability estimates for each of the verbal and nonverbal codes, using kappa statistics provided by the computer program ComKappa (Version 1.0, 1997, © Roger Bakeman & Byron Robinson).

Initial results showed that kappa reliability for all codes averaged 0.62 for verbal codes and 0.64 for nonverbal codes. Based on these initial results and on further discussion of coding disagreements, the P-CAMPIS was refined again and a set of 5 of these transcriptions were tested by an independent members task force for coding reliability. Kappa values for this last set of transcriptions averaged 0.803 for verbal codes and 0.787 for nonverbal codes. As a reference point, verbal codes of the original CAMPIS showed kappa of 0.79 (range of 0.53 to 0.94), and codes added to the P-CAMPIS showed an average kappa reliability of 0.84 (range of 0.78 to 0.89).

2.3 Transportability and Reliability

Finally, in order to assure that P-CAMPIS codes were transportable to researchers who were not part of the original P-CAMPIS task force, a second set of independent naïve coders were then trained to use the P-CAMPIS. A last set of 20 new transcripts (total transcriptions used at this point = 45) was then coded. Kappa values for the final version of the P-CAMPIS averaged 0.87 for verbal codes characterizing adult vocalizations, 0.92 for verbal codes characterizing child vocalizations, and averaged 0.88 for nonverbal codes. The following interpretations of clinical significance apply to individual weighted kappa values: kappa (W) <0.40 = poor, 0.40-0.59 = fair, 0.60-0.74 = good, and 0.75-1.00 = excellent [11].

2.4 Description of the P-CAMPIS Instrument

The final version of the P-CAMPIS contains 40 codes that characterize verbal and nonverbal interactions between children, parents and medical personnel in the perioperative setting. Coders can apply the most appropriate codes to either written transcriptions of interactions (verbal codes) or videotaped interactions (verbal and nonverbal codes) using specific criteria as described in the 23-page P-CAMPIS manual. Five of these codes describe adult-to-adult communication, including behaviors such as use of humor, commands for managing their child's behavior, etc. Eighteen codes describe communications between and adult and child and include behaviors such as commands to engage in procedural activities, reassurance, empathy, and giving control to the child. One code, vocalizations that refer to the child's future health status, can be applied to either adult-adult interactions or adult-child interactions. Nine codes describe child vocalizations, including distress behaviors such as crying, and coping behaviors such as nonprocedural talk and request for support. Finally, seven codes describe nonverbal behavior such as empathic touch and nonverbal resistance.

2.5 Preliminary Validity

To establish preliminary evidence for validity of the P-CAMPIS codes, we next examined partial construct validity of the P-CAMPIS by determining the relationship of some of the codes to the child's anxiety. Specifically, we hypothesized that increased verbalized fear, resistance and crying from the child as assessed on the P-CAMPIS should be associated with increased preoperative anxiety as assessed by the m-YPAS. We identified all children who exhibited at least one of the following behaviors: fear, resistance and crying, and then compared this group of children to children who did not exhibit any of these three behaviors. An independent t-test examining state anxiety differences between these two groups of children found that indeed, children who verbalized resistance, fear or crying as coded on the P-CAMPIS were scored by observers (mYPAS) as significantly more anxious during induction of anesthesia (75.5 ± 18.2 vs. 44.2 ± 22.4, p=0.001).

We used independent t-tests to analyze the relationship between P-CAMPIS subscale scores on children's anxiety. Data were analyzed using SPSS 13.0 (SPSS Inc, Chicago, IL) and are reported as mean ± standard deviation. Significance was accepted at $p<0.05$.

3 Discussion

This investigation supports the reliability and initial validity of the newly developed P-CAMPIS. Kappa statistics showed excellent reliability and confirmed expected associations between P-CAMPIS codes and children's perioperative anxiety, indicating construct validity. The creation of this scale is the first necessary step in understanding the building blocks that underlie behavioral interactions in the perioperative environment. This understanding is crucial to be able to develop empirically based behavioral preparation programs that target interactions between parents, children and health care providers.

Previously, Blount and colleagues developed a working model to conceptualize how various factors impact anxiety and distress behavior in children undergoing painful invasive medical procedures [12, 13]. We suggest that this model is valid for parental presence during induction of anesthesia. This model suggests that there are primarily two types of factors that predict the reaction of children in acute painful medical situation. Proximal factors, are the parental and health care staff behaviors that occur immediately prior/ during the medical procedure and distal factors which are baseline variables such as temperament, child's age, and level of distress during past medical procedures [14-17]. While distal factors may strongly impact child's anxiety and distress, these factors are difficult or impossible to change. In contrast, proximal factors, such as parental behaviors and health care provider behaviors, can be modified, thus resulting in decreased child's distress. In a study of children receiving immunizations, Frank found that proximal factors accounted for 38% of the variance in child distress and 55% of the variance in child coping [18]. Given the strong contribution of proximal factors to children's reactions during painful medical procedures, it is not surprising that a number of recent experimental studies involving children undergoing immunizations, VCUG, and lumbar puncture have demonstrated that parental behavior training programs effectively reduce children's distress [19-22]. Despite methodological difficulties in some of these preliminary interventional studies it is clear that once the appropriate parent-child interaction patterns are identified, training parents and health care providers to change their state-like behaviors during invasive medical procedures is a valid and effective approach to reducing children's distress. We suggest that children's anxiety and distress during induction of anesthesia also can be reduced in large part by modifying proximal factors such as parental and health care providers behaviors.

It is important to note that sequential analysis can best determine which specific adult behaviors are likely to actually prompt particular child behaviors. That is, simply looking at correlations or frequencies between parental behaviors and anxiety in children will only indicate an association, not a cause and effect relationship. In contrast, sequential analysis can identify specific child behaviors that most often immediately precede and follow parent and health care provider behaviors. Briefly, to address such questions the investigator defines windows of time around the first (given) behavior and then asks whether onsets of the second (target) behavior are more likely within such windows than not. An odds ratio is used to describe such relations for individual dyads, and a log odds ratio is used for subsequent analyses [23, 24]. Separate sequential analyses are be conducted for different adults (e.g., parents vs. health care providers) and for different phases (preoperative, induction) of the procedure. Given bi-directional effects between individuals during interactions, identifying the immediate behavioral precedents and antecedents of specific adult behaviors provides a more complete understanding of when and why an adult may engage in a potentially negative behavior and allows for more targeted intervention efforts with children parents, and health-care providers. The sample size of such a future study needs to be sufficiently high, as moderating variables such as age of the child are likely to affect the response of a child to any particular adult behavior (Figure 1).

Finally, some methodological limitations of the P-CAMPIS and the approach described in this manuscript have to be discussed. We recognize that some parent-child interactions cannot be fully captured by the P-CAMPIS. However, based on the body of work that was done over the past decade by Blount and others, as well the successful intervention programs developed for procedures such as VCUG, we are confident that the parent-child interactions that are captured by the newly developed P-CAMPIS coding system are powerful enough to enable us to effectively modify children's distress during induction of anesthesia.

In conclusion, we have provided initial validation for a new scale that is directed at measuring behavioral interactions that occur during the perioperative period between children, their parents and health care providers. We suggest that this scale should be utilized to identify specific parental and health care behaviors that lead to increased or decreased children's anxiety. Once such causal sequential analyses are conducted and these behaviors identified, empirically based, cost-effective, PPIA preparation programs can be developed.

References:

1. Hannallah, R. S. and Rosales, J. K. (1983) Experience with parents' presence during anaesthesia induction in children. *Can. Anaesth. Soc. J.*, 30, 286-9
2. Schulman, J. L., Foley, J. M., Vernon, D. T. and Allan, D.(1967) A study of the effect of the mother's presence during anesthesia induction. *Pediatrics*, **39**, 111-114
3. Bevan, J. C.; Johnston, C.; Haig, M.J.; Tousignant, G.; Lucy, S.; Kirnon, V.; Assimes, I.K.; Carranza, R. (1990) Preoperative parental anxiety predicts behavioral and emotional responses to induction of anesthesia in children. *Can. J. Anaesth*., **37**, 177-182
4. Kain, Z. N.; Mayes, L. C.; Caramico, L.A.; Silver, D., Spieker, M., Nygren, M. M., Anderson, G. and Rimar, S.(1996) Parental presence during induction of anesthesia. A randomized controlled trial. *Anesthesiology*, **84**, 1060-7
5. Kain, Z., Mayes, L., Wang, S., Caramico, L. and Hofstadter, M.(1998) Parental presence during induction of anesthesia vs. Sedative premedication: Which intervention is more effective? *Anesthesiology*, **89**, 1147-56.
6. Kain, Z., Mayes, L., Wang, S., Krivutza, D. and Hofstadter, M.(2000) Parental presence and a sedative premedicant for children undergoing surgery: A hierarchical study. *Anesthesiology*, **92**, 939-946.
7. Hickmott, K. C., Shaw, E. A., Goodyer, I. and Baker, R. D.(1989) Anaesthetic induction in children: The effect of maternal presence on mood and subsequent behaviour. *Eur. J. Anaesthesiol*., **6**, 145-155
8. Blount, R., Corbin, S., Sturges, J., Wolfe, V., Prater, J. and James, L.(1989) The relationship between adults behavior and child coping and distress during bma/lp procedures: A sequential analysis. *Behavior Therapy*, **20**, 585-601.
9. Blount, R., Cohen, L., Frank, N., Bachanas, P., Smith, A., Manimala, M. and Pate, J.(1997) The child-adult medical procedure interaction scale-revised: An assessment of validity. *J. Pediatr. Psychol*., **22**, 73-88.
10. Kain, Z., Mayes, L., Cicchetti, D., Bagnall, A., Finley, J. and Hofstadter, M.(1997) The Yale preoperative anxiety scale: How does it compare to a gold standard? *Anesth. Analg*., **85**, 783-788.
11. Cicchetti, D. V., Volkmar, F., Sparrow, S. S., Cohen, D., Fermanian, J. and Rourke, B. P.(1992) Assessing the reliability of clinical scales when the data have both nominal and ordinal features: Proposed guidelines for neuropsychological assessments. *J. Clin. Exp. Neuropsychol*., **14**, 673-86.
12. Blount, R., Bunke, V. and Zaff, J.(2000) Bridging the gap between explicative and treatment research: A model and practical implications. *J. Clin. Psychology in Medical Settings*, **7**, 79-90.
13. Varni, J., Blount, R., Waldron, S. and Smith, A.(1995) Management of pain and distress. *Handbook of pediatric psychology*, pp. 105-123.
14. Dahlquist, L., Armstrong, F., DeLawyer, L., Greene, P. and Wuori, D.(1986) Preparing children for medical examinations: The importance of previous medical experience. *Health Psychol*., **5**, 249-259.
15. Frank, N., Blount, R., Smith, A., Manimala, M. and Martin, J.(1995) Parent and staff behavior, and previous child medical experience, and maternal anxiety as they relate to child distress and coping. *Journal of Pediatric Psychology*, **20**, 277-289.
16. Blount, R., Davis, N., Powers, S. and Roberts, M.(1991) The influence of environmental factors and coping style on children's coping and distress. *Clinical Psychology Review*, **11**, 93-116.
17. Pate, J., Blount, R., Cohen, L. and Smith, A. (1996) Childhood medical experience and temperament as predictors of adult functioning in medical situations. *Children's Health Care*, **25**, 281-298.
18. Frank, N. C., Blount, R. L., Smith, A. J., Manimala, M. R. and Martin, J. K.(1995) Parent and staff behavior, previous child medical experience, and maternal anxiety as they relate to child procedural distress and coping. *Journal of Pediatric Psychology*, **20**, 277-289.
19. Zelikovsky, N., Rodrigue, J. and Gidycz, C.(2000) Cognitive-behavioral and behavioral interventions help young children cope during a voiding cystourethrogram. *J. Pediatr. Psychol*., **25**, 535-543.
20. Cohen, L., Blount, R. and Panopoulos, G.(1997) Nurse coaching and cartoon distraction: An effective and practical intervention to reduce child, parent and nurse distress during immunizations. *J. Pediatr. Psychol*., **22**, 355-370.
21. Blount, R., Bachanas, P., Powers, S., Cotter, M., Franklin, A., Chaplin, W., Mayfield, J., Henderson, M. and SD., B.(1992) Training children to cope and parents to coach them during routine immunizations: Effects on child, parent and staff behaviors. *Behavior Therapy*, **23**, 689-705.
22. Powers, S., Blount, R., Bachanas, P., Cotter, M. and Swan, S.(1993) Helping preschool leukemia patients

and their parents cope during injections. *J. Pediatr. Psychol.*, **18**, 681-95.

23. Bakeman, R., Deckner, D. and Quera, V.(2004) Analysis of behavioral streams. *Handbook of research methods in developmental psychology*. Teti D. (Ed.), Oxford, UK, Blackwell Publishers.
24. Bakeman, R. and Quera, V. (1995) *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. York: Cambridge University Press.

Testing for Mate Preference and Mate Choice in *Drosophila pseudoobscrua*

Y.-K. Kim[1], W.W. Anderson[1], P.A. Gowaty[2]
[1] *Department of Genetics, University of Georgia, Athens, U. S. A.*
[2] *Institute of Ecology, University of Georgia, Athens, U. S. A.*

Abstract
Natural selection is expected to adjust behavior to increase fitness. We tested this hypothesis with *Drosophila pseudoobscura* while observing the relationship between mate preference and mate choice. Mate preference was determined in *Amsterdam* arenas where a single fly was allowed to choose between two flies of the opposite sex. The time the chooser flies spent in front of partners, A and B, respectively, was measured. Our offspring viability studies have demonstrated that freely expressed mate preference affected the fitness of offspring. Matings of females with non-preferred partners showed compensation for significantly lower viability from these matings by an increase in the number of eggs or sperm. Preferred partners, regardless of sexes, were significantly bigger in body size than NP partners. In attempts to investigate correlations between mate preference and mate choice, the same fruit flies collected from the mate preference tests were observed in Elens-Wattiaux mating chambers where behavioral interactions between sexes and individuals were possible. Current data show that 1) females discriminated between P and NP males, mating significantly more with P males; 2) males, however, randomly mated with P and NP females; 3) both sexes mated significantly more with first courting or encountered partners; and 4) the presence of another individual during copulation significantly reduced copulation duration between mates, and so fewer sperm were delivered to females. These results suggest that under ecological and social constraints, individual *Drosophila* males and females vary behavior facultatively.

Keywords
Mate preference, mate choice, *Drosophila*, natural selection

Introduction
Mate choice is defined as any behavioral pattern displayed by members of one sex that leads to the greater likelihood of their mating with certain members of the opposite sex [2]. When choosing a mate, an animal utilizes various estimates of potential mate quality in terms of short-term benefits or long-term benefits or both. Picking a mate with high quality increases the chooser's fitness, and so such selectivity should be favored by natural selection [3,9,10]. Partridge [5] demonstrated that offspring from *Drosophila melanogaster* females raised in populations have significantly higher viability than offspring from females randomly paired with males, and these results show that mate choice affects offspring viability. However, interactions between or within the sexes could account for the observed differences. As part of collaborative studies on five different taxa, we controlled and eliminated reproductive competition such as male-male competition, female-female competition, female-male mate choice, and coercive interactions between the sexes. Then we tested whether *D. pseudoboscura* females or males mated to preferred partners had fitness benefits. Anderson *et al.* [1] have demonstrated that freely expressed mate choice was favored by variation in offspring viability, and for both female and male *D. pseudoobscura* mating with preferred partners resulted in significantly enhanced offspring viability. Subsequently, by estimating the number of sperm in the uterus, we measured fitness of males mated to preferred females and of males mated to non-preferred females. As observed in female fecundity data, NP males compensated for low offspring viability by delivering more sperm than P males. Both sexes preferred to mate with partners bigger in body size, but there was no significant difference in quantities of cuticular hydrocarbons between the two types of flies. In this study we investigated whether there was a correlation between free mate preference and mate choice in *D. pseudoobscura*. We observed the same flies in both the mate preference (Amsterdam) arenas and in Elens-Wattiaux mating chambers.

Methods

1) Mate Preference Tests
The arenas were made with tygon tubes, screens and Effpendorf tubes following Anderson *et al.* [1], as depicted in Figure 1. The distance between the two Eppendorf tubes where individual flies were placed was 5 cm. One side of each tube was taped to block visual contact and/or competition between the same sex during observation. The Eppendorf tubes were separated by a screen from the main tygon tube. Thus only visual, acoustic and olfactory contact between sexes could occur during observations. In the female preference test, we randomly picked two males from vials containing 7-day old virgins using an aspirator and placed each male in a tube. We put male A in the left side (position A) and male B in the right side (position B). We gently introduced a female of the same age through a slit in the middle of the arena. Female movements were observed with The Observer 5.0 [4] for 20 minutes per trial. We measured the time the female chooser spent in front of each male, T_A and T_B, until the 20 minute observation period ended. Our criteria on the preference test were as follows: 1) the choosing female had to spend at least 60% of a total trial time available for choice in the parts of the chamber near the males; and 2) the choosing female had to spend 60% of the total visiting time, $T_A + T_B$, with one of the two males. To determine repeatability, the same flies were individually stored in vials and tested again the next day. The positions of the males were switched in Test 2: male B in position A and male A in position B. If the female

chose the same male, the preferred and non-preferred male were determined on the basis of the time spent with each male. In the male preference test, one male was observed with two females following the same procedure.

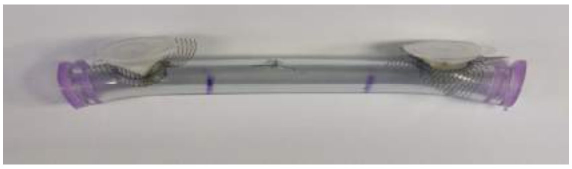


Figure 1. *An "Amsterdam" mate preference arena.*

2) Mating observation

After the preference tests, the same flies were placed in Elens-Wattiaux mating chambers. Flies of the same sex were simultaneously placed into the chambers using an aspirator. The chooser fly of the opposite sex was then entered. For identification, a wing of one of the two discriminatees was notched under CO_2. It has been shown that wing notching does not affect *Drosophila* courtship and mating. Observations continued for 20 minutes or until mating occurred. During observation, courtship latency and copulation duration were measured. We also scored which male first courted a female and the female response to the male courtship. After mating, we measured the length of the wings of the three test flies (see Figure 2).

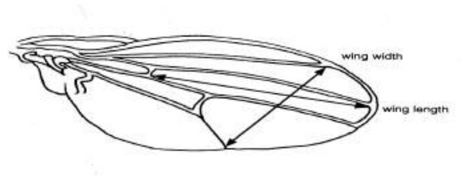


Figure 2. *Measurement of wing length in D. pseudoobscura.*

3) Observation of sperm in the uterus

After copulation, the female was removed from the vial and dissected in PBS [6,8]. The uterus, ventral receptacle and spermathecae were removed. The uterus was scissored away from the other two parts creating an opening through which sperm could be rolled out using dissecting pins and pin vises. The uterine tissue was removed and the sperm was left to be spread evenly throughout the 20 µL of PBS. One µL of sample was taken using a positive displacement pipette and placed on a microscope slide. The slide was dried at 60°C for 10 minutes. Then the sample was fixed in 3:1 methanol and acetic acid for 5 minutes and rinsed once in PBS. A DAPI/PBS (1:1000) stain was applied for 5 minutes and the sample was rinsed in PBS and allowed to air dry. After adding a drop of glycerol, a coverslip was placed over the sample and sealed with fingernail-polish. The slides were stored in dark until the sperm were counted under a fluorescent microscope/digital camera (Figure 3).

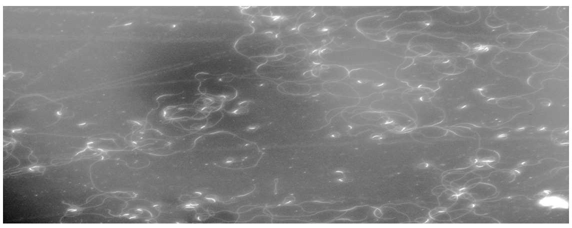


Figure 3. *DAPI-stained sperm from D. pseudoobscura*

Results

1) Behavioral observation

In both female and male choice tests, the repeatability, the frequency of flies that chose the same partners twice in the mate preference test, was 36.4% and 41.2%, respectively (Table 1). There was a significant difference in numbers of females mating with P males and with NP males that were previously collected in the preference test. Males, however, did not discriminate between P and NP females in the mating chambers. During courtship/mating observation in the chambers, both sexes mated significantly more often with first courting or encountered partners (Table 2). There was no significant difference in numbers of first courting males who were P and NP, and no significant difference between the numbers of P and NP among females encountered first.

Table 1. Results from D. pseudoobscura mate preference and mate choice tests.

Female Choice	**N**	**# females who chose the same ♂ twice in preference**	**# females mated with P♂ vs. NP♂**
	416	151 (36.4%)	80:56*
Male Choice	**N**	**# males who chose the same ♀ twice in preference**	**# males mated with P♀ vs. NP♀**
	342	141 (41.2%)	54:63

*p<0.05

Table 2. Mating discrimination between P and NP partners in Elens-Wattiaux mating chambers.

Female Choice	**N**	**# females mating with 1^{st} courting ♂ vs. 2^{nd} ♂**	**# 1^{st} courting ♂ who were P ♂ vs. NP ♂**
	116	90:26*	49:41
Male Choice	**N**	**# males mating with 1^{st} courted♀ vs. 2^{nd} ♀**	**#1^{st} courted ♀ who were P ♀ vs. NP ♂**
	75	57:18*	26:31

*p<0.001

2) Observation of sperm

After the preference tests, the mean numbers of sperm observed in the uterus after mating with P partners and NP partners, respectively, are shown in Table 3. In the female choice tests, females mated with P males had fewer sperm in their uterus than females mated with NP males (p=0.01). In the male choice situations, males that mated with P females delivered significantly fewer sperm than males that mated with NP females (p=0.0003). However, when mating occurred in the mating chambers, there were no significant differences in numbers of sperm between P and NP partners, which are shown in parentheses.

Table 3. Average numbers of sperm in the uterus of the females mated with P and NP partners (mean±S.E.).

Test	**N**	**P**	**NP**
Female Choice	20	2,060.95±237.81 (994.88±99.40)	3,077.60±297.41* (922.62±77.65)
Male Choice	20	1,648.15±131.76 (810.93±87.96)	2,498.05±166.70** (875.53±105.18)

* p<0.05; ** p<0.001

3) Observation of wing length

We measured wing length of P and NP partners collected from both mate preference tests and mate choice tests (see Table 4). In both female and male choice tests, there were significant differences in wing length: P flies have significantly longer wings than NP flies. However, there was no significant difference in wing length between the two types of flies actually mating in the chambers (shown in parentheses).

Table 4. Mean wing length (μM) of P and NP partners.

Test	N	P	NP
Female Choice	20	119.10±1.16 (74.54±0.44)	113.95±1.32* (74.54±0.37)
Male Choice	20	139.20±0.80 (81.04±0.45)	131.43±0.85** (81.54±0.51)

* p<0.01; ** p<0.001

Discussion

Behavioral interactions between the sexes significantly affected mate preference between the P and NP partners we observed. When mate choice was observed in the mating chambers, where interactions between the sexes and within each sex were possible, *Drosophila* females showed discrimination and mated more frequently with the P males that they chose in the preference tests. However, males were undiscriminating and mated randomly with individuals of the opposite sex. Males mated more frequently with the females encountered first, and females also mated more frequently with males to court them first.

Sperm transfer in matings with NP partners, was significantly higher than in matings with P partners. These results were consistent with the female's compensatory behavior we observed in the previous study [1]: females mated with NP males laid more eggs than females mated with P males. They appeared to compensate for reduced viability by increasing the number of eggs. Likewise, in the sperm observations, males mated with NP females delivered significantly more sperm than males mated with P females. Therefore, males also compensated for low viability by increasing the number of sperm transferred to females. However, when given a choice of mate in the chambers, competition and interference significantly affected the numbers of sperm transferred to females. Consequently, there was no difference in the mean numbers of sperm observed between matings with P and NP partners.

In the mate preference tests, both sexes preferred to choose larger partners. However, in the mate choice tests, there was no significant difference in body size between P and NP partners who actually mated with the chooser flies, regardless of the chooser sex. Overall, these results suggest that behavioral interactions between sexes significantly influence *Drosophila* mate preference.

We are grateful to A. Green, K. Kim, J. Laverentz, and D. Prosser for their technical assistance with the experiment. This project was supported by NIH Grant 5R01 GM048528-06 to WWA.

References:

1. Anderson, W.W., Gowaty, P.A. Kim, Y.-K. (2005). Mate choice and offspring viability in *Drosophila pseudoobscura*. (ms for *Proc. Natl. Acad. Sci.USA*).
2. Halliday, T.R. (1983). The study of mate choice. In Bateson, P. (ed.), *Mate choice*, Cambridge University Press, Cambridge, 3-32.
3. Hamilton, W.D., Zuk, M. (1982). Heritable true fitness and bright birds: a role for parasites? *Science* **18**, 384-387.
4. Noldus Information Technology (2003). The Observer 5.0. Noldus Information Technology, Wageningen, The Netherlands.
5. Partridge, L. (1980). Mate choice increases a component of offspring fitness in fruit flies. *Nature* **283**, 290-291.
6. Price, C.C., Dyer, K.A., Coyne, J.A. (1999). Sperm competition between *Drosophila* males involves both displacement and incapacitation. *Nature* **400**, 449-452.
7. Robertson, F.W., Reeve, E. (1952). Studies in quantitative inheritance. I. The effects of selection on wing and thorax length in *Drosophila melanogaster*. *J. Genet.* **50**, 414-448.
8. Snook, R.R., Markow, T.A., Karr, T.L. (1994). Functional nonequivalence of sperm in *Drosophila pseudoobscura*. *Proc. Natl. Acad. Sci.USA* **91**, 11222-11226.
9. Trivers, R.L. (1972). Parental investment and sexual selection. In Campbell, B. (ed.), *Sexual selection and the descent of man*, Aldine Press, Chicago, 136-179.
10. Zahavi, A. (1975). Mate selection: a selection for a handicap. *J. Theor. Biol.* **53**, 205-214.

Influence of breeding practices on 6- and 12-month-old human infants' attentional behavioral repertoire

G. Ortiz [1], C. Cañedo [2] and M.F. Gallardo[3]
[1] *Centro de Estudios e Investigaciones en Comportamiento, Universidad de Guadalajara, México*
[2] *Escuela de Psicología, ITESO, México*
[3] *Facultad de Psicología, Universidad de Guadalajara,* México

Abstract
Parents play an important role in development by regulating and organizing their infants' behavior through their own behavior [2]. Parents control the infant's attentional repertoires until the child is capable of naming and signaling objects and the repertoire becomes independent [3]. However, these attentional repertoires are often conceptualized as a matter of physiological maturation and not as something learned through mother (caregiver)-child interactions [5]. In order to identify the influence of breeding practices on the attentional repertoires in 6- and 12- month old infants, we videotaped mother-child interactions in a 30-min game situation and registered the frequency of the different behaviors that compose a repertoire of attention: a) attention getting (i.e. object showing, object touching, body movement, object manipulation), attentional responses (i.e. to direct sight to object, to direct sight where mother signals, to direct sight to another object), spatial proximity gradient (i.e. mother and child together touching the same or different objects) and breeding strategies (i.e. to give feedback as rewarding or praising, to model, to exemplify, to warn). The results show consistency between indications given by the mother and the child's attentional responses.

Keywords
Mother-child interaction, development, attentional behavior.

1 Introduction
During growth, parents play an essential role in the child's behavior development, because its through them that disorganized behavior starts to take shape and structure [2]. Parents are the first contact the child has with the world, and it is through them that it will know the environment that surrounds it. It has been suggested [5] that behavior repertoires are acquired during childhood, through the person in charge of breeding; due to the fact that this learning is done in an informal way.
In this way, although attention is commonly referred as a matter of physiological maturity and not as something that is learned through mother-child interactions, different authors agreed that at the beginning, it is the parents who control and direct the child's attention. Later, as soon as the child is capable of naming and pointing to objects, its attention will become independent from its parents, which means, voluntary [3].
The objective of the present project is to study the influence of breeding practices on the behavior repertoire of attention in children of 6 and 12 months old.

2 Method

2.1 Participants
For this study a sample of 3 dyads of mother-child was taken, with the child between 6 and 12 months of age. Pairs were as follows:
a) Mother- 6-month old girl
b) Mother- 12-month old boy
c) Mother- 12-month old boy

2.2 Materials
The materials used were:
- Video camera (Canon, model ES189A)
- Lap top (Compaq Pentium IV)
- Various toys
- The Observer Video-Pro (Noldus Information Technology)
- Theme 5.0 (Noldus Information Technology)

2.3 Behavioral catalogue
This catalogue was based on those proposed by Ribes & Quintana (2002).
1. Calling attention:
a. With verbalization
b. Without verbalization
For example: Show object, show two objects, indicate object, touch object, facial expression, instigate, body movement, manipulating an object.

2. Attention responses:
For example: Direct look to object, direct look at mother, direct look out of object and mother, direct look at other object, motor agitation, touch object, manipulate object, babbling.

3. Gradient of spatial proximity:
For example: (1) Mother and child together, touching each other or touching the object, looking at each other; (2) Mother and child together, touching each other or touching the object, one of them looking at the other; (3) Mother and child together touching each other or touching the object both looking at the same object or in the same direction.

4. Teaching strategies:
For example: Feedback, Adapting, Modeling, Illustrating, Notifying.

2.4 Procedure
A focal continuous method of observation was used [1]. Mother-child interaction was videotaped in a play situation, during which the frequency of different behaviors that compose the attention repertoire were registered. The recordings were done in the usual place

where the mother and child play, and had a 40 minute duration for the dyads “a” and “b” and 25 for the dyad “c”; in all cases, the first ten minutes were only for the habituation of the participants to the videotaping condition.

Figure 1. *Frequency of observed behaviors for child and mother.*

HIDDEN INTERACTION PATTERNS

Dyad "a"

→ Mother, manip obj v / mother, show obj v
→ Child, looking at the obj / motor agitation

Dyad "b"

→ Mother, manip obj / manip obj v
→ Child, looking at the object.

→ Mother, instigate v / adaptate
→ Child, babbling

→ Mother, facial expression v
→ Child, looking the mother

→ Mother, show obj v
→ Child, looking other obj / looking the obj.

→ Mother, show obj.
→ Child, looking the object

Dyad "c"

→ Mother, show the obj / manip obj / touch the obj.
→ Child, looking the obj

Figure 2. *Hidden patterns most frequently registered in all dyads.*

2.5 Data analysis

Data obtained by observing the categories described before, were analyzed using Theme 5.0 in order to find the hidden patterns of mother-infant interactions. Sequences of different types of events consistently repeated along a data sample are recognized as patterns by this program.

3 Results and conclusions

Figure 1 shows the frequency of observed behaviors for child and mother of all dyads. In general, we observed that the children executed a greater amount of behaviors than the mothers, although the number of registered categories was similar for both sides of the dyads. The category with the highest frequency in the children was look at an object; while manipulate the object was the highest frequency category in the three mothers, two of which were joined with verbalization.

The hidden patterns of behavior that were present most frequently in the mother-child relationship are shown in Figure 2. In dyads "a" and "c" there was only one relevant pattern while in dyad "b" there were several relevant patterns (n=5). In almost all cases, the mother's behavior was related to the manipulation, showing or touching the object, while the child's behavior was related to directing the look to the object and to some kind of body expression (i.e. babbling, facial expression, motor agitation).

The children showed the expected attentional behavioral repertoire for the age. At the same time, the interaction patterns between mother and child were different in each dyad, although the number of registered behaviors was similar. The frequencies of registered behaviors in children were higher than those of the mother. From the hidden pattern analysis we can observe that the children's behaviors were initiated by the mother's action.

References:

1. Altmann, J. (1974). Observational study of behaviour. *Behaviour*, **49**, 227-267.
2. Quintana, C. (2002). Desarrollo de la conducta durante la primera infancia. In: *Psicología del aprendizaje*. Ribes-Iñesta, E. (Ed.). México: Manual Moderno, pp. 113-129.
3. Rosselló, J. (1997). *Psicología de la Atención. Introducción al estudio del mecanismo atencional.* Pirámide: Madrid.
4. Ribes, E.; Quintana, C. (2002). Mother-child linguistic interactions and behavioral development: a multidimensional observational. *Behavior Analyst Today*, **3**(4), 442-454.
5. Staats, A. (1983). *Aprendizaje, lenguaje y cognición*. Trillas: México.

Visitors as stimuli for the expression and distribution of behavioral patterns in captive vervet monkeys (*Cercopithecus aethiops pygerythrus*): a pilot study

G. Ortiz[1], E. Cañedo[2] and C. Rosario[2]
[1] *Universidad de Guadalajara, México*
[2] *ITESO, México*

Abstract

The presence or absence of observing visitors might be meaningful stimuli for the expression of behavioral patterns and differential use of enclosure space in captive vervet monkeys (*Cercopithecus aethiops pygerythrus*). In order to assess this hypothesis, five vervet monkeys were videotaped 60 minutes per day, during nine days (two days in December and seven days in February) in which the number of visitors, the weather and the group composition varied. These observations were made at the zoo of Guadalajara (Mexico). The behavior patterns (i.e. Resting, Locomotion, Grooming others, Agonist behavior, Self cleaning, Playing) and locations of adult monkeys inside their enclosure were scanned every 30 s. The results show differential patterns of behavior and use of space as a function of the aforementioned variables (i.e. presence–absence of observing visitors, weather, and group composition).

Keywords

Primates, behavioral patterns, spatial distribution, captivity, *Cercophitecus aethiops*.

1 Introduction

It is evident that for animals of species that remain in captivity, these conditions have drastically changed compared to what might be found in their natural environment, as in the case of primates kept in zoos. The usual function of a zoo in a city is recreation and tourism, the decision for the structuring of enclosures is mainly given by matters such as zoo image or accessibility to observing the animals by the spectators and not necessarily the behavioral or habitat characteristics of each animal. The zoo's recreational function does not represent an obstacle for research to be made in it. On the contrary, it facilitates the study of many aspects that in the wild would not be possible to observe. At the same time, it would be possible to improve the environment in which the animals live, with the corresponding improvement in time and quality of the animal's life span (e.g. [7,10]).

In general, the research made in some zoos has mainly been made in the veterinary and biological areas [2,8], and the study of animal behavior has not received the attention that it should get [4]. An aspect that has to be studied is the behavior of animals interacting with theitr environment, created and modified by humans, conceptualizing the organism as an entity that behaves and not only as a receptor of its new environment [6]. For example, the influence of certain visual and auditory stimuli, and of the interaction of monkeys with their new habitat inside a zoo, coincides with one of the topics that has been of highest interest for behavior analysis: the identification of certain stimuli that control the behavior of organisms. These stimuli are part of the organism's environment and could be proximal signals that control mainly terminal behavior, like taking and ingesting some meals and copulation; while, other stimuli are distal signals, that control diverse behaviors like the direction of movement, and other relevant behaviors for their spatial location (e.g. [9,13]). It is also known that certain stimuli work as unleashing signals of instinctive behavior and allow a deployment of behavior characteristic to the species, and other signals that are learned by organisms given their interaction with their current environment (e.g. [5,11,16,17]).

A main characteristic of zoos is that spectators are part of the interactive environment of the animals, and the presence–absence of these spectators can be a stimulus that modifies the behavior of animals in their environment. In the present study, a systematic observation of displayed behaviors by vervet monkeys (*Cercopithecus aethiops pygerythrus*) was made in order to register the influence that spectators have as potential stimuli on temporality and spatiality of the behavioral catalogue in their captive condition.

2 Method

2.1 Subjects and environment

The observations were made on a captive group (n=5) of vervet monkeys (*Cercopithecus aethiops pygerythrus*), belonging to the Guadalajara Zoo, located in the city of Guadalajara, Jalisco, México. The group was formed by an adult male, an adult female, two juveniles and an infant. During the day, these primates were located in an enclosure area about 5 m long, 4 m wide and 2 m high, with a down slope of 3 degrees toward the front of the enclosure. The area was limited by three concrete walls and a metal grill fence with short bushes about 50 cm high on the spectators´ side, as well as a top closing area made by the same material as the fence in the front.

In Figure 1, a schematic representation of the enclosure is displayed. Inside, a "Y" shaped log rests on the floor and against the wall on the far end of the enclosure (A). The floor on the rear part of the enclosure (1 m width) is made of concrete; the rest of the enclosure's floor is made of compacted dirt. On this dirt part at about 1 m from the fence, there is a concrete "island" on the right-hand side for the visitors (I), as well as several rocks. At the back of the enclosure in the central area, the access to the animals' quarters can be found.

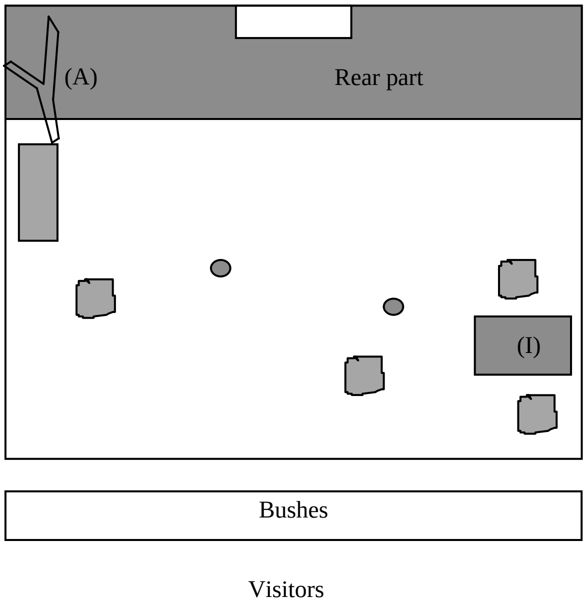


Figure 8. *Schematic representation of the enclosure.*

2.2 Materials

A Sony mini DV TRV-18 camcorder was used to record the behavior of the subjects during an hour each day over the observational span. For the registration of behaviors, The Observer Video-Pro (Noldus Information Technology) was installed and used on two laptop computers, a Toshiba with a Pentium 4 processor and a Sony with the same characteristics.

2.3 Procedure

The behavioral pattern and spatial use of the male and female in the enclosure were registered during the recording time (Monday through Sunday from 15:00 to 16:00 hours). To register the behavior of the subjects and their location, a focal animal observation method, with a 30 s scan was used [1]. The enclosure was divided in 6 sections, resulting from the combination of sectors near (up to 1 m from the fence) and far, as well as the right, centre and left sectors: the near left (NL), centre (NC), and right (NR), far left (FL), centre (FC), and right (FR). A schematic representation of the division made in the enclosure is shown in Figure 2.

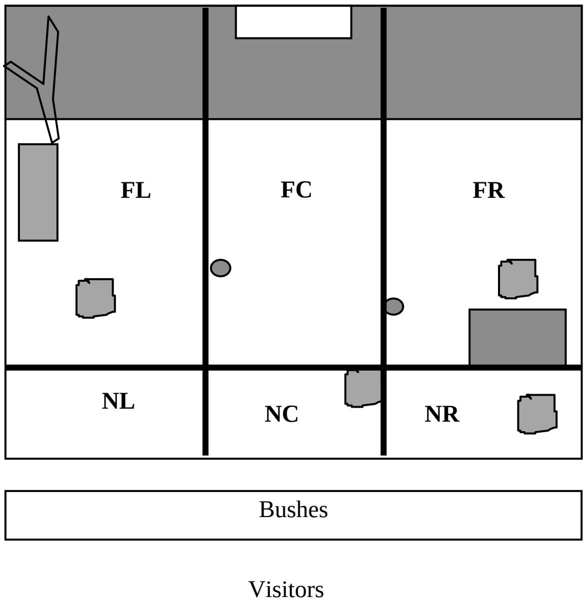


Figure 2. *A schematic representation of the division made in the enclosure.*

On days with a high affluence of visitors (i.e. Saturday and Sunday), the recording was made with the camera in hand due to the need of movement caused by the interference of spectators in the visual field of the camera, trying to keep each subject in this field.

The days with less affluence of visitors (Monday and Tuesday, days on which the zoo is closed, and Wednesday, Thursday and Friday) the camera was placed on a tripod, trying to keep as much of the space in the enclosure in focus as possible. At this time, the observers moved out of the visual field of the subjects by placing themselves at about 10 m to the right of the enclosure to avoid causing an influence on the behavior of the observed subjects. Due to this, an *ad libitum* registration of the subjects´ behavior could not be made.

2.4 Definition of behavioral catalogue

The behavioral patterns considered relevant to the study were defined based on the behavioral catalogue presented by Carrera (1994): Locomotion, resting, agonistic behavior, grooming, playing, and grooming others.

3 Results

In Figure 3, the behavioral patterns (upper graphics) and location (lower graphics) for the male and female are shown, with visitors and without visitors. Independently of the presence/absence of visitors, the most common behavioral pattern was resting, with a significant decrease in the female without visitors. The second most important pattern for the female was grooming others and locomotion; while the male showed a pattern of locomotion. On the other hand, we observed a preference for the distant locations (FL and FR) in the female in the presence of visitors, and a change in the distribution on days without visitors: an almost equal pattern of distribution in all areas. In the male, we observed a preference for the near locations in the presence of visitors, while on the days without visitors his pattern of distribution became almost equal (even more than in the female) in all areas.

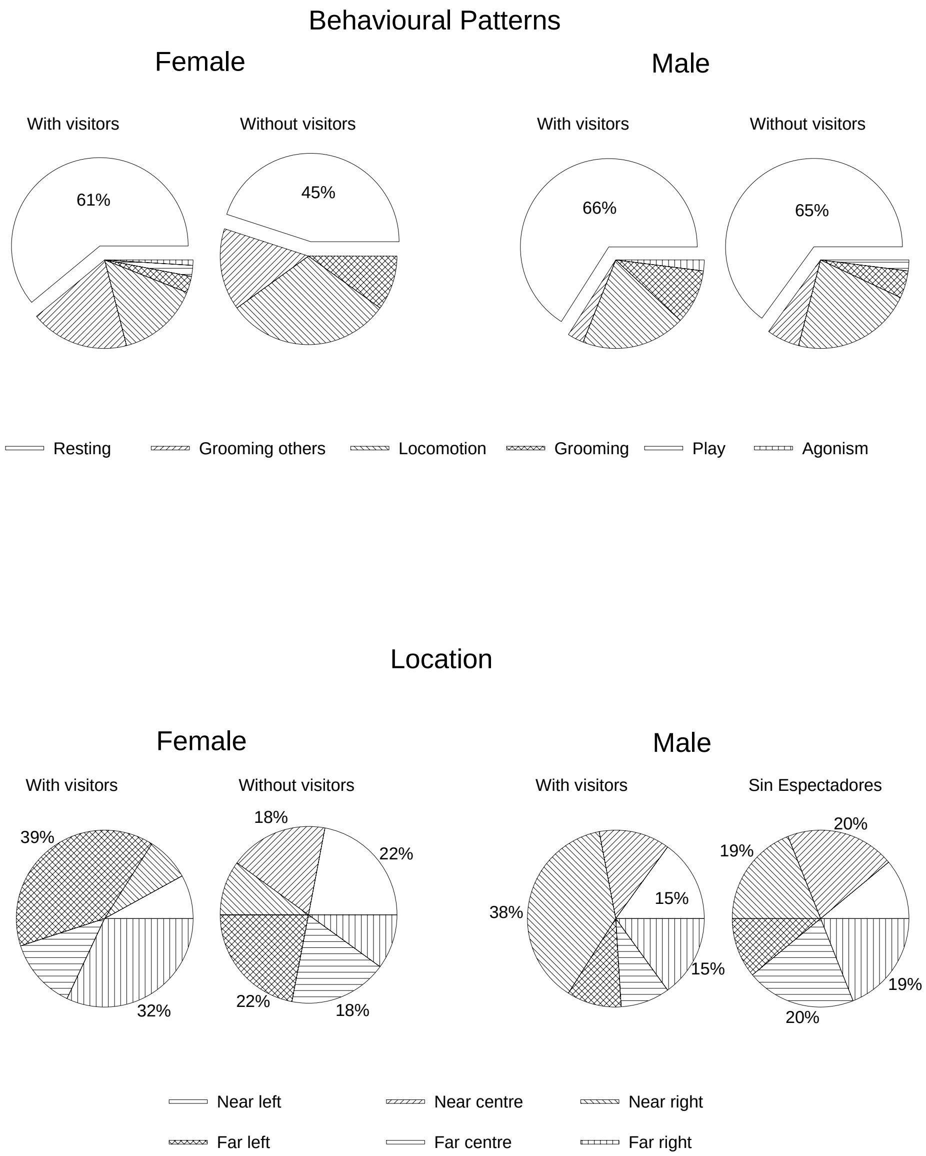


Figure 3. *Behavioral patterns (upper graphics) and location (lower graphics) for the male and female, with visitors and without visitors.*

4 Discussion

As it could be observed, there is an existing change in both the expressed behaviors by the subjects in the study and the spatial location inside the enclosure. This seems to indicate that visitors have a relevant stimulus function for captive vervet monkeys. One of the possible reasons is that spectators are usually associated with food grants, as it was observed in an anecdotic way during the recording of the observations. This was also observed in a contiguous enclosure, in which another species of monkeys (i.e. Capuccino monkey) were kept, this type of behavior was observed at the time the monkeys stretched out their hand to receive food from the visitors.

On the days the zoo was open to the public, the use of near spaces was practically null for the female. She was usually located on the log that lies against the wall in the back of the enclosure, or at sites that kept her at a large enough distance from visitors. At the same time, a preference for the log was observed during the days without spectators. This could be explained by studies made in natural habitats in which it is mentioned that this species tends to be semi-terrestrial although they most likely accomplish their activities on tree branches (e.g. [12]).

Another aspect to consider is the age of the infant, and the activities displayed by each of the members of the group, as well as the use of space. As the infant grows the male tended to distance himself from the visitors, it could be related to protection and surveillance patterns of some kind displayed by the male [14,15]. A situation to be considered in further studies, in which behavioral changes could be more specifically identified (i.e. pattern expression and spatial distribution concerns) produced by the birth of an infant, which implicates the introduction of a new member to the group and could imply a reduction of space and access to food.

Finally, it is important to comment on the fact that the highest activity frequency is centered on resting; in the female it may not be so relevant, because of the infant's age, and this could considerably limit her activities even in natural conditions. It would be important to carry on with more comparative studies between the behavioral patterns expressed in captive and natural conditions, as well as the relevant stimulation for such patterns, to be able to identify, define, and in a given moment, modify the conditions of the enclosure in which not only this species of primates but all the captive animals live in.

References:

1. Altmann, J. (1974). Observational study of behavior, Behaviour, **49**, 227-267.
2. Caba-Vinagre, M.S.; López-Domínguez, R.; Silva-López, G. (1993). Estudio comparativo de primates en cautiverio en México. In: Estrada, A.;. Rodriguez-Luna, E.; López-Wilchis, R. Coates-Estrada, R.. (Eds.). *Estudios Primatológicos en México*. Vol 1. Veracruz: Biblioteca Universidad Veracruzana, pp. 201-213.
3. Carrera-Sánchez, E. (1994). Desripción del comportamiento de un grupo de mono aullador. *La Ciencia y el Hombre*, **18**, 127-149.
4. Crockett, C.M. (1996). Data collection in the zoo setting, emphasizing behavior. In: Kleiman, D.G.; Allen, M.E.; Thompson, K.V. and Lumpkin, S. (Eds.). *Wild Mammals in Captivity. Principles and Techniques*, Chicago: The University of Chicago Press, pp. 545-565.
5. Eibl-Eibesfeldt, I. (1979). *Etología. Introducción al estudio comparado del comportamiento*. Barcelona: Omega.
6. Gibson, J.J. (1986). *The ecological approach to visual perception*. New Jersey: Lawrence Erlbaum Associates, Publishers.
7. Guerrero, D. (1997). Enrichment 101: A basic overview, http://www.arkanimals.com/E/Enrich101.html
8. Guichard-Romero, C.A. (1994). Consideraciones acerca de los programas de conservación en los Zoológicos: El caso de los primates. *La Ciencia y el Hombre*, **18**, 31-35.
9. Honig, W.K. (1987). Local cues and distal arrays in the control of spatial behavior. In: Ellen, P. and Thinus-Blanc, C. (Eds.). *Cognitive processes and spatial orientation in animal and man*, Lancaster: Martinus Nijhoff Publishers, pp. 73-88.
10. Krause, M.A.; Gordon, M.A.; Burghardt, M.; Lentini, A. (1999). Object provisioning for nile soft-shelled turtles (*Trionyx triunguis*), *Lab Animal*, **28**(7).
11. Lorenz, K. (1979). *Evolution et modification du comportement*, Paris: Petite Bibliothèque Payot.
12. Melnick, D.J.; Pearl, M.C. (1986). Cercopithecines in multimale groups: Genetic diversity and population structure. In: Smuts, B.B.; Cheeney, D.L.; Seyfarth, R.M.; Wrangham, R.W.; Struhsaker, T.T. (Eds.). *Primate Societies*, Chicago: University of Chicago Press, pp. 121-134.
13. Menzel, E. W. Jr. (1987). Behavior as a locationist views it. In: Ellen, P.; Thinus-Blanc, C. (Eds.). *Cognitive processes and spatial orientation in animal and man*, Lancaster: Martinus Nijhoff Publishers, pp. 55-72.
14. Ortiz, G.; Cabrera, F. (2002). Papel de los espectadores en la expresión y distribución espacial de patrones conductuales en monos vervet en cautiverio: un estudio piloto. *Enseñanza e Investigación en Psicología*, **7**(2), 375-392.
15. Ortiz, G.; Cabrera, F. (2003). Los espectadores como estímulo para la expresión y distribución espacial de patrones conductuales en monos vervet (*Cercopithecus aethiops pygerythrus*) en cautiverio: un estudio piloto. In: Carvajal, S. (Ed.). Avances en la Investigación Científica en el CUCBA- XIV Semana de la Investigación Científica, Guadalajara, México: Universidad de Guadalajara, pp. 217-218.
16. Tinbergen, N. (1975). *Estudios de Etología I*. Madrid: Alianza Editorial.
17. Tinbergen, N. (1981). *El Estudio del Instinto*. México: Siglo XXI.

An isolation protocol for testing anti-depressants and anxiolytics in the marmoset monkey

M.J. Jongsma [1], R.A.P. Vanwersch [1], S.A.M. van Vliet [1,2], I.H.C.H.M. Philippens [1, 2]
[1] *Dept. Diagnosis and Therapy, TNO Defence, Security and Safety, Rijswijk, The Netherlands*
[2] *Dept.of Psychopharmocology, Utrecht University, Utrecht, The Netherlands*

Abstract

An isolation protocol in a nonhuman primate has been developed in which a 'natural' expression of fear and depressive behavior is present. Separation from the social group is stressful for nonhuman primates, elevating anxiety- and depressive-like behaviors.
Different readout-systems were used resembling depressive-like behavior in human: observational recordings of anxiety- and depressive-like behavior, human threat test (HTT), startle reflex, body weight and activity.
In the new protocol animals are tested during normal housing, during two weeks of isolation, and after re-socialization. There was a correlation between the observational and quantitative parameters confirming the anxiety related behavior during the isolation. After replacing the animals in the colony, most parameters recovered. Only depressive-like behavior was still present.

Keywords

Marmoset monkey, behavior, isolation, anxiety, depression.

1 Introduction

Testing anti-depressants or anxiolytics remains dependent to a great extent on the use of proper animal models of anxiety and depression. An ideal model would be one reproducing all features of human anxiety and/or depression. Therefore, an isolation protocol in a nonhuman primate has been developed in which a 'natural' expression of fear and depressive-like behaviour is present.
Separation from the social group is stressful for nonhuman primates, elevating anxiety- and depressive-like behaviors. It seems that patients suffering from depression elevated cerebrospinal fluid corticotrophin-releasing factor (CRF). Infusion of intracerebroventricular CRF did not affect this typical anxiety and depressive-like behavior in the separated monkeys in contrast to the social housed monkeys, indicating the natural development of depressive-like behavior [4].
To test the behavioral effects of anxiety and depression in this animal model, different readout-systems were used resembling depressive-like behavior in human: observational recordings of anxiety-like behavior and depressive-like behavior, HTT, startle reflex, body weight, and activity. These tests and observations are performed during normal housing, during two weeks of isolation, and during normal housing after the separation.

2 Materials and methods

2.1 Animals

Eight adult male and female common marmosets *(Callithrix jacchus)*, aged 2-4 years with body weights of 300-450 g were obtained from BPRC, The Netherlands. Normally the animals are housed together in one room. During the isolation period, they were kept in separate rooms. The ambient temperature was regulated at 25 ± 2 º C. The relative humidity was >60%. A 12-h light-dark cycle was maintained. Animal care is described in Standard Operating Procedures in agreement with the European Community guidelines. An independent committee on Animal Care and Use approved this study.

2.2. Behavioral observation

The used scoring list is designed to record in detail the occurrence symptoms correlated to central effects and the condition of the animal. The following symptoms were taken into account: 1) Lack of appetite, 2) inadequacy of grooming, 3) apathy (i.e. the responsiveness of the animal to its surroundings), 4) mobility, 5) stereotypy (any repetitive, patterned and rhythmic movement), 6) stypic behavior (idiosyncratic non-locomotive stereotyped actions), 7) passive (maintenance of a non-locomotive behavior with no simultaneously scoreable behaviors except for self-directed behaviors), 8) huddle (fetal-like self-enclosed position with head lower than shoulder level), 9) back of cage (location of the animal in the cage).
Some parameters were grouped in anxiety-like signs: stereotypy and stypic behavior, and in depressive-like signs: mobility, passive, apathy, and huddle. The rates of severity were coded from 0 to 4 (0: normal, 1: extreme normal, 2: mild, 3: moderate, 4: severe). Besides these items the bodyweight was registered.

2.3 Human Threat Test

For the marmoset monkey the effects of human interaction can be used to detect the effects of anxiolytic drugs [1]. The responses to a human observer were observed over 2 min. During this time, the human observer stands in close proximity to the marmoset's home cage. The observer is the only human in the room to prevent distraction. The behavior was registered by video monitoring.
The following parameters were recorded: 1) time spent in front of the cage, 2) time spent in nest box, 3) number of forward jumps, 4) body postures: number of tail postures (i.e. genital presenting), scent making, arched pilo, 5) Fear postures: slit stares (flattening of the ear tufts and staring at the threat through eye lids reduced to slits), rearing, twisting.

2.4 Spontaneous exploratory behavior

The levels of activity, alertness, and exploratory behavior play an important role in practically all measurements of animal behavior. To assess automatically and quantitatively the combined loco-motor and exploratory activity of the marmoset the so-called 'bungalow' task is used [5].
The apparatus consists of four horizontally placed non-transparent PVC boxes (23 x 23 x 23 cm) with a mesh-wire top, interconnected by photo-cell-guarded PVC tubes. Four lights are mounted on the closed ceiling of the apparatus. A curtain surrounds the whole apparatus to avoid distraction of the animal. The animals could freely move and change from one compartment to another during a 20-minute session.
A video tracking system (EthoVision, Noldus) was used to register the movements of each animal. The loco-motor activity is expressed as the number of compartment changes.

2.5 Auditory startle reflex

In the auditory startle reflex the stretching movement of the legs is used to reflect the reaction of the animal on a startle signal [2]. Animals are exposed to 20 auditory startle pulses (120 dB, pink noise, 20 ms) while standing on a platform in a PVC cage. The startle response of 200 ms duration is measured by a transducer connected with the platform, registering the force exerted by the animal upon presentation of the stimulus. The area under the curve (AUC) and amplitude and of the startle response are registered and used to express the motor reaction of the startle reflex [6].

2.6 Study design

Baseline values from all read-out systems were collected during normal housing. Thereafter, all eight animals were isolated in separate rooms to induce depressive-like behavior [5]. During two weeks of isolation the same parameters were checked, like the daily observational scoring and bodyweight, and the weekly Human threat test and Bungalow test. After these two weeks the animals were re-socialized in the common marmoset facility. For the period of one week the animals were re-tested.

2.7 Methods for statistical analysis

The results are presented as mean ± SEM. An ANOVA analysis was applied followed by a Newman- Keuls test where appropriate and a t-test with Welch's correction. In all tests p values < 0.05 were considered significant.

3 Results

During isolation, most parameters were significantly affected (see Table 1, Fig. 1 and 2). There was a correlation between the different parameters (observational and quantitative parameters) confirming the anxiety related behavior.

After resocialization, most parameters recovered to normal. Only the depressive-like behavior was still present.

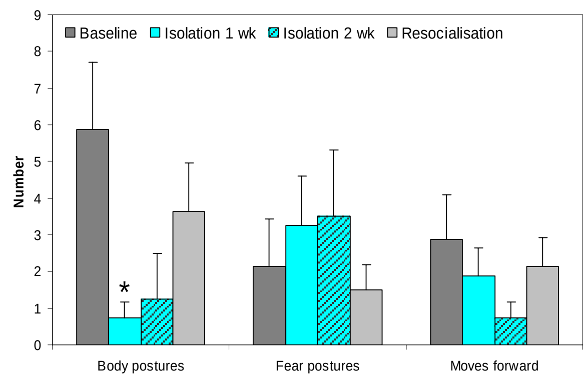


Figure 1. *Human Threat Test: mean (± SEM, n=8/group)*

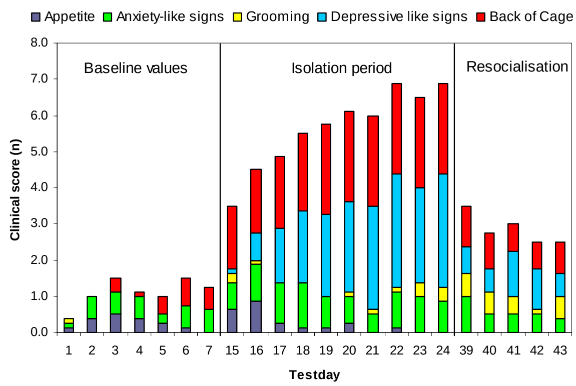


Figure 2. *Observational parameters (mean, n=8/group)*

	Baseline	Isolation	Resocialization
Startle reflex: AUC	3985 (587)	5277 (1030)	4515 (655)
Startle reflex: Amp	3.3 (0.8)	4.3 (1.4)	3.0 (0.7)
Time in nest box in sec	2.4 (1.8)	46.4 (21.6)	0.4 (0.4)
Time in front in sec	46 (12)	19 (2)	72 (17)
Body weight in g	371 (1)	357 (1)	369 (1)
Bungalow # shifts	68 (26)	38 (9)	56 (22)

Table 1. ***E****ffects of isolation on startle reflex, HTT, body weight and activity (mean (SEM) n=8/group).*

4 Discussion

During isolation a significant increase of depressive-like and anxiety-like signs were found (p < 0.001). These effects were found in all test systems used. The fear postures and the time spent in the nest boxes in the HTT increased, which is in accordance with the direction tendency on the startle response. Also the body weight decreased during isolation, although the activity -both in the home cage and in the bungalow task- decreased. Furthermore, the moves forward and time spent in the front of the cage both decreased as a result of the isolation. These parameters are correlated with anxiolytic behavior, indicating more fear.

Otherwise, the body postures almost disappeared in the HTT. Body postures are mainly territorial postures which are normal in the presence of other marmoset monkeys.

Remarkably, during resocialization, after the separation period, most parameters recovered to almost normal. Only depressive-like behavior was still present and grooming became more inadequate. Grooming is correlated to depressive-like behavior. CRF leads to excessive grooming behavior [4]. Indeed, after resocialization the CRF levels probably decreased in these animals, leading to rebound inadequate grooming effects.

In conclusion: The isolation protocol of these social animals, result in severe depressive and anxiety-like behavior. Therefore, this protocol can be used as a model for testing anti-depressants and anxiolytics in the marmoset.

References:

1. Carey, G.J.; Costall, B. (1992) Behavioural effects of anxiogenic agents in the common marmoset. *Pharmacol Biochem. and Behav*, **42,** 143-153
2. Davis, M.; (1992) Neurochemical modulation of Sensory-motor reactivity: Acoustic and tactile startle reflexes. *Neurosci. biobehav. rev*, **4** , 241-263
3. Dunn, A.J.; Berridge, C.W. (1990) Physiological and behavioural responses to corticotropin releasing factor administration: is CRF a mediator of anxiety or stress responses? *Brain Res. Rev*, **15**, 71-100
4. Dunn, A.J.; Lai, C.W. (1987) CFR-induced excessive grooming behavior in rats and mice. *Peptides*, **8**, 841-844
5. Philippens, I.H.C.H.M.; Melchers, B.P.C. (2000b) Behavioral test systems in marmoset monkeys. *Behavior Res. Methods, Instruments, and Computers*, **32**, 173-179
6. Strome, E.M.; Wheler, G.H.T. (2002) Intracerebroventricular corticotrophin-releasing factor increases limbic glucose metabolism and has social context-dependent behavioral effects in nonhuman primates. *PNAS*, **99**, 15749-15754

Nighttime care of infants by first-time adolescent and adult mothers

W.A. Wetherall [1], J. J. McKenna[2] and the Centers for the Prevention of Child Neglect
[1] *Center for Children and Families and the Department of Psychology, University of Notre Dame, Notre Dame, USA*
[2] *Department of Anthropology, University of Notre Dame, Notre Dame, USA*

Abstract
While anthropological inquiry on parenting behaviors has contributed a cross-cultural perspective on the diverse approaches to childrearing, fewer studies have focused on nighttime caretaking practices and their effects on child development. Even less is known about the nighttime parenting behaviors of adolescent mothers, including how specific patterns of behaviors may contribute to, or protect against, poor child outcomes.

The present study conducted at The University of Notre Dame Mother-Baby Behavioral Sleep Laboratory analyzed the nighttime parenting behavior and interactions of first-time adolescent and adult mothers and their infants. Using overnight infrared video recordings, questionnaires, and direct observations during in-home assessments, data were collected from a sample of 23 adolescent (mean 17.7 years) and 14 ethnically-matched adult (mean 25.2 years) mother-infant dyads when the infants were four months old. During overnight sleep studies participants were allowed to maintain normal nighttime caregiving routines and sleeping arrangements.

Using the Observer 5.0, overnight video recordings were coded for mother and infant sleep state, sleep location and position, infant feeding method and frequency, amount of physical proximity, affectionate care, crying duration and maternal response, sleep-related risks and other mother-infant interactions.

Group differences between adolescent and adult mothers and between solitary sleeping, bed-sharing, and room-sharing dyads in the form and frequency of particular behaviors, particularly affectionate care of infants and sleep-related risks were examined.

Implications for child development, sleep safety and design of effective parenting interventions are considered.

Keywords
Nighttime parenting, adolescent mothers, infant sleep, infant sleep safety, child development

Introduction
The inquiry into how first-time adolescent and adult mothers care for their infants at night is part of a larger, multi-site longitudinal parenting study funded by the National Institutes of Child Health and Development aimed at identifying the precursors to abuse and neglect and improving child outcomes.

In general little research on nighttime parenting behavior has been conducted although previous studies have shown that sensitive and responsive nighttime parenting may be as important as daytime care when it comes to the healthy physiological development of infants. [2, 3] Previous research on adolescent parenting has shown that children born to teen mothers are more likely to experience adverse developmental outcomes than children born to adult mothers. Such problems range from delayed or diminished cognitive and language development to socioemotional dysfunction. [4]

Overnight sleep studies, lasting on average 10-12 hours, offer researchers a unique and extended naturalistic observation of mother–infant interaction and give insight into how adolescent parenting styles differ from those of adult mothers. By combining this data with daytime parenting measures, in-home observations and on-site assessments, we are able to examine the overall quality and consistency of daytime and nighttime care of children.

Methods
Participants in the overnight sleep studies were drawn from a larger, multi-site, longitudinal study on the transition to parenting over the first three years. (N=684)

Data collection for the sleep lab occurred between July 2002 and May 2005. Sleep studies took place at two time points, once when the infants were 4 months old and again at 8 months. Mothers were instructed to arrive shortly before their infant's normal bedtime so that pre-sleep activities and bedtime routines could be observed. Mothers and were asked to maintain normal patterns of infant care and sleeping arrangements.

The Mother-Baby Behavioral Sleep Laboratory looks much like a small two-bedroom apartment and allows for a variety of different sleeping arrangements. In one bedroom there is a standard double bed for mothers or for mothers who bedshare with their infants. This room is also equipped with a bassinette for mothers who room-share. The adjacent bedroom has a crib for mothers who sleep separately from their infants. There is also a lounge area and private bathroom. Microphones and infrared cameras are located in the lounge and bedrooms and audio and video feeds were monitored continuously by a sleep lab technician. A digital mixing board allowed for split screen recording when the mother and infant were in separate rooms.

Completed overnight recordings were coded by pairs of researchers in real time using the Observer 5.0. The coding taxonomy consisted of 12 behavioral classes and 16 modifier classes. Our complex taxonomy allowed us to capture innumerable variations in mother-infant interactions. Behavioral classes included mother and infant sleep state, infant sleep location, infant sleep position, feeding method, crying duration and maternal response, proximity of infant to mother, affectionate care, and risks to the infant.

Additional information on actual practices and attitudes about nighttime care was collected using 7-day sleep diaries and short questionnaires. Multiple in-home and on-site assessments beginning prenatally and conducted at regular intervals over the course of three years provide a comprehensive picture of parenting behavior and child outcomes.

Results

Preliminary analysis of 4 month data reveals interesting patterns and group differences. Adult mothers were more likely to breastfeed and sleep separately from their infants while adolescent mothers were more likely to bottlefeed and bedshare. This finding is in contrast to previous sleep studies of adult mothers that have shown a strong connection between breastfeeding and bedsharing. [1]

Both teen and adult mothers tended to modify their infant's sleep environment by adding items such as blankets, pillows, and soft toys. These materials increased potential breathing risks to the infant at night and are generally not recommended for use at 4 months.

Since maternal sensitivity, warmth, and responsiveness are important to a child's social, emotional and cognitive development, particular attention was given to behavior that constituted affectionate care, including kissing, patting, rubbing, rocking, and tickling. Affectionate care was coded as an event lasting 10 seconds or less. Frequency of affectionate care was divided by the total potential time affectionate care could have occurred. For example, time mothers spent asleep was not included since they needed to be awake to engage in affectionate care of their infants. Overall rates of affectionate care were grouped in low and high categories based on a median split.

When comparing teen and adult mothers, neither group showed a significantly lower or higher rate of affectionate care. However, both groups demonstrated a clear split between mothers with low rates of affectionate care, and those with high rates. The rates of observed affectionate care ranged from a low of 0 which indicated no affectionate care to a high of 1.75 events of affectionate care per minute.

Interestingly, mothers who rated high and low on affectionate care were evenly distributed among the three primary sleeping locations. Thus, for example, bedsharing mothers were no more or less affectionate than were those mothers who placed their infant to sleep in a bassinette or crib.

A similar pattern was observed for primary feeding method. More mothers in our sample bottlefed their infants than breastfed, but neither group was more or less affectionate with their infants.

Differences in patterns of high and low rates of affectionate care were seen when examining sleep-related risks to infants. Infants of mothers rating low on affectionate care were more likely to experience specific breathing risks and feeding risks and were twice as likely to have self-fed over the course of the overnight.
Mothers with high rates of affectionate care were significantly more likely to place their baby at risk of overheating by overblanketing, which may be related to a baby's risk of SIDS.

Discussion

Low and high rates of affectionate care can be indicators for differing levels of maternal responsiveness, sensitivity and warmth but may also indicate differences in child temperament and self-regulation. Better distinguishing between types of affectionate care, such as care that is elicited from the mother by the baby in response to a particular need versus that which the mother spontaneously initiates as an expression of enjoyment of her baby may help us better understand our nighttime measures of sensitive and responsive care in relation to current daytime measures. Analysis of 8 month data will show consistency or changes in patterns.

Further analysis is needed to compare nighttime measures of affectionate care with similar daytime parenting constructs and to assess correlations with child outcomes. As we begin that analysis we will be able to determine whether or not rates of affectionate care will help identify sub-threshold levels of physical and cognitive neglect. This information will aid in the develop of appropriate interventions to improve child outcomes.

Preliminary findings on infant sleep safety, reveals that both adult and adolescent mothers modify their infant's sleep environment in unsafe ways, introducing excess soft bedding and toys that present dangerous breathing and overheating risks. Understanding what factors compel mothers to add unsafe materials to the crib, bassinette, and adult bed will assist clinicians in designing more effective educational programs to improve infant sleep safety.

Research supported by NICHD HD 39456, Centers for the Prevention of Child Neglect 2001.

References:

1. Ball, H. (2003). Breastfeeding, bed-sharing and infant sleep. *Birth* **30 (3)**, 181-188.
2. McKenna, J. (2000). Cultural Influences on Infant and Childhood Sleep Biology, and the Science that Studies It: Toward a More Inclusive Paradigm. *Sleep and Breathing in Children: A Developmental Approach*, G. Loughlin, J. Carroll, and C. Marcus (Eds.). New York: Marcel Dekker, Inc.
3. McKenna, J. et al. (1999). Breastfeeding and Mother-Infant Cosleeping in Relation to SIDS Prevention. *Evolutionary Medicine*, W.R. Trevathan, E.O. Smith and J.J. McKenna, (Eds.). New York: Oxford University Press.
4. Whitman, T.; Borkowski, J.; Keogh, D.; and Weed, K. (2001). *Interwoven Lives: Adolescent Mothers and their Children.* Mahwah, N.J. : Lawrence Erlbaum Associates.

Quantifying social asymmetric structures

Antonio Solanas[1], Lluís Salafranca[1], Carles Riba[1], Vicenta Sierra[2] and David Leiva[1]
[1] *Departament de Metodologia de les Ciències del Comportament, University of Barcelona, Barcelona, Spain*
[2] *Department de Mètodes Quantitatius i de Gestió, ESADE-Ramon Llull University, Barcelona, Spain*

Abstract

In this paper we describe a mathematical procedure for analyzing dyadic interactions among agents in social systems. The proposed technique mainly consists of decomposing asymmetric data into their symmetrical and skew-symmetrical parts, so that the asymmetric matrix is decomposed into two matrices. A quantification of skew-symmetry for a social system can be obtained by dividing the norm of the skew-symmetrical matrix by the norm of the asymmetric matrix. This calculation provides researchers with a quantity related to dyadic reciprocity in the social system. The procedure enables researchers to identify the agents whose contribution to the asymmetry is greater and also to quantify dyadic reciprocity for each agent in a dyad. These values are symmetric, making it possible to analyze data by multivariate statistical methods, which can help social researchers to capture the underlying structure.

Keywords

Dyadic reciprocity, asymmetric data, social systems structure

1 Introduction

Social phenomena have been investigated by individual research methods, although social studies often involve more than one agent. In general, these research methods focus on agents' attributes, but are not concerned with their dyadic relationships. It should be noted that, if individual research methods are used, the independence of each agents' behavior from other agents' actions is assumed. Although there is no doubt that classical statistical tests allow social researchers to discover significant patterns in social systems, it is not obvious how far individual research methods are useful for extracting regularities from agents' interactions. That is why our study deals with research methods founded on dyadic relationships.

Regarding indices proposed for describing social structures, Landau's index of dominance [3] requires each agent of a dyad to be categorized as dominating or not dominating. This transformation suppresses quantitative information and does not take into account the numeric difference between two-way measurements in a dyad. The Social Relations Model (SRM) [2] has been proposed to estimate effects from dyadic research. SRM analysis is related to a two-way analysis of variance, agents being simultaneously considered as actors and partners, and several coefficients can be calculated from interval, ratio and absolute scales of measurement. SRM enables researchers to quantify dyadic and generalized reciprocity among agents, although these indices are correlation coefficients. The statistical model for the SRM is:

$$x_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \tag{1}$$

where x_{ij} expresses any directional measurement of the interaction between the agent i and the agent j, μ denotes the mean level, α_i represents the actors' effect, β_j symbolizes the partners' effect and γ_{ij} corresponds to the agents relationship effect. The dyadic reciprocity is quantified by standardizing Cov(γ_{ij}, γ_{ji}) and assessed by a correlation coefficient. Note that a measure of discrepancy of agents' behavior is not taken into account, though this might be an adequate way of founding a quantification of dyadic reciprocity.

The main purpose of this work is to explain a new technique that allows social researchers to quantify the degree of skew-symmetry in a social system. By analyzing data with this technique, social researchers can also divide the overall skew-symmetry into parts, each assigned to one agent.

2 A quantification of skew-symmetry

Constantine and Gower [1] proposed decomposing asymmetric matrices into their symmetrical and skew-symmetrical parts in order to analyze them. An independent geometrical representation of each part of a matrix can then be obtained as a solution to the problem of representing entities or objects in a Euclidean space.

Any matrix **X** can be decomposed into its symmetric and skew-symmetric parts, here denoted by **S** and **K**, respectively. The decomposition of **X** is unique as follows:

$$\mathbf{X} = (\mathbf{X} + \mathbf{X}')/2 + (\mathbf{X} - \mathbf{X}')/2 = \mathbf{S} + \mathbf{K} \tag{2}$$

The following equality relates skew-symmetry and symmetry values:

$$1 = \frac{tr(\mathbf{S}'\mathbf{S})}{tr(\mathbf{X}'\mathbf{X})} + \frac{tr(\mathbf{K}'\mathbf{K})}{tr(\mathbf{X}'\mathbf{X})} = \Psi + \Phi \tag{3}$$

where, respectively, Ψ and Φ denote the proportions of the symmetrical and skew-symmetrical parts of a matrix. Each trace of a matrix in formula (3) returns the sum of squares with respect to the origin point. These indices allows researchers to quantify the skew-symmetry or, if preferred, the symmetry in a social system. Perhaps formula (3) becomes more understandable if its terms are expressed as a sum of squares:

$$tr(\mathbf{X}'\mathbf{X}) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}^2,\ tr(\mathbf{S}'\mathbf{S}) = \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^2 \ \text{and}\ tr(\mathbf{K}'\mathbf{K}) = \sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}^2 \tag{4}$$

where n denotes the number of agents. The latter expressions show that Ψ and Φ correspond to ratios of variability with respect to the origin point. An index relating Ψ and Φ, whose admissible values range from 0 to 1, is δ, being $\delta = \Phi/\Psi$. If δ equals to 0, social systems are symmetric and if $\delta = 1$, the skew-symmetry is at its maximum level.

3 Decomposing the skew-symmetry

Although a quantification of skew-symmetry allows social researchers to describe the whole social system as a unity, most researchers are interested in extracting information about individual agents. That is why it is necessary to decompose the total skew-symmetry into the agents' contributions. The effect of each agent on the skew-symmetry and symmetry of a social system can be obtained by decomposing the terms of the previous expression. We can obtain the contribution of each agent to the skew-symmetry and symmetry of the social system by decomposing the terms of formula (3) into the agents' source of variation:

$$\psi_j = \frac{\mathbf{s}_j^{'}\mathbf{s}_j}{tr(\mathbf{X'X})} \; and \; \varphi_j = \frac{\mathbf{k}_j^{'}\mathbf{k}_j}{tr(\mathbf{X'X})} \tag{5}$$

The magnitude or frequency of activity agents has not been standardized, and this is the main problem with ψ_j and φ_j quantifications. That is, one agent could only contribute to the measure of skew-symmetry more than others as a consequence of a higher level of activity. To obtain standardized measurements, the following equality can be defined: $\eta_j = \psi_j + \varphi_j$. Note that η_j includes the whole contribution of agent j. Now, the skew-symmetry and symmetry ratio for each agent can be written as follows:

$$\lambda_j = \frac{\psi_j}{\eta_j} = 1 - \frac{\varphi_j}{\eta_j} = 1 - \upsilon_j \tag{6}$$

The latter expression takes into account the whole contribution of each agent. Now, by means of algebraic calculus:

$$\sum_{j=1}^{n}\frac{\psi_j+\varphi_j}{\eta_j} = \sum_{j=1}^{n}\frac{\psi_j}{\eta_j} + \sum_{j=1}^{n}\frac{\varphi_j}{\eta_j} = \sum_{j=1}^{n}\sum_{i=1}^{n}\frac{s_{ij}^2}{\mathbf{s_j^{'}s_j}+\mathbf{k_j^{'}k_j}} + \sum_{j=1}^{n}\sum_{i=1}^{n}\frac{k_{ij}^2}{\mathbf{s_j^{'}s_j}+\mathbf{k_j^{'}k_j}}$$
$$= \sum_{j=1}^{n}\sum_{i=1}^{n}\lambda_{i\leftarrow j} + \sum_{j=1}^{n}\sum_{i=1}^{n}\upsilon_{i\leftarrow j} = \sum_{j=1}^{n}\lambda_j + \sum_{j=1}^{n}\upsilon_j = n \tag{7}$$

The ratios $\square_{i\leftarrow j}$ y $\square_{i\leftarrow j}$ correspond to the contribution of agent j to agent i values of symmetry and skew-symmetry, respectively. Finally, a symmetric matrix, denoted by $\mathbf{\Omega}$, can be obtained by noting the following equality:

$$\omega_{i\leftarrow j} = \frac{\upsilon_{i\leftarrow j}}{\lambda_{i\leftarrow j}} = \frac{\dfrac{k_{ij}^2}{\mathbf{s_j^{'}s_j}+\mathbf{k_j^{'}k_j}}}{\dfrac{s_{ij}^2}{\mathbf{s_j^{'}s_j}+\mathbf{k_j^{'}k_j}}} = \frac{\dfrac{k_{ji}^2}{\mathbf{s_j^{'}s_j}+\mathbf{k_j^{'}k_j}}}{\dfrac{s_{ji}^2}{\mathbf{s_j^{'}s_j}+\mathbf{k_j^{'}k_j}}} = \frac{\upsilon_{j\leftarrow i}}{\lambda_{j\leftarrow i}} = \omega_{j\leftarrow i} \tag{8}$$

The elements $\omega_{i\leftarrow j}$ of the matrix $\mathbf{\Omega}$ correspond to the common and reciprocal dyadic reciprocity among all agents. $\mathbf{\Omega}$ is a proximity matrix and can be analyzed by means of multidimensional scaling to determine the underlying dimensions in the social group and to represent the agents in a Euclidean space.

4 An example

A group of *grey mangabeys* was recently studied in the Barcelona Zoo [4]. Its composition is described in the Table 1. We have only analyzed avoiding behavior. The variable of interest considered is the number of times agent i avoided agent j in cases in which j tried to get close to i. We obtained $\delta = 0.7779$, a value that suggests the social system differs from symmetry, at least if avoiding behavior is analyzed.

If the symmetrical matrix $\mathbf{\Omega}$ is analyzed by multidimensional scaling, two underlying dimensions are found (see Figure 1). The first dimension seems to be associated with rivalry, while the second could be described as social hierarchy.

Table 1. *Composition of the group of grey mangabeys.*

Agent	Description
V	Dominant adult male
Mo	Dominant adult female
Jo	Adult female
K1	Juvenile male
K2	Juvenile female
M2	Juvenile male
M3	Infant male
J3	Infant male

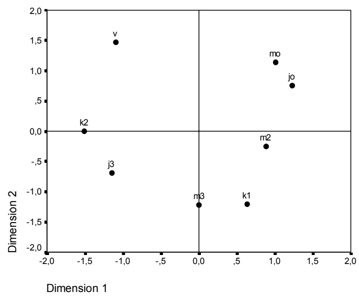


Figure 1. *Two underlying dimensions: rivalry and hierarchy.*

5 Conclusions

The technique described enables social researchers to quantify the skew-symmetry of social systems, taking into account dyadic relationships among agents. Several indices for describing social systems are derived from the discrepancies in agents' behaviors. Regarding this last point, the SRM quantifies dyadic reciprocity and generalized reciprocity as a correlation coefficient value. It should be noted that the dyadic correlation value depends on specific relationship effects among all agents, and does not correspond to the reciprocal interaction between each pair of agents. The dyadic reciprocity, as estimated in SRM, refers to the whole dyadic reciprocity in social systems, although social researchers are mainly interested in each pair of agents' dyadic reciprocity. The procedure that has been described in this work allows common and specific dyadic reciprocities to be quantified. These values, included as elements of a symmetrical matrix $\mathbf{\Omega}$, can be used to identify the characteristics of the social structure.

References:

1. Constantine, A. G., & Gower, J.C. (1978). Graphical representation of asymmetric matrices. *Applied Statistics*, **27**, 297–304.
2. Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.
3. Landau, H. G. (1951). On dominance relations and the structure of animal societies: effect of inherent characteristics. *The Bulletin of Mathematical Biophysics*, **13**, 1–19.
4. Maté, C. (1999). *Caracterización del juego social de los mangabeys grises (Cercocebus atys lunulatus) en cautividad*. University of Barcelona: Unpublished doctoral thesis.

Mobility and movement in chimpanzees (*Pan troglodytes)*: preliminary approach to the landscape use of early Homo on a local scale

T. M. Alcantarilla
Department of Prehistory and Ethnography, Universidad Complutense de Madrid, Madrid, Spain

Abstract
The reality of movement behaviour in chimpanzees (*Pan troglodytes*) is rather more complex than the image that has been proposed in the current literature on hominid ranging behaviour. The current idea is that one of the traits which defines our genus in contrast to non-human hominoids is a novel use of resources. However, the analysis of ranging patterns in great apes is an intricate issue involving many factors which interact on different levels.

We used published data from Kibale Forest Reserve, Budongo Forest Reserve, Gombe National Park, Taï National Park, Mahale Mountains, Kahuzi-Biega National Park, Kalinzu Forest Reserve, Mont Assirik, Semliki, and Ugalla to analyze the ranging patterns in chimpanzees. The results show that food availability, food distribution, grouping pattern and activity are in constant dynamic interaction, and these factors cannot be ordered into a simple hierarchy. Our statistics show that none of the factors involved plays a predominant role, rather that all four are intrinsically interconnected.

Keywords
Chimpanzee, movement, ranging patterns, early Homo.

1 Introduction
Animal movement is one of the more attractive subjects in Social Science. For example, why do animals move away, the relation between movement and speciation, and a long list of questions can be asked. However, the evaluation of ranging patterns in our remote ancestors has not received adequate attention from palaeoanthropology researchers.

When and *how* hominids colonized different parts of the world has been a key focus of interest since the 19^{th} century [1]. The majority of studies has evaluated their movement behaviour from a global point of view instead of on a smaller scale. Regarding early Homo, the analyses have focused on their capacity to occupy areas in Eurasia and their expansion through the African continent [2]. However a focus on a local scale, or in other words, the ranging patterns of early Homo has played a secondary role until the present moment. The current view is that *Homo* was the first hominoid able to leave the security of the wet rainforest to enter open and seasonal environments, nevertheless how they used the surrounding resources is still unknown [3].

The species *Pan troglodytes* has been chosen in this study to evaluate the complexity of hominoid movement behaviour, for two main reasons: *i)* their close genetic relationship with the genus *Homo; ii)* their environmental conditions are remarkably similar to the landscape reconstructions of early Homo sites. The aim of this project has been to evaluate the factors actively involved in and affecting the ranging pattern of existing wild chimpanzees.

2 Methodology
In the present poster, only published data on chimpanzees has been used. The use of published data implies the employment of material collected by other researchers and as such is always problematic. The study of ranging patterns in chimpanzees is a new issue in ecological studies, which have been focusing mainly on associative patterns, feeding, or cultural behaviour. In the cases where information exists about the use of space in a local scale, it is always indirect because of the fact that the main focus has been placed on other areas. Consequently, it has been sometimes difficult to find adequate information for the present study. Every project has a research question to answer, which is different in each case, and even when the research question is similar, each research project has its own particular approach to the problem. This fact greatly affected data collection for the present study and demanded an especially scrupulous analysis of the data employed (as is exposed below).

Data from ten chimpanzee sites were compiled into a database. The statistical tests were performed with the SPSS 12.0 software. This program was chosen for its clear advantages over other programs available on the market. SPSS allows statistical tests to be performed in a quick and easy way. In addition, SPSS can be used as a brief classical database (this is a clear advantage over Excel). Of course, as a database it has several limitations with respect to Access for example, but because of the type of data that were used, and the range of statistical tasks, SPSS was considered the most suitable computer program for meeting all our requirements.

3 Discussion
The utilization of the environment from a local approach is a complex issue to analyse because it requires the assessment of a wide number of aspects. Furthermore this subject can be approached from several ways (**Figure 1**).

Our analysis shows that there are four factors participating the most actively in chimpanzee ranging patterns. These are: food availability, food distribution, grouping patterns and activity. In opposition to several reports in favor of one of these as the principal element affecting the chimpanzee movement behaviour, our results show that none of them plays a predominant role with respect to the remaining three [4][5][6][7][8].

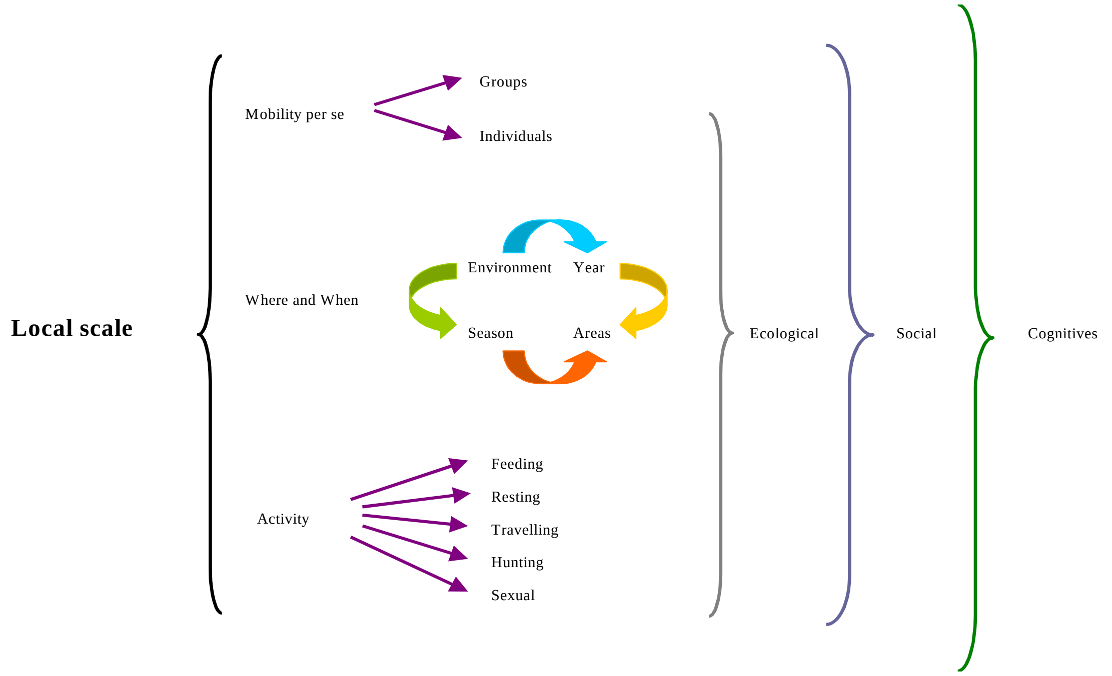


Figure 1. *Relevant factors from a local approach*

According to our results, the hypothesis of a main determinant factor is untenable. Our statistics show that none of the factors involved plays a predominant role, rather that all four are intrinsically interconnected.

The current hypotheses concerning early Homo landscape use are based on the evaluation of only one of the factors (such as activity or diet) involved in the ranging patterns. However the reality of movement behaviour in chimpanzees (*Pan troglodytes)* is rather more complex than the image that has been proposed in the hominid literature [9][10]. Many different factors underlie the concept of movement and these components interact in different levels. It is essential to avoid observing the factors separately, rather to consider all factors conjunctively to have a more complete view. If we only take a few factors into account, we can only make a partial reconstruction of the behaviour of our earliest ancestors. This is important to have an overall view of the real panorama.

With this analysis we propose that early Homo ranging patterns in a local scale have been overestimated while the chimpanzee pattern has been underestimated, and we suggest the need to review current interpretations of the archaeological data from Plio-Pleistocene early Homo hominids.

References

1. Pavlov, P.; Roebroeks, W.; Svendsen, J.I. (2004). The Pleistocene colonization of northeastern Europe: a report on recent research. *Journal of Human Evolution* **47**(1), 3-17.
2. Dennell, R. (2003). Dispersal and colonisation, long and short chronologies: how continuous is the Early Pleistocene record for hominids outside East Africa? *Journal of Human Evolution* **45**(6), 421-440.
3. Alcantarilla, T.-M.; Uomini, N. (in press). A brief theoretical impression of *Homo* abilities. *Postgraduate Forum Interdisciplinarity Conference*, University of Southampton, UK, 18 March 2005.
4. Chapman, C. A.; Wrangham, R. W. (1993). Range use of the forest chimpanzees of Kibale: implications for the understanding of chimpanzee social organisation. *American Journal of Physical Anthropology* **31**, 263-273.
5. Chapman, C. A.; White, F.J.; Wrangham, R.W. (1994). Party size in chimpanzees and bonobos. A reevaluation of theory based on two similarly forested sites. *Chimpanzee cultures*. R. W. Wrangham, W. C. McGrew, F. B. M. de Waal and P. G. Heltne. Harvard, Harvard University Press**,** 41-57.
6. Matsumoto-Oda, A. (1999). Mahale chimpanzees: grouping patterns and cycling females. *American Journal of Primatology* **47**(3), 197-207.
7. Balcomb, S. R.; Chapman, C.A.; Wrangham, R. W. (2000). 'Relationship between chimpanzee (*Pan troglodytes*) density and large, fleshy-fruit tree density: conservation implications'. *American Journal of Primatology* **51**(3), 197-203.
8. Newton-Fisher, N. E.; Reynolds, V.; Plumptre, A. J. (2000). The relationship between food supply and chimpanzee (*Pan troglodytes schweinfurthii*) party size in the Budongo Forest Reserve, Uganda. *International Journal of Primatology* **21**(4), 613-628.
9. Blumenschine, R. J., et al. (2003). Late Pliocene *Homo* and hominid land use from western Olduvai Gorge, Tanzania. *Science* **299**, 1217-1221.
10. Potts, R.; Behrensmeyer, A.K.; Ditchfield, P. (1999). Paleolandscape variation and Early Pleistocene hominid activities: Members 1 and 7, Olorgesailie Formation, Kenya. *Journal of human evolution* **37**(5), 747-788.

Analysis of individual learning sequences - an alternative methodological approach

F. Schott, S. Azizighanbari & T. Schubert
Institut für Pädagogische Psychologie I, Technische Universität Dresden, Germany

Abstract

A proposal and first results of an alternative methodological approach for analyzing interindividual differences and intra-individual constancy within the ongoing learning processes by means of dismantling learning processes into a content aspect and a behavioral aspect are presented.

Keywords

analyses of learning; learning tasks; learning processes

1 Introduction: Needs and results of research on learning processes

A main aim of this approach is to overcome the problem of describing and analyzing complex learning tasks and processes. For that purpose observation of ongoing individual learning processes [5] and learning strategies seems to be promising.

Previous research [3] which was based on such observation of learning processes suggested that there is intra-individual constancy in learning from text despite of interindividual differences. In more complex instructional settings one could expect even larger interindividual differences.

2 Goals and questions of this approach

(1) Are there any inter- and intra-individual action patterns to identify in complex learning settings?

(2) If there is intra-individual constancy in learning sequences, is there a connection within individual learning sequences and success in complex learning?

(3) Individual constancies and a correlation within such and learning success may lead to the question, which influence structures of tasks exert on learning sequences and learning success?

3 Conceptual basics for the analysis of tasks and processes of learning

3.1 The state-change-concept - A conceptual base for an observation and coding system

The State-Change-Concept (Zustands-Änderungs-Konzept) [2, 4, 6, 7] is based on educational task analysis procedures. Following this concept tasks can be divided into states: the chosen formulation of the task ("Aufgabenstellung") and the solution of the task ("Aufgabenlösung"). These states reflect the content aspects of the task. The change from one state to the other refers to the behavioral aspect of the task. An operation consists of at least two states and one operator (for transferring state 1 into state 2).

The shown concept makes an analytical non-empirical quantification and detailed description of learning processes possible. It is used as a basis of the observation and coding system of this approach that allows identification and comparison of domain specific operations and generic operations. Each sequence (State 1, Operator, and State 2) is coded as an operation. The main advantages of this approach are the facts that (a) specific operations (bound to task) can be coded in the same way, and (b) similar operations in different tasks are comparable, because they are coded the same way. An example of analysis of learning processes based on the introduced concept is shown in figure 1.

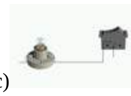
(1) Verbal Description		
light bulb and switch are not connected	connecting light bulb and switch	light bulb and switch are connected
(2) Computer Simulation (Virtual)		
(2a) 	(2b) 	(2c) 
(3) Construction Task (Real Objects)		
(3a) 	(3b) 	(3c) 
(4) Generic Operation (a form of CONNECT)		
2 not connected objects; 1 element to be used for connection	connecting	2 connected objects; 1 element used for connection
(5) Possible Learning Sequence		
By means of repetition (of the operation *CONNECT*) this operation is practiced and improved.		

Figure 1. *Verbal description, virtual and real representation and generic description of an operation and an example of a possible learning sequence within the State-Change-Concept [1, in press].*

3.2 Finite Deterministic Automata - A theoretical base for description of learning sequences

Finite deterministic automata (FDA) are models of computation that allow simulations of processes as changes of discrete states. Each next state depends on the current state and current input character. Here two main components of FDA are mentioned: the state alphabet (Σ_Q) and the terminal or input alphabet (Σ_T). The state alphabet of the automaton refers to single states the process can adopt, inclusive the initial state and the final state. The input alphabet describes inputs or actions. The productions of an automaton as different state sequences can be presented in different ways, in tables or diagrams. An example for such a production which shows one possible state sequence of a test subject who convert a series connection into a parallel connection is given in figure 2.

Even learning processes as focused in this approach can be modeled by means of a FDA. So FDA can be used to compare individual sequences to each other, individual sequences with an ideal sequence, to quantify the similarity of sequences and probabilities for expected sequences and finally to identify clusters of sequences out of a number of individual sequences.

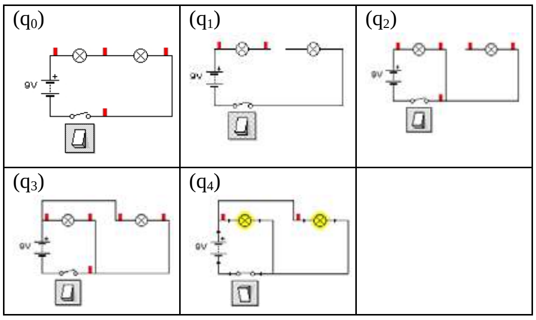


Figure 2. States appearing within the conversion from a series connection into a parallel connection

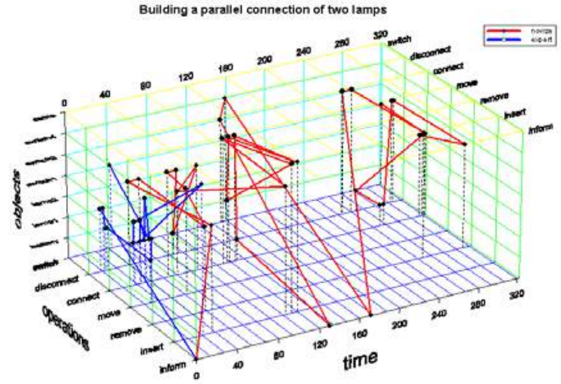


Figure 3. The sequence of operations with different objects over time by an expert and by an novice when building a parallel connection of two lamps

4 First results of Preliminary Studies

Several variations of experimental designs are currently developed, not yet carried out. Here only results of preliminary studies to investigate interindividual differences and intra-individual constancies within the ongoing learning process are presented.

In these studies the tasks learners had to deal with comprised four different electric circuits: a simple circuit, a serial connection, a parallel connection, and a change-over connection. These tasks were presented in two different a “virtual” and a “real” learning setting. The virtual setting demand subjects to construct electric circuits with use of learning software for Physics in a computer simulation. In the “real” learning setting subjects had to construct electric circuits with real objects with use of a construction kit.

The shown behavior of the subjects while solving tasks has been coded with a coding system which was based on the mentioned state-change-concept and included several categories of behavior (inform, connect, disconnect, move, insert, switch) and objects (paper, lamp 1, lamp 2, cable, switch 1, switch 2). Required time for operations with objects was taken.

Results revealed differences within sequences of actions of same subjects in different turns. While second turns subjects solved tasks faster than in first trials. They were reading the information only a short time and did not insert not needed objects.

In this study approximately ideal operation sequences by an expert got compared to the sequence found in observing a novice too. The expert was disconnecting, ordering and connecting fast and took nearly no information. Instead the novice took information in 3 long lasting phases and inserted not necessary objects. It took him 8 times more time to solve the task which is shown in Figure 3.

First empirical results in this investigations revealed useful insights in the analyses of learning processes and encouraged in going on with this approach.

References:

1. Hillebrandt, D.; Schott, F. & Schubert, T. (in press). Analysis and Prediction of Individual Learning Pathways – Are there Specific Advantages of Virtual Learning Environments Compared to their Counterparts in Reality?
2. Schott, F. (1986). *Lehrzielorientierte Unterrichts-psychologie. Bausteine einer psychologischen Theorie lehrzielorientierter Lehr-Lern-Prozesse.*
3. Schott, F. (1992). The Useful Representation of Instructional Objectives: A Task Analysis of Task Analysis. *Instructional Models in Computer-Based Learning Environments*, Berlin, Springer.
4. Schott, F.; Hillebrandt, D.; Al-Diban, S.; Schubert, T.; Heyne, N.; Herrmann, U. (2002). Complex Learning in multimedia Environments - Which kind of Learning and How to do research on it?. *EARLI SIG 6 Biannual Workshop*, Erfurt, Germany, 27 th - 29th June 2002.
5. Schott, F. & Seidel, P. (1997). PLANA: An ID Model Focusing on Instructional Task Analysis. *Instructional design: International perspectives - Theory, research, and models*, **1**, 79-112.
6. Schott, F.; Neeb, K.-E. & Wieberg, H.-J. W. (1981). *Lehrstoffanalyse und Unterrichtsplanung - Eine praktische Anleitung zur Analyse von Lehrstoffen, Präzisierung von Lehrzielen, Konstruktion von Lehrmaterialien und Überprüfung des Lehrerfolgs,* Braunschweig, Germany.

Methods for diagnostic lying behavior: "The Carlson's button"

A. Kankova
Department of Psychology, Belarusian State Pedagogical University, Minsk, Belarus

Abstract
The present abstract describes the phenomenon of truthful-lying behavioral and methods for measuring truthful-lying behavior in children between 3 and 7 year of age. Lying is considered as behavioral form, which includes lying pronouncement; recognition that this pronouncement is lying; and the striving attaching wittingly lying thought kind of truth. Although lying behavior has received attention from developmental psychologist with diverse theoretical orientations, a relatively smaller number of studies have proposed systems for diagnostic lying behavior to be based on the association between ability to lie and cognitive mechanisms. In this abstract technique for diagnostic truthful-lying is proposed behavior. Methodological basis of the technique named "The Carlson's button" is the function-stage model of ontogenetic development (Karandashev, 1991).

Keywords
Truthful behavioral, lying behavioral, realization, relation, understanding, reflection.

1 Introduction
The truth and lie concern to group of those notions which are widely used in a science, in pedagogical and social practice, and also in daily human life. Wide application of the given notions inevitably washes away their contents. Till now there isn't common opinion about distinctive attributes of lie and border between lie and the truth.

The question on children's lying has the important theoretical and practical value. From the practical point of view it is important to understand a root, function and conditions of lying. From the theoretical point of view is also very interesting. The phenomenon of children's lying adjoins to a lot of facts of childhood psychology. These facts are extremely wide: features emotional, cognitive and social development of the child of preschool age. So, in order to define the facts which should be subject to the scientific analysis, it is necessary to determine notion of lying and to create specific complex system of diagnostics for studying truthful-lying behavior in children between 3 and 7 year of age.

2 Theory
Lying is considered as behavioral form, which includes lying pronouncement; recognition that this pronouncement is lying; and the striving attaching wittingly lying thought kind of truth [1]. Thus, for studying genesis and dynamics of truthful and lying behavior of preschool children it is necessary to study cognitive mechanisms in their basis.

In spite of the fact that the phenomenon of lying behavior draws attention of representatives of various psychological directions, there are a relatively smaller number of studies have proposed systems for diagnostic lying behavior. Therefore we offer the technique, allowing studying laws of display of truthful and lying behavior children of preschool age depending on the certain mental functions.

3 Methods
The name of a technique: the Technique for diagnostic truthful-lying "The Carlson's button".
Contingent of examinees: children in the age of 3-7 years.
Subject of diagnostics: truthful and lying behavior.
The operational status: an experimental situation.
Competence of the executor: the psychologist of a various professional standard.
Completeness and structure of a technique: the technique represents the experimental situation including the contradiction which should be reflected in consciousness of the child and it is solved of them.
The equipment: the text of the description of a technique; two identical boxes with buttons of dark blue color taking place in them; a set including 5 red, 5 dark blue and 5 green buttons; a sweet, the report of fixing of results.
Theoretical positions: the theoretical basis of the technique is the idea of V.S. Muhina who considers lying as conscious distortion of truth which occurs when the child appears in a situation in which there is a discrepancy of his desires and showed requirements or norms. Here the child solves for himself a dilemma: to satisfy his desire and to pass test by alienation on the part of the adult, or to remain in kind attitudes with the adult, but to refuse his desire. Such situation becomes for the child a situation of double motivation [3]. The choice, as experimental, situations of double motivation is caused by ideas of V.V. Zenkovski about lie, as behavioral form, which includes lying pronouncement; recognition that this pronouncement is lying; and the striving attaching wittingly lying thought kind of truth. So, about conscious lie as a specific form of behavior directed on solving a contradicted situation, it is possible to speak only provided that the child is capable to reflect this contradiction. And it is possible only at the certain stage of his mental development.

In connection with that ability to truthful and lying behavior is a phenomenon of mental development, as a methodological basis of a technique the function-stage model of ontogenetic development [2]. In this model a child psychic development is considered as a development of system of psychic functions. In preschool age the following leading psychic functions are: realization, relation, understanding and reflection. In the first phase, as a leading function is realization. It is ability of the child to present him in the given position of another, to allocate it in variety of vital displays. In the second phase – the relation – as a control and management of reflection of the reality from the not own point of view. In the third phase – understanding - construction of an image of the reality from the not own point of view. That means that the child builds an internal position as imagined without a support on subject activity. In the fourth phase – a reflection – as independent construction of an image of the reality from the not own point of view. That means that the child actively builds an internal position. The internal position is determined as a position occupied by child in the cogitative plan which is distinct from his own position [2]. Being in the certain phase of psychic development, the child shows specific behavior.

Work: the Technique of studying of truthful-lying behavior consists that the experimenter tells the child a story about Carlson who has lost his button. Without it he may not start a propeller and return to home on a roof. The child is offered to help Carlson find his button.

The child is presented two identical boxes, containing dark blue buttons and a few red buttons. The experimenter informs the child that anybody except for him may not look in boxes. The experimenter gives the child a sweet and tells him that he may eat it, only provided that he will execute tasks explained by Carlson: 1. It is possible to open only one box. 2. If the child sees a red button He can keep the sweet for himself which was given by the experimenter. 3. If the child sees a dark blue button he should give a sweet to another child which will wait for him in the next room. 4. After the task with boxes will be executed, the child should go in the next room and speak another child what color of button has seen.

After the child receives the instruction, the experimenter leaves a room for 1 minute. The behavior of the child in the meantime is recorded with a video camera. When the experimenter comes back, he asks the child, how many boxes he opened also what color of button he has seen. . Then the experimenter compares statements of the child and his behavior recorded. ***The approximate text of a technique: "***Now I shall tell you a story about Carlson. Listen attentively and remember.

Today in the morning I have met Carlson. He was very sad. And when I have asked him why he is such sad, he has told me a story which has happened to him. He has lost a button which started the propeller. And then suddenly Carlson said that only two boys (girl) *Nick* and *Mike* (names of children participating in experiment) can help him. But they should make all how Carlson will tell Further for investment of perception of the text children participating in experiment will be named *Nick* and *Mike*.

Carlson has passed two boxes and some red, dark blue and green buttons. They lay before you (to pay attention children to two boxes and a set of multi-colored buttons). Carlson searched everywhere but so anything and has not found. There were only these two identical boxes in which Carlson did not look. Anybody cannot open these boxes, except for *Nick*. So, Carlson asks you: *Nick* and *Mike* to help him to find a button. Now *Mike* (one of children) goes with me in the next room and wait *Nick* (other child) which should come and speak what color button was seen.

The task for first child (*Mike*):

Mike should ask the *Nick* the following questions:

1. What color of button he (*Nick*) has seen? - 2. If *Nick* will tell that he has seen a red button he can leave to himself a sweet and eat it. - 3. If *Nick* tells that he has seen a dark blue button he should give a sweet to you (*Mike*).

The task for second child (*Nick*):

Nick should search for a red button in boxes. Only *Nick* may look in boxes. *Nick* can help Carlson find a button if he will satisfy all conditions about which has told Carlson:

1. It is possible to open only one box. - 2. Anybody, except for *Nick* may not look in boxes. Carlson will be at this time in other room and to wait *Nick*, - 3. *Nick* should come to *Mike* and tell what color of button he has seen. - 4. If *Nick* will see a button of red color he can leave to himself a sweet which has passed Carlson. - 5. If *Nick* will see a button of dark blue color he should give Carlson a sweet. You should be very attentive because you will carry out these tasks independently without me. I should go to call. You don't wait for me, carry out the task and go in group. Let's repeat conditions, which need to be executed to find the Carlson's button. Let's begin *Nick.* (In brackets we answer the experimenter's questions which are significant about understanding of the technique's instruction).

1. How many boxes you may open? (One) - 2. What color of button you should find? (Red) - 3. What should you make after will open a box and will see color of button? (To go in other room to *Mike*) - 4. What you should tell *Mike*? (What color of button I have seen) - 5. What you will make with a sweet if you will see red button? (I shall take away with myself in group) - 6. What you will make with a sweet if you will see a dark blue button? (I shall give *Mike*).

And now we shall repeat what *Mike* should make.

1. What you should ask at *Nick*? (What color of button has seen) - 2. If *Nick* will tell that has seen a red button what he will make with a sweet? (Will take away with itself in group) - 3. If he will see a dark blue button what he will make with a sweet? (He will give me)

Processing of results: each aspect of truthful and lying behavior shown by the child is marked is familiar "+". Set of separate aspects determines truthful and lying behavior.

Interpretation of results: Combination of separate aspects of children behavior permits to conclude the specificity of truthful and lying behavior of children who are taking place in one of four phases of preschool age.

Truthful behavior. *Disciplined truthful behavior*: the child opens one box, informs that he has opened one box and has seen a dark blue button. *Undisciplined truthful behavior*: the child opens two boxes, informs that he has opened two boxes and has seen dark blue buttons.

Lying behavior. *Strictly lying behavior:* the child opens two boxes, informs that he has opened one box and has seen a red button. *Deceitful behavior:* the child opens two boxes, changes a dark blue button on red, and informs that he has seen a red button.

4 Results

Results of approbation of a technique show that being in 1st phase (3-4 years), children open 1 box and inform, that saw a dark blue button, i.e. show the disciplined truthful behavior. Being in 2nd phase (4-5 years), children open two boxes but do not hide that saw dark blue a button, i.e. also show truthful behavior but undisciplined. Being in 3rd phase (5-6 years) preschool age, children open two boxes and inform that saw a red button, i.e. show strictly false behavior. Children who are taking place in 4th phase (6-7), open two boxes, change a dark blue button on red, inform that opened one box and saw a red button, i.e. show deceitful behavior.

5 Conclusions

Results show that the ability to lie arises in second half of preschool age. This ability is linking with such mental functions as understanding and reflection.

References:

1. Zenkovski V.V. (1995) *Psychology of the childhood*.
2. Karandashev Y.N. (1991) *Age function-stage model of ontogenetic development of children*.
3. Muhina V.S. (1999) *Phenomenology of development and being of the person*.

Video-based ethnographic approaches to defining an emotional connection between caregivers and care recipients in nursing homes

L. Levy-Storms
Departments of Social Welfare and Medicine/Geriatrics, University of California, Los Angeles, USA

Abstract

The purpose of this paper will be: 1) to present preliminary results on the affective quality of verbal and non-verbal communication of caregivers/care recipients' interactions during mealtime from established non-relational measurement protocols and 2) to define relational qualities of caregivers/care recipient interactions that can be measured. Eleven volunteer caregivers and care recipients were video- and audio-taped during mealtime interactions. The largest proportion of affective verbal communication was social behavior (32%) and then agreement (10%). The largest proportions of total observed time that caregivers' displayed affective non-verbal communication behaviors included eye gaze (20%) and smiles (10%). The largest proportions of observed time that caregivers displayed positive affect were "attentiveness" (38%) and "involved attitude" (27%). Using excerpts of video and audio from these interactions, relational measures of caregivers "emotionally connecting" with care recipients will be illustrated and defined operationally by specific observational criteria.

Keywords

video, audio, in-person observations, aging, long-term care

1 Introduction

Several weaknesses exist in well-established approaches to measuring the affective qualities of verbal and non-verbal communication. The main limitation is the focus on either one person or the other in a dyadic interaction instead of the affective quality of the relationship between them. However, the affective quality of the relationship can have important effects on behaviors of one or both of the dyad members. The purpose of this paper will be: 1) to present preliminary results from adapting established coding protocols that capture the affective quality of verbal and non-verbal communication of either physicians/patients or teacher/students to caregivers/care recipients' interactions during mealtime in nursing homes and 2) to define relational qualities of caregivers/care recipient interactions that can be measured.

2 Methods

All recruitment and study procedures were approved by the UCLA Human Subjects Review Board. Subjects came from two long-term care facilities and included paid caregivers and care recipients. Caregivers were recruited from staff meetings. Care recipients were self-consented or assented after legal guardian consent. Average consent rates were low: 33% (caregivers) and 55% (residents) across the two facilities. The sample for this pilot study was non-random and included a total of ten dyads (caregiver and care recipient) and one triad (one caregiver and two residents). Some caregivers and/or residents participated in more than one dyad with one dyad being filmed twice. All except one caregiver were female. All of the residents were female with varying levels of cognitive impairment. In all, eleven interactions compose the study sample with five from one facility and six from the other. Caregivers and care recipients were informed ahead of time of the day and mealtime of video- and audio-taping. All interactions took place during lunch or dinner. Four out of six sessions were recorded in the facility main dining room with others taking place in a small private dining room. Most caregivers sat next to the resident during the meal, except 2 residents who were independent and 1 that required minimal assistance from caregivers. The sample included six care recipients and seven caregivers from two long-term care facilities. Some care recipients/caregivers participated in more than one observation, although none were paired with the same person more than once. The camcorder was a Sony DVR30; the audio device was a Sony digital voice recorder (ICD-BP150). A tripod was used for all videotaping, and it was placed such that the caregiver's face was optimally viewable with as much of the resident's face and food tray kept as visible as possible. Although the observation protocol called for the video- and audio-taping to begin when the tray arrived and to continue until the tray is removed, during one session the meal had already begun when recording began.

2.1 Measures of Affective Verbal Communication

All verbal communication between caregivers and care recipients during feeding assistance was audio recorded and transcribed verbatim. Each transcript was imported into *Ethnograph* [9], a qualitative analysis software program, for coding. Coding will entail several levels. The most basic level includes the identification of 24 specific verbal behaviors based on the labeling of verbal utterances such as a word, phrase, or complete sentence (e.g., empathy/legitimize; reassurance/encourage). These 24 categories have been adapted from Roter's Interaction Analysis System's (RAIS) 34 categories [8], which were originally designed to code physician-patient communication and were found to be reliable [3]. For example, inter-rater reliabilities ranging from a low of .69 to a high of 1.0 across the 24 categories has been found in similar research on older adults in long-term care settings [3]. These 24 verbal communication behaviors have been clustered into two sub-types of affective communication behavior based on correspondence analyses (CA), which is similar to a factor analysis [3]. The two affective sub-types or classification schemes include the following: *Affective communication*: pertains to utterances that express verbal attentiveness, concern, and providing of empathy from caregiver. *Social communication*: contains information on the caregivers' use of social conversation such as personal statements, small talk and banter with no particular function related to care. The frequency of these two sub-types of verbal communication will be counted and divided by the total number of utterances (the smallest possible complete unit, excluding sentence fragments, phrases and fillers) [2] spoken by caregivers per feeding assistance episode [3].

2.2 Measures of Affective Non-verbal Communication

The following six non-vocal, nonverbal categories and definitions are based on previous research on older adults in long-term care settings: *Patient-directed eye-gaze* occurs when the caregiver looks directly at the care recipient [4]. *Head nods* are defined as vertical movement of the caregiver's head as a sign of attentiveness in conversation or as reinforcement of the spoken word [4]. *Smiling* is a gesture of friendliness primarily exhibited by a distinct facial grin [4]. *Affective Forward leaning* is a posture that a caregiver exhibits by bending towards or sitting closer to a care recipient when doing so is not necessary to carry out care [4]. *Affective touch* is expressive and relatively spontaneous with emotion and not necessary for completing care [7]. Using the video-recorded interactions, the total frequency of each nonverbal gesture by caregivers only across one-minute time intervals will be noted and divided by the total number of intervals during observation time [4].

2.3 Measures of Affective Verbal Communication

Global affect ratings of communication behaviors have been utilized in previous research to assess the quality of communication between caregivers and patients in a variety of settings [1,4,5,6,8]. Using the video-recorded interactions, the total frequency of each display of affect by caregivers only across one-minute time intervals will be noted and divided by the total number of intervals during observation time [4].

3 Results

Mealtime interactions averaged 32 minutes and 28 seconds (32:28) with a range of 06:25 to 55:55. Preliminary results on the specific types of affective verbal and non-verbal communication of these caregivers and their global affect follow. The proportion of the total verbal utterances that were affective by subtype included: social behavior (32%), agreement (10%), paraphrase (4%), shows concern or worry (1%), empathy (3%), and reassurance (.2%). The average proportions of total observed time that caregivers' displayed affective non-verbal communication behaviors included: smiles (10%), eye gaze (20%), head nods (8%), touch (5%), and forward leans (4%). The average proportion of observed time that caregivers displayed affect was as follows: "attentiveness" (38%), "involved attitude" (27%), and "warm" (14%).

4 Conclusions

Although these data suggest that affective levels of caregivers' communication vary by channel and specificity of the unit of observation (i.e., particular behaviors or global affect), they do not explicitly incorporate the care recipients' communication behaviors simultaneously as they relate to the caregivers' behaviors. For example, is not always clear whether when the caregiver smiles, h/she is smiling at the care recipient or whether the care recipient even notices. Such a measure would be relational.

4.1 Relational Measures of Emotional Connectedness

From strategies to emotionally connect caregivers with care recipients, relational behaviors can be targeted and measured. For example, techniques to "get the attention of the care recipient" include 1) making eye contact and keeping it; 2) Positioning yourself in front of the person; 3) Using the care recipient's name, a social greeting or phrase to help the person "get ready" to listen; 4) Being patient and giving time for processing; 5) directing and redirecting their attention often [10]. These relational behavioral techniques will be illustrated in a 5 minute DVD video, and a configuration for coding them in The Observer will be described.

References

1. Bensing, J.M.; Dronkers, J. (1992). Instrumental and affective aspects of physician behavior. *Medical Care* **30(4)**, 283-298.
2. Caporael, L.R. (1981). The paralanguage of caregiving: baby talk to the institutionalized aged. *J. Pers. Soc. Psychol.*, **40(5)**, 876-84.
3. Caris-Verhallen, W.M.C.M.; Kerkstra, A.; van der Heijden, P.G.M.; Bensing, J.M. (1998). Nurse-elderly patient communication in home care and institutional care. *International Journal of Nursing Studies*, **35**, 95-108.
4. Caris-Verhallen, W.M.C.M.; Kerkstra, A.; Bensing, J.M. (1999). Non-verbal behaviour in nurse-elderly patient communication. *Journal of Advanced Nursing,* **29(4)**, 808-818.
5. Caris-Verhallen, W.M.C.M.; Gruijter I, M.; Kerkstra, A.; Bensing, J.M. (1999). Factors related to nurse communication with elderly people. *Journal of Advanced Nursing,* **30(5)**, 1106-1117.
6. Harris, M.J.; Rosenthal, R.; Snodgrass, S.E. (1986). The effects of teacher expectations, gender, and behavior on pupil academic performance and self-concept. *Journal of Educational Research*, **79(3)**, 173-179.
7. LeMay, A.C.; Redfern, S.J. (1987). A study of non-verbal communication between nurses and elderly patients. In: Fielding P, ed. *Research in the Nursing Care of Elderly People*. New York: John Wiley & Sons, 171-190.
8. Roter, D. (1989). *The Roter Method of Interaction Process Analysis,* Baltimore: Johns Hopkins University Press.
9. Seidel, J.; Friese, S.; Leonard, D.C. (1995). *The Ethnograph*, 4.0 ed. Amherst: Qualis Research Associates.
10. Kohler, S. (2004). *How to Communicate with Alzheimer's: A Practical Guide and Workbook for Families*. Los Angeles: GRP, Co.

Supported by National Institute on Aging (Grant# 5K01AG020561) and the National Institute on Aging UCLA Claude D. Pepper Older Americans Independence Center number AG10415. Special appreciation to Susan Kohler, Courtney Martin, Maria Alonzo, and Benjamin Wheeler.

Assessing methods of data capture in pediatric oncological settings

M.C. Naughton[1], T.S. Foster[1], T.L. Abrecht[1,2], A.M. Peterson[1], L.A. Penner[1,2] and J.C. Ruckdeschel[1]
1. Communication and Behavioral Oncology, Karmanos Cancer Institute, Wayne State University, Detroit Michigan
2. Department of Family Medicine, Wayne State University, Detroit Michigan

Abstract

Traditional methods for coding human interactions are the use of pen and paper. Electronic systems, such as Noldus Observer, provide an alternative method. The application and value added of electronic coding in human interaction, especially in a medical context, has not been extensively explored. The goal of this study was to determine whether an electronic methodology is a more accurate and efficient means of coding than traditional hardcopy coding. In this study, we examined the differences in coder accuracy, time, efficiency, reliability and ease of data processing and analysis. Four coders analyzed nine tapes of interactions between pediatric oncology patients and their parents. The proximity of the parent to the child and the amount of tactile contact was recorded using either a pen/paper coding method or the Noldus Observer system. The findings supported the value added by using the electronic coding methodology.

Keywords

Manual, Electronic, Data Collection, Coding Methodology

Introduction

The key question in this investigation was whether electronic forms of data collection represent an improvement over manual data collection. Very little research has been done to determine that there are any benefits to capturing data electronically. This study sought to identify what benefits, if any, are gained by using an electronic method.

The manual method of coding, using a pen and paper to record visual data, is the standard method of collecting data. Electronic methods, such as Noldus Observer©, have entered laboratories and are being used on a multitude of research projects. Studies using electronic methods of data collection span viewing animal behavior, emotional expression of pre-school children, parents of children, real-time interaction patterns in soccer, and patient doctor interactions in an oncological setting.[1,2,3,4]

We conducted a pilot study comparing traditional coding methods, using pen and paper in conjunction with Windows Media Player, to electronic coding methods using the Observer Video-Pro. This study utilized data gathered as part of a larger investigation of caregivers' role in pediatric cancer pain and survivorship. Participants were consented via IRB approved guidelines and their clinic visit was recorded using a non-intrusive video recording system described by Albrecht et al.[4] Utilizing these data, we compared traditional and electronic coding methods. Two specific research questions were addressed:

- Is there a difference between the two methods in the accuracy of coding?
- Is there a difference in time efficiency?

Methods

Sample

Our sample was selected from pre-existing video recordings of interactions among of nine children receiving cancer-related treatments and their primary adult caregivers during the clinic visit (parents, other relatives, guardians). Five male and four female children, ranging in age from 3 to 12 years comprised the sample. Two male and seven female caregivers were coded. Five children were Caucasian and four were African American.

Coding System

The coding system was a modified version of one developed by Peterson et al.[5] in a previous study of caregivers' interpersonal distance (proximity) and touch behavior. Coding categories were simplified to include two behavior classes (proximity and touch) with three types of behavior elements in each class. See Table 1.

Proximity	Touch
• Within Reach • Outside Reach • Unknown	• Supportive • Instrumental • No Touch

Table 1. *Behavior Classes and Elements*

Proximity or interpersonal distance was based on approximate distances from the parent to the child's upper torso or specifically between the navel and the top of the head. Codebook definitions were provided to determine if the caregiver was within reach of the specified area, outside of reach or unknown (i.e. caregiver left frame, video was blocked or lights were turned out). Touch behaviors were defined as any instance of tactile contact specifically from the parent to the child. Two categories of touch were defined: touch that is supportive (e.g. hugging, kissing, holding, etc.) and touch that is functional (e.g. helping, examining, cleaning, etc.) in nature, as well as absence of touch or no touch.

Coder Training

A codebook was developed for both manual coders and electronic coders that included definitions, explanation of codes and providing specific examples. Both sets of coders were trained on behavioral classes and elements concurrently, while addressing any questions regarding procedure. Coding method specific information was then given to each group independently (e.g. saving of files, layout of coding schemes, etc.).

Coding Procedures

Four independent coders viewed videotaped interactions of parents and children during pediatric oncology treatments. Two of the coders used a manual coding system consisting of paper coding sheets and the other two used the Observer Video Pro, a computerized system for the collection, analysis, presentation and management of observational data. Tapes and conditions (manual, electronic) were randomly assigned to coders. All coding assignments were timed and compared to the actual length of the video. Upon completion of coding, the first two authors independently reviewed and verified results and indicated variations in coded behaviors. One author verified all manual coding and all electronic coding was verified by the second. The authors looked for accuracy in coding, missed behaviors and coding of behaviors that did not actually occur. After this, the first two authors collaborated to develop a consensus judgment of behaviors that actually occurred and did not occur - for each interaction. This information was double checked for quality control and accuracy.

Results and Discussion

Nine videotaped interactions were used in the analysis. Results are presented by research question.

Accuracy and Inaccuracy

Accuracy was determined from three different measures: The number of behaviors coded correctly; the number of behaviors missed; and the number of behaviors coded incorrectly (i.e. coding behaviors that did not actually occur).

No significant difference was found in the number of behaviors coded correctly (t (16) = .079, ns), indicating that the manual coding method and the electronic coding method captured similar data within the coding scheme. Coders using the manual system correctly coded 679 of the 806 behaviors that actually occurred, (84.24%); 663 (82.25%) were accurately coded using electronic methods. In addition, no significant difference was found in the number of missed behaviors indicating that the incidence of missing actual behaviors was no greater in one method than the other. Specifically, of the 417 missed behaviors, 195 (46.76%) were missed by manual and 222 (53.23%) were missed by electronic methods.

There was a significant difference in the amount of inaccurately coded behaviors between the two methods (t (16) = 2.830, p>.05). Across the nine tapes, an average of 17.44 behaviors were incorrectly identified using the manual method and 5.11 using the electronic method. An alternative way of looking at these differences is, of the 203 coded inaccuracies, 157 (77.33%) were coded using the manual system and 46 (22.66%) were coded using the electronic system. See Figure 1.

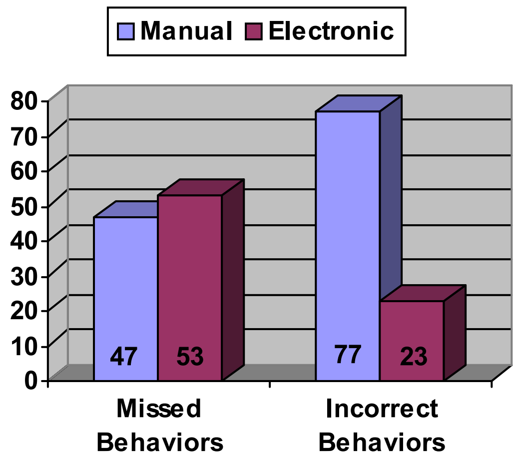


***Figure 1**. Percent Comparison of Errors*

Time and Resource Efficiency

Coding time for manual was noticeably more than electronic coding time, in some cases as much as twice as long than coders using the electronic method. Although the difference was not significant (t (16) = .1.400, ns), in every case manual coding required more time to complete than electronic coding. Total coding time for each method is depicted in Figure 2. Further, this comparison does not include the additional time needed when coding manually to input data into a format that is conducive to data analysis, a process which is automated with electronic methods. Consequently more time and resources are needed to complete project specific tasks.

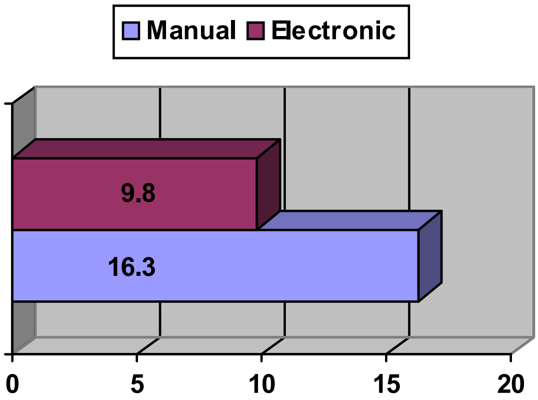


***Figure 2**. Coding Time in Hours*

Conclusion

Although manual methods have been primarily used in the past, this study demonstrates the value of coding using an electronic method. Time and resources needed when planning and implementing a project can be reduced by coding electronically without losing accuracy in data capture while significantly reducing the number of errors in data capture. This supports the accuracy and efficiency of using electronic methods. Further research will continue to look at differences in coder accuracy and ease of use across methodologies.

Supported by National Cancer Institute (RO1 CA 100027-01)

References:

1. Bischof, A., Stark, H., Blumenstein, R., Wagner, T., Brechmann, A., Scheich, H. (2001) A semiautomatic method for adding behavioral data to videotape records in combination with the Noldus Observer system. *Behavior Research Methods, Instruments & Computers,* 33,549-555.
2. Agliati, A.A., Anolli, L. Jonsson, G.K. (2002) Time patterns of emotional communication in children. Proceedings of Measuring Behavior 2002, 4th International Conference on Methods and Techniques in Behavioral Research (Amsterdam, The Netherlands, 27-30 August 2002), 3-4.
3. Jonsson, G.K., Bjarkadottir, S. H., Gislason, B., Borrie, A, Magnusson, M.S. (2003) Detection of real-time patters in sports interactions in football. In: C. Baudoin (ed.), *L'Éthologie Appliquée Aujourd'hui. Volume 3. Éthologie Humaine*, pp.27-45. Levallois-Perret, France: Editions ED.
4. Albrecht, T. L., Ruckdeschel, J. C., Ray, F. L., Pethe, B. J., Riddle, D. L., Strohm, J., Penner, L. A., Coovert, M. D., Quinn, G., Blanchard, C. G. (in press). A portable, unobtrusive device for video recording clinical interactions. *Behavior Research Methods, Instruments, & Computers.*
5. Peterson, A.M., Cline, R.J., Foster, T.S., Penner, L.A., Albrecht, T.L., Keller, C.M., Naughton, M.C., Jones , E.M., Ruckdeschel, J.C. (2005). Quieting Strategies for Pediatric Cancer Patients Undergoing Painful Procedures. Poster session presented at the annual meeting of the American Psychosocial Oncology Society, Phoenix, AZ. January 2005.

Searching for diel rhythmicity in behavioral data collected by The Observer

R. Zemek [1], R. Socha [2]
[1] *Department of Experimental Ecology, Institute of Entomology, České Budějovice, Czech Republic*
[2] *Department of Physiology, Institute of Entomology, České Budějovice, Czech Republic*

Abstract

The paper describes how to perform simple chronobiological analysis of data obtained by The Observer using feeding behavior of the firebug, *Pyrrhocoris apterus* (Heteroptera) as an example. Data on feeding activities were obtained from videotaped recordings using The Observer 3.0 with Support Package for Video Tape Analysis. For further analysis raw observation data files (*.ODF) were used. The method is based on circular statistics and randomization test applied to examine whether there is statistical evidence of non-randomness of activity or whether the activity is uniformly distributed throughout the day-night cycle. Computations and drawing of graphs were performed using Fudgit and Gnuplot programs under Linux and a PASCAL program RanDir.

Keywords

Biological rhythms; data analysis; circular statistics; feeding activity; insects.

1 Introduction

Biological rhythms are usually studied using various activity monitors based e.g. on infrared sensors or video image analysis systems. These electronic devices have been also successfully applied in a long-term automated recording of insect locomotion [7]. Study of rhythmicity in complex behavioral activities of insects like feeding or mating can be hardly automated but it is possible to take the advantage of computer-assisted event recording in connection with a time-lapse video recorder. In this paper we show how data collected by means of The Observer can by analyzed to reveal diel changes in feeding activity of the firebug, *Pyrrhocoris apterus* (Heteroptera).

2 Materials and Methods

2.1 Recording of feeding activity

Two-weeks study of bugs' behavior was carried out under constant laboratory conditions. Since direct continuous observation would not be possible for such long period including scotophase, the arena with bugs was recorded by infra-red-sensitive CCD video camera connected to a time-lapse video recorder via VITC generator. Each video cassette represented 24-hours record. Data on feeding of bugs on linden seeds were obtained from videotaped recordings using The Observer 3.0 with Support Package for Video Tape Analysis [5, 6]. Since maximum duration of event recording is 24 hours, each video tape was analyzed separately.

2.2 Data analysis

The obtained data were analyzed with respect to the distribution of feeding activity of bugs within a day. For this purpose methods of time series analysis [1] were applied. Raw observation data files (*.ODF) were used. They were first merged into single file containing information on activity during two weeks. The time data were then converted into angles where a period of 24 hours corresponds to a full turn of 360 degrees. Hence time t is converted to angle ϕ as $\phi = 360t/1440$ where t is in minutes. Mean vector (r) which determines a mean or peak hour, was calculated from time data. Its rectangular coordinates were calculated using the following equations:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\cos\phi_i$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}\sin\phi_i$$

The mean vector length, r, was then calculated as

$$r = \sqrt{\bar{x}^2 + \bar{y}^2}$$

The mean or peak time of feeding activity was derived from a mean angle, $\bar{\phi}$, which was in turn calculated from the rectangular coordinates of the mean vector as

$$\bar{\phi} = \begin{cases} \arctan(\bar{y}/\bar{x}) & \bar{x} > 0 \\ 180^\circ + \arctan(\bar{y}/\bar{x}) & \bar{x} < 0 \end{cases}$$

The mean time $\bar{t}$ is obtained by converting $\bar{\phi}$ back to the time scale. Calculations and plotting of graphs were made using macro written in FUDGIT [2] and GNUPLOT [8] software.

2.3 Statistical test

We further examined whether there is a statistical evidence of one-sidedness or non-randomness of feeding activity. This was tested against the null hypothesis that the activity is uniformly distributed throughout the day. The usual parametric tests for circular data (e.g., Rayleigh's test [1]) require that the individual measurements are independent of each other. This assumption is likely to be invalidated by the fact that data were acquired during two weeks by observing the same individuals. To circumvent this problem we used randomization test [3]. An additional advantage of applying randomization is that it makes no demands with respect to a specified underlying distribution of the data. For this purpose count of the events (start of feeding) observed in each of 24 hours (C_0, C_1, ..., C_{23}) were permuted within the individual bug. If there were no preferred time of activity (the null-hypothesis), any combination of the C's should be equally likely. The null hypothesis was rejected if the proportion of permutations leading to a mean vector equal to or larger than the observed one (called r_{obs}) is less or equal to the level of significance 0.05. Randomization tests were performed using a PASCAL program RanDir originally developed by G. Nachman (Dept. Popul. Biol., University of Copenhagen, Denmark) for testing orientation of organisms [9].

3 Results

The ethograms showing distribution of the feeding activity (Fig. 1) in four individual diapausing brachypterous males of *P. apterus* throughout the first 14 days of their adult life demonstrated that the activity predominantly occurred throughout the photophase.

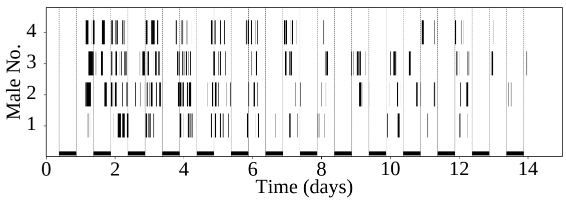


Figure 1. *Ethogram of temporal feeding pattern in four individual diapausing brachypterous males of* P. apterus *during the first 14 days of their adult life. Vertical bars means the numbers and durations of feeding. Vertical dashed lines indicate start and end of scotophase.*

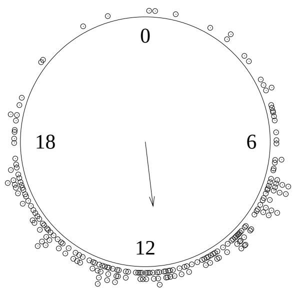


Figure 2. *The circular graph showing the distribution of the onset of feeding within a day in diapausing brachypterous males of* P. apterus*. Arrow indicates mean vector. Note that the circle scale is on a 24-h basis.*

The distribution of feeding within a day and mean vector are shown in the circular graph (Fig. 2). It is obvious that the time of feeding activity is not random. The acrophase (mean time) of the activity occurred at 11:32 MET with the mean vector length r=0.518 (n=205).

The randomization test revealed that the distribution of feeding activity was significantly different from random distribution (N=1000000, P<0.001) giving statistical evidence of directedness in the activity.

4 Conclusions

The described method provides basic chronobiological analysis of behavioral data collected by means of The Observer. The analysis of example data on feeding activity of diapausing brachypterous males of *P. apterus* revealed diel rhythmicity with the peak of activity found in the photophase.

Supported by the Grant Agency of the Czech Republic (Grant No. 206/03/0016).

References:

1. Batschelet, E. (1981) *Circular Statistics in Biology*. Academic Press, London.
2. Lacasse, M.D. (1994) *FUDGIT. A multi-purpose data-processing and fitting program.* User's manual. Version 2.42, McGill University, Montreal, Canada.
3. Manly, B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2nd edn. Chapman & Hall, London.
4. Maxová, A.; Kodrík, D.; Zemek, R.; Socha, R. (2001) Diel changes in adipokinetic and walking activity of *Pyrrhocoris apterus* (Heteroptera) in relation to physiological status and wing dimorphism. *Eur. J. Entomol*., 98, 433-438.
5. Noldus Information Technology (1993) *The Observer, Base Package for DOS*. Reference manual, version 3.0 edition. Wageningen, the Netherlands.
6. Noldus Information Technology (1994) *The Observer, Support Package for Video Tape Analysis*. Reference manual, version 3.0 edition. Wageningen, the Netherlands.
7. Socha, R.; Zemek, R. (2000) Locomotor activity in adult *Pyrrhocoris apterus* (Heteroptera) in relation to sex, physiological status and wing dimorphism. *Physiol. Entomol*., 25, 383-389.
8. Williams, T.; Kelley, C. (1999) *GNUPLOT. An interactive plotting probram.* User's manual, Version 3.7.
9. Zemek, R.; Nachman, G. (1999) Interaction in a tritrophic acarine predator-prey metapopulation system: prey location and distance moved by *Phytoseiulus persimilis* (Acari: Phytoseiidae). *Exp. Appl. Acarol.* 23, 21-40.

Observation of food choice in catering before and after introduction of organic cheese and meat slices

A.A.M. Poelman [1] , C. Glorie [2], J. Mojet [1]
[1] *Agrotechnology & Food Innovations, Wageningen UR, Wageningen, the Netherlands*
[2] *Sodexho Food & Management Services, Capelle aan den IJssel, the Netherlands*

Abstract

Observational research in a real life setting offers great potential for studying eating and drinking behavior, but (reported) applications are still scarce. An exploratory study was carried out in a food service restaurant in Wageningen where the consumer lunch choices for cheese and meat slices were observed. The study was part of a larger study, in which the effect of different marketing strategies on the success of organic meat and cheese slices sales was studied by monitoring sales and individual consumer lunch choices during a 6 weeks period. The objective of the observational study was to determine if the marketing strategy was successful in gaining the consumers' attention. The second objective was to measure the product handling time. It was concluded that the strategy was successful in gaining consumers' attention, but that product handling might hamper repeat purchase. The observational study proved a valuable addition to the other instruments used in the study to understand the consumers' choice process.

Keywords

Food choice; consumer; natural environment; humans.

1 Introduction

Observational research in a real life setting offers great potential for studying eating and drinking behavior, but (reported) applications are still scarce. In a food service restaurant in Wageningen the consumer lunch choices for cheese and meat slices were observed. The observations were part of a larger study aimed at investigating the introduction of organic products in catering. The effect of different marketing strategies on the success of organic meat and cheese slices sales was studied by monitoring sales and individual consumer lunch choices during a 6 weeks period. Camera recorded observations were explored to obtain an insight in the consumers' choice process.

1.1 Objectives

For organic products to be successful, consumers have to consider them while choosing a product. If products are incorporated in the consumers' consideration set, this presumably will be reflected by a prolonged choice process. The objectives of this study were twofold:

- To measure the time to decide for meat or cheese slices before and after introduction of organic slices
- To examine the time for product handling of regular and organic products after the choice was made.

2 Methods and materials

2.1 Subjects

The study was carried out in a food service restaurant of an applied research institute in Wageningen during a 6 weeks period.

Camera observations were made one hour during lunch time on two consecutive days at: three weeks before introduction of the organic products, the week the organic products were introduced and three weeks after introduction of the organic products. The behavior of every subject buying a meat or cheese product was coded, as were the number of visitors of the restaurant.

2.2 Procedure

Camera recorded observations were made which were coded and analyzed using The Observer (Noldus Information Technology, v. 5.0.23, 2003).

Eye and hand behaviors were coded separately. Recorded eye movements included looking at the counter, reading product information and looking away from the counter, e.g. while talking to someone or looking at the menu. Recorded hand movements included touching products, picking up but subsequently putting back products and picking up products. The gender of the consumer and the product choice were also recorded.

2.3 Products

Meat and cheese slices were studied, which are part of a traditional Dutch (cold) lunch. The non-organic meat and cheese slices were part of the regular assortment of the food catering restaurant, and were presented as usual in portion-sized transparent, plastic wrappings. The organic meat and cheese slices were new in the assortment and were presented unwrapped on a white porcelain tray in the same counter. In the introduction week the consumers had the opportunity to taste the products. In the four weeks following the introduction the organic products were available daily to the consumers.

2.4 Data analysis

Only exploratory data analyses were carried out so far. The choice behavior was set to start when a person was looking at the counter of the meat and cheese products and was set to end when the person had put the product on his/her tray and looked away from the counter. A t-test was carried out to analyze the differences in decision time before introduction and one week after introduction, as well as the differences one and three weeks after introduction. Product handling was set to start when the consumer touched the product and was set to end when the consumer had placed the product on his/her tray. A two-way analysis of variance was carried out with product handling time as a dependent variable, and product and week of measurement as independent variables. The Bonferroni test was used as a posthoc test. The analyses were carried out on the first product a consumer had chosen. A level of $p \leq 0.05$ was set for statistical significance. Data were analyzed using SPSS (SPSS 12.0.1, 2003).

3 Results

3.1 Choice behavior

Man and women did not differ in the time it took them to choose a product, so their data were combined in subsequent analyses.

The average decision time before introduction of the organic products was 5.0 seconds (see Figure 1). In the week the organic products were introduced, the consumers needed, on average, one second longer to chose a meat or cheese product (t (1, 276) = -2.1, p = 0.04). Although the time had decreased again three weeks after the organic products were introduced, this effect was not significant.

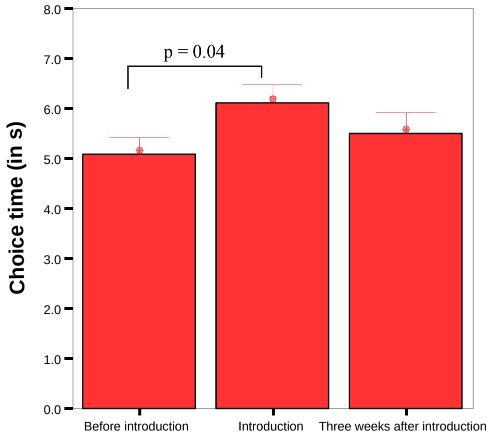


Figure 1. Mean decision time (+SEM) before, during and after introduction of organic meat and cheese slices.

Figure 2 shows that it required less time to chose cheese slices than meat slices. The decision time did not differ for consumers choosing organic or regular cheese. However, the consumers that chose organic meat slices needed more time to decide than consumers that chose regular meat slices.

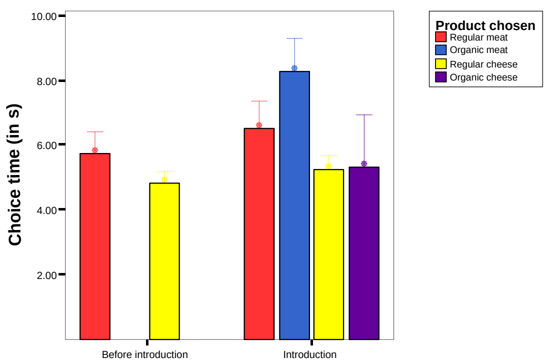


Figure 2. Mean decision time (+SEM) by product before and at introduction of organic meat and cheese slices.

In the introduction week, 27% of the meat buyers chose the organic variety, whereas 5% of the cheese buyers chose the organic cheese.

Four weeks after introduction the percentage of organic meat sales had decreased to 7%, whereas the percentage of organic cheese sales did not change.

3.2 Product handling

Figure 3 shows the time to pick up the product once a decision was made. The differences between the product categories are significant (F (3, 311) = 42.750, p< 0.001). Whereas it took on average nearly 3 seconds to pick up a regular product, the consumers needed over 8 seconds for an organic cheese and over 10 seconds for an organic meat slice portion.

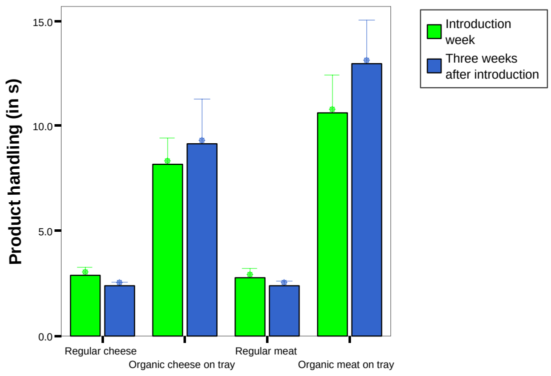


Figure 3. Mean duration (+SEM) to pick up a product by category in the week of introduction and three weeks after introduction.

Post-hoc testing shows that all differences were significant, with the exception of the difference between regular meat and cheese slices. Two reasons could account for the longer time to pick up the organic products. One reason is the packaging, as the organic products were presented unwrapped and needed a fork to lift. A second reason might be the heterogeneous product consistency of the organic products, especially in the case of the meat slices, which often tore during product handling.

4 Conclusion and discussion

The strategy was successful in gaining consumers' attention. The decline in sales of the organic meat products, the low initial sales of the organic cheese products and the long product handling time are factors that need to be addressed to further improve the sales of the organic products.

We found this exploratory observational study a very useful addition to the other instruments used in the study (sales data, questionnaires) to gain an understanding in actual consumer behavior at the moment of product choice. Several elements, such as product handling, were critical to the product success, and would otherwise perhaps have been overlooked.

In subsequent analysis attention has to be paid to differences in attention level during the choice process, as well as the analysis of multiple product choices by one consumer.

Using The Observer Video-Pro to analyze mother-child interactions in language development

C. Quintana and E. Ribes-Iñesta
University of Guadalajara, Mexico

Abstract
Observation of mother child-interactions is a central issue in the analysis of language acquisition. We present data on the observation of mother-child interactions. A girl and her mother were video-taped while they were playing across several ages (4-16 months). Video-recordings were analyzed using The Observer Video-Pro 5.0 and Theme 5.0 with a categorical system developed to observe linguistic interaction patterns and their relation with social and cognitive skills. We discuss the usefulness of the system in identifying mother-child interaction patterns and the relationship of those patterns with the acquisition of cognitive skills.

Keywords
Mother-infant interactions, cognitive development, observational techniques.

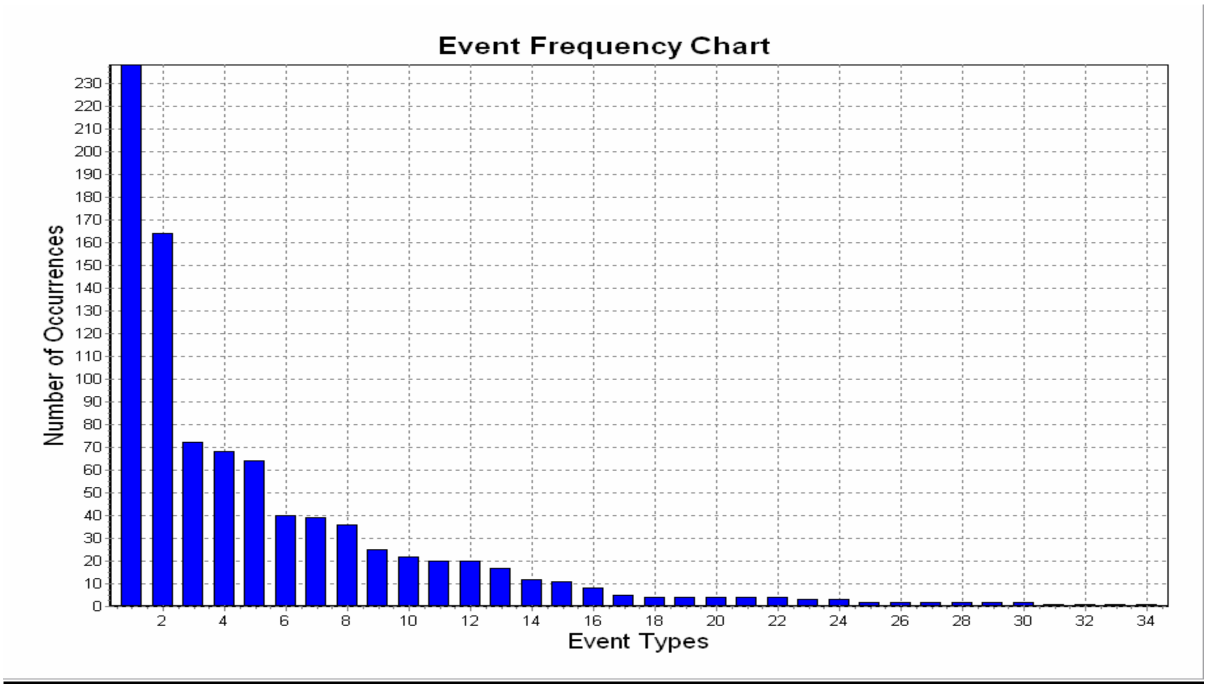
1 Introduction
We used The Observer 5.0 and Theme 5.0 to obtain data based on an observational system designed to analyze the development of language. The system is conceived to measure the interaction of the child with the mother in real time, and in general settings or more specific situations. In this paper we evaluated the dimension of cognitive achievements. This aspect of linguistic interactions has to do with how the mother establishes different adjustment criteria and how the child adapts to them along different stages of his (her) development.

2 Objectives
1. To test the usability of an observational multidimensional system created for the analysis of mother-child linguistic interactions.
2. To identify patterns of development by determining the existence of hidden patterns of cognitive achievements along different stages.

3 Method
We present longitudinal data of one mother-child dyad when the child was between 4 and 16 months old. The mother and child were filmed at home when they were playing together. Video-recordings were analyzed applying a special multidimensional system created for the observation of mother-child linguistic interactions. Categories of the system were configured in The Observer Video Pro and observed on a continuous basis. Hidden patterns of interaction were then established using Theme 5.

4 Results
Data obtained applying the categories of the cognitive achievements dimension of the observational multidimensional system [1] were analyzed using Theme 5.0 in order to find the hidden patterns of mother-infant interactions. Data were categorized according to the task demands settled by the mother in the playing situation. Each achievement of the child corresponds to one demand of the mother. There were 9 categories of achievement (see Table 1).
On the other hand, the child's behavior was recorded as: a) accomplishing the demand (achievement), b) doing a behavior non-compatible with the demand (non achievement), c) doing a mistake while trying to reach the demand (error), or d) rejecting to execute the demand (rejection).
Figure 1 shows the frequency of the types of events resulting from the combination of an achievement category and a modifier (achievement, non achievement, error and rejection). Events of non achievement (non achievement, error and rejection) were the most frequent type of events.
Most of the achievement categories were related to non-verbal tasks specified by instructions from the mother (i.e. throwing a ball, jumping, running). Other frequent demands of the mother were: to pay attention to an object or situation and reproduce movements or actions.
Sequences of different types of events consistently repeated along a data sample are recognized as patterns by Theme 5.0. Patterns identified in the analyzed data showed four different patterns involving the category of paying attention (see Table 2). In addition, there were 5 different trial and error patterns (see Table 3). In these patterns, the same achievement category was repeated twice in a sequence. The child was not able to accomplish a demand in the first place, then, she was successful in accomplishing it in a second trial. For example, the child made a mistake placing a puzzle piece, then the mother prompted her a second time with the same piece and this time the child did it successfully.
Most of the patterns identified were distributed uniformly along the developmental stages analyzed except the patterns involving verbal and manipulating skills. This type of patterns appeared by the age of 12 months.

5 Conclusions
- Event types of non achievement were more frequent than event types related to achievement or adjustment to specific demands of the mother.
- Even though the category of attention was not the most frequent, it was involved in a number of patterns (Table 2 and 3). It seems that getting the child's attention was an important component of many cognitive activities, whether the child succeeded or not in accomplishing different environmental demands.
- Even though the non achievement categories were the most frequent, it did not mean that the child had a poor execution regarding her mother demands (Table 3). It seems that in many mother-child interactions, the child needed more than one trial to succeed in achieving a criterium (trial and error patterns).



Achievement	*Non Achievement*
1. Non-verbal instructions	2. Non-verbal instructions
2. Pay attention	3. Reproducing
6. Reproduction	5. Non-verbal isntructions, error
11. Naming	7. Pay attention
13. Manipulating	8. Naming
15. Choosing	9. Verbal instructions
16. Orientating	10. Naming, error
19. Indicating	12. Inespecific instructions
21. Verbal instructions	14. Verbal instructions, rej
27. Specifying	17. Reproducing, error
	18. Choosing
	20. Non-verbal instructions
	22. Non-verbal instructions, rej
	23. Manipulating
	24. Indicating
	25. Orientating
	26. Reproducing, rej
	28. Choosing, error
	29. Transcribing
	31. Orientating, error
	32. Manipulating, sugerir
	33. Indicating, error
	34. Manipulating, error

Figure 1. *Frequency of different types of events.*

Category	Modifiers
To Pay attention	a) Achievement
	b) Non-achievement
To compute	c) Error
To discriminate	d) Rejection*
To follow instructions a) Non-verbally b) Verbally	
To identify a. Non verbally i. Indicating ii. Choosing iii. Manipulating iv. Orienting b. Verbally i. Designating i.i Naming (generic names, proper names) i.ii Denoting i.iii Assigning ii. Describing ii.i Specifying ii.ii Representing ii.iii Stipulating	
To imagine	
To remember	
To reproduce	
To transcribe	

***Table 1**. Categories of cognitive achievements.*

	Patterns			
	1	**2**	**3**	**4**
Event Type 1	Paying attention (achievement)	Paying attention (achievement)	Paying attention (non achievement)	Paying attention (achievement)
Event Type 2	Reproducing (non achievement)	Paying attention (non-achievement)	Paying attention (achievement)	Manipulating (achievement)

***Table 2.** Patterns involving the category of attention.*

	Patterns				
	1	2	3	4	5
Event Type 1	Naming (error)	Reproducing (non achievement)	Manipulating (non achievement)	Paying attention (non achievement)	Non verbal instructions (non achievement)
Event Type 2	Naming (achievement)	Reproducing (achievement)	Manipulating (achievement)	Paying attention (achievement)	Non verbal instructions (achievement)

Table 3. *Trial and error patterns.*

- A closer analysis will permit to determine the effectiveness of the mother's behavior in promoting the emergence of new kinds of competences and behaviors in the child related to the observed dimensions.

References:

1. Ribes, E.; Quintana, C. (2002). Mother-child linguistic interactions and behavioral development: A multidimensional observational system. *Behavior Analyst Today*, **3**(4), 442-454.

Capturing usability data on-the-go

P.J. Blignaut[1] , T. McDonald[1]
[1] *Department of Computer Science and Informatics, University of the Free State, South Africa*

Abstract
Traditional usability analysis with an observer making notes while watching a user does not allow for multi-user settings and exact time measurements for each task executed. An event-logging tool was developed to capture users' responses while browsing the Internet. Questions are displayed that prompt the user to select the answer from a mutually exclusive set of possibilities or type free text into a box. URLs visited, the time of activation of each page, and the host name of the PC were saved to a MySQL database, allowing for queries that can be used in empirical analyses regarding correctness and speed of task execution and comparisons between various user groups.

Keywords
Usability analysis, event logging.

1 Introduction
Nielsen [3] motivated that testing five users is enough to identify 80% of usability problems in a system. Traditional usability metrics, e.g. percent of tasks completed, time to complete a task, occurrences of errors, users' subjective satisfaction, etc. can be measured in a usability laboratory, testing three to five users [4].
Usability testing is, however, not about the system only but also about the users and the different needs of users with different profiles. In order to pick up differences in usage patterns between users with different cultures, reading and writing skills, IT experience, age, gender, etc. a number of representatives from each homogenous user group is needed. Furthermore, to do inferential statistical analysis, five users in a group are not enough.
Tools such as N6 (formerly NUD*IST) (QSR International), The Observer (Noldus Information Technology), and Morae (TechSmith) allow the researcher to record, manipulate and search the usability data. While these tools are invaluable to analyze usability data from one user at a time, they lack the ability to record all events from all users in a multi-user setup in such a way that the data can be easily analyzed afterwards.
Event logging, on the other hand, is the process of recording user-initiated activities and has been a practice of software usability testing for decades [1]. Event logs look at how website users are using the web interface and generate significant insights without the partiality of observer or participant biases and without the limitations of focus groups or one-on-one user sessions.

2 Event logging tool

2.1 General description
While several event-logging tools are commercially available, they do not always cater for the exact needs of a specific usability test. In a study to determine differences in web-searching patterns between various user groups, 500 users were seated in a computer laboratory in several sessions one after the other [2]. In order to analyze the data it was essential to capture users' demographic details, instruct users to do specific tasks and to obtain feedback on user satisfaction.
A tool was developed to capture users' responses while browsing the Internet. URLs visited, the time of activation of each page, and the host name of the PC were saved to a MySQL database. The tool was designed to fill the bottom 10% of the screen with the Internet browser filling the remaining 90% (Figure 1).

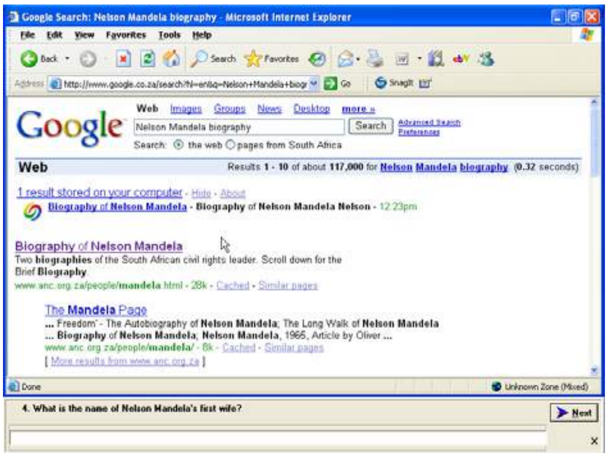


Figure 1. *Screen print of the usability tool together with an Internet browser showing some Google results.*

2.2 Catering for three types of data
Questions are displayed that prompt the user to select the answer from a mutually exclusive set of possibilities or type free text into a box. Three categories of questions can be posted into the database beforehand to accommodate a researcher's needs regarding the specific software application to be tested:

A. Questions to capture a user's profile, e.g. gender, age group, computer experience, etc.

B. Instructions with regard to specific tasks that must be done with the software that is tested. For example, the question visible in Figure 1 refers to the name of Nelson Mandela's first wife. The user must use a search engine, e.g. Google, and surf the web to find the answer.

C. Questions to capture qualitative responses from the user regarding the usability of the website and his/her satisfaction while using it. Instead of providing the user with a paper-based questionnaire that must be entered into a database or spreadsheet afterwards, the questions are posted on the screen, while the user is browsing the Internet, providing him/her with a space to record his/her answers directly into the database.

2.3 System architecture
The tool connects to a MySQL database with a structure as in Figure 2.

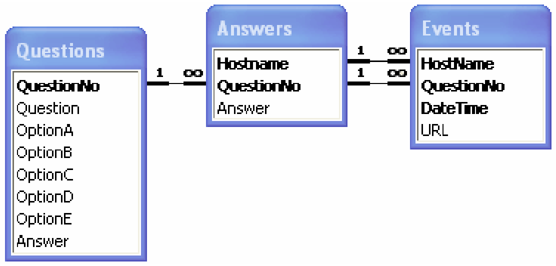


Figure 2. *Database structure.*

This structure allows for usability items in each of the above-mentioned categories to be saved beforehand. Table 1 shows some typical entries in the *Questions* table.

Cate-gory	No	Question	Options				Correct Answer
			A	B	C	D	
A	1	Gender	M	F			
A	2	Home language	Sotho	Tswana	Afr	Eng	
A	3	Previous accesses to the internet	0	1-10	11-20	21+	
B	4	Mandela's first wife					Evelyn Mase
C	5	Did you enjoy the session?	No	So-so	Yes		
C	6	Did you find the questions easy?	No	So-so	Yes		

Table 1. *Typical usability items.*

For the items in Table 1, a typical set of answers can look as in Table 2.

Host	Date/Time of Answer	Question	Answer
PC001	2005/01/24 10:08:42.515	1	M
PC001	2005/01/24 10:08:46.121	2	Sotho
PC001	2005/01/24 10:08:58.208	3	11-20
PC001	2005/01/24 10:15:07.234	4	E Mase
PC001	2005/01/24 10:15:55.325	5	Yes
PC001	2005/01/24 10:16:21.221	6	So-so

Table 2*. Typical entries in Answers table.*

For the question on Nelson Mandela's first wife, the typical events registered could look as in Table 3.

Host	Time	URL	Answer
PC001	10:18:01:031	http://www.google.co.za/	
PC001	10:18:51:343	http://www.google.co.za/search?hl=en &q=nelson+mandela+first+wife	
PC001	10:20:36:046	http://nobelprize.org/peace/laureates/ 1993/mandela-bio.html	
PC001	10:21:37:875	http://www.google.co.za/search?hl=en &q=nelson+mandela	
PC001	10:22:45:234	http://almaz.com/nobel/peace/1993a.html	
PC001	10:23:04:078	http://www.google.co.za/search?hl=en &q=nelson+mandela	
PC001	10:24:38:328	http://www.time.com/time/time100/ leaders/profile/mandela.html	E Mase

Table 3. *Typical entries in Events table.*

2.4 Research facilitation

Unlike the traditional approach to usability testing where a single user is observed at a time by an observer and notes made manually, this tool allows for several users to be seated in a computer laboratory at once. They can be given specific tasks to perform while all web sites accessed together with the exact date and time of access are saved into the database.

The database structure showed in Figure 2 allows for queries that can be used to return the number of URLs visited per user and question, the number of items answered correctly per user, the average time (seconds) taken per question, etc. Users can be grouped according to gender, internet experience, and their reading and writing skills in English.

These query possibilities create opportunities to various empirical analyses. Search patterns and problems experienced can be identified per user group based on gender, culture, experience, etc. Also, usability issues on a specific web site can be identified if the percentage of users who executed a specific task correctly and the average time to do so is compared with that of some experienced and well-trained users.

2.5 Proof of the pudding

The tool was used successfully during research to determine the differences in web-searching patterns between various groupings of students on a university website [2]. More than 500 users were observed during practical sessions of a computer literacy course over a period of two weeks without any assistance or physical presence of a researcher or lecturer.

3 Summary

Event logging can be a valuable alternative form of usability testing. Though it does not allow for facial expressions and/or user frustrations to be captured, it does facilitate inexpensive quantitative analysis of the performance of large numbers of users with little effort from the researcher. If the ethical issues of capturing users' responses without them knowing it are dealt with, the unobtrusiveness can ensure that true real-time data are captured.

The large quantity of data and the quantitative nature thereof allows easy and reliable statistical analyses to be made. Hypothesis testing can be done regarding difficulty of tasks and time taken to complete tasks. Also, results from users of one particular profile (e.g. gender, age group, race, computer/internet experience, etc.) can be compared with results from users of another profile.

References

1. Hilbert, D.M.; Redmiles, D.F. (2000). Extracting usability information from user interface events. *ACM Computing Surveys*, **32**(4), 384-421.
2. McDonald, T.; Blignaut, P. (2005). The effect of cultural differences on the efficiency of searches on a university website. *Proceedings of the 11th International Conference on Human-Computer Interaction*, Las Vegas, Nevada, 22-27 July 2005.
3. Nielsen, J. (2000). Why you only need to test with 5 users. http://www.useit.com/alertbox/20000319.html
4. Nielsen, J. (2001). Usability metrics. http://www.useit.com/alertbox/20010121.html.

Measuring the effect of distractions on mobile users' performance

M. Crease
Human-Web Group, NRC-IIT e-Business, 46 Dineen Drive, Fredericton, New Brunswick, Canada

Abstract
Users of mobile applications in everyday life are required to be aware of their surroundings yet evaluations of such applications often allow the participants to focus entirely on the evaluation task. This is largely due to practical considerations, as the undertaking of an evaluation in an uncontrolled environment - such as a busy street - is fraught with difficulty. These difficulties may be logistical - it may be hard to control the environment and capture data - and ethical – participants may be exposed to injury if being asked to perform a complicated task in a real life environment. This paper describes the usability lab our group has built which will allow us to investigate the effect different forms of distraction have on mobile users performing different tasks under different conditions.

Keywords
Human-Computer Interaction, usability, mobility, distractions

1 Introduction
The use of mobile computing devices - such as handheld computers and next generation mobile phones - is becoming increasingly popular. Mobile Human-Computer Interaction (HCI) is the research field related to the design and evaluation of the interaction between users and mobile computing devices. The nature of these devices, however, implies that a different approach is required for the evaluation of applications running on mobile devices as compared to those running on standard, desktop devices.

A desktop application is typically run in a reasonably quiet environment with few distractions meaning the user is able to pay full attention to the task at hand. A mobile application, on the other hand, is run in a variable environment where there maybe multiple distractions meaning that the user's attention on the task at hand is limited.

The evaluation of mobile applications, however, has typically not reflected this change in usage environment. A recent survey showed that only 41% of research in Mobile HCI evaluated the system being researched [2]. Furthermore, the majority of these evaluations (71%) were undertaken in a laboratory setting with very few of these using special evaluation techniques designed for a mobile application. The authors suggested that this was for two reasons: the background of most practitioners is computer science and so this is the most familiar technique and; the laboratory environment is the easiest to manage.

This paper outlines the design of a Mobile HCI lab designed to allow the evaluation of mobile applications in the most realistic environment. In particular, the paper concentrates on the features of the lab that are designed to distract the experimental participants in a controlled fashion.

2 Background
Kjeldskov & Sage developed a framework which described mobility in terms of physical motion and the amount of attention required to navigate [4]. Using this framework five techniques for simulating mobility in a laboratory environment were developed and evaluated. A sixth technique – which divided the users attention between conscious actions and the use of a mobile system while stationary - was also evaluated. In all six cases the techniques were compared with the results achieved by participants walking in a pedestrian street. The participants reported the highest workload when walking in the pedestrian street but uncovered most usability problems when sitting. The nature of the problems discovered when sitting, however, were mostly cosmetic indicating that mobility is a significant factor in the evaluation process. The authors attempted to mimic a real-life scenario by forcing the participants to follow a continuously differing path through the lab but this had no effect on the results obtained. This may have been due to the nature of the evaluation where the participants followed an evaluator. This meant that the participants were only required to maintain their position behind the evaluator who set the speed and was watching for obstacles.

There are some other examples of evaluation design which consider the mobility of users. Mustonen *et al.* found that reading speed is affected by walking speed although the reading task also has an effect [5]. It was found that the walking speed had a greater impact on the participants' performance when they were reading the text rather than searching the text for a particular pattern. Similarly, Brewster found that users' abilities to enter numeric codes on a graphical user interface was significantly reduced when in a real-life environment as compared to a laboratory environment [1]. Pirhonen *et al.* used a step machine to simulate a user walking when evaluating a gesturally controlled media player [6]. The use of the step machine had the advantage of simulating movement while keeping the user stationary to allow for easy video capture.

An investigation into where evaluations should take place surprisingly showed that there is little or no benefit in undertaking evaluations in the field compared with in the lab [3]. In the study the laboratory evaluation uncovered more usability issues than the equivalent field study; and the field study only uncovered one issue that the lab study did not. This issue concerned the validity of the data entered in a safety critical situation and reflected the fact that the laboratory environment was not as realistic as the field environment. Furthermore, it was found that the lab environment allowed for more control and easier data capture than when in the field. It was also found that the participants in the lab environment uncovered as many context-aware related problems as in the field. This is surprising as it would be expected that the richness of the context in the field compared to the lab would enable participants to uncover more of these types of problems.

3 Lab Design

The previous section outlined research that has been undertaken on mobile evaluation. This research showed that while laboratory based studies are suitable, mobility is a significant factor when evaluating mobile applications. At first glance, this would imply that an open space with scope for participant movement would be sufficient but as Kjeldskov & Sage reported this does not necessarily force the participant to monitor their surroundings to any great extent [4].

It is necessary, therefore, to design the lab in such a way that evaluators are able to distract participants in a controlled fashion. There are three main types of distraction that need to be supported:

Passive Distractions distract the participants but require no active response. An example of such a distraction in real-life is a billboard on a wall.

Active Distractions require the user to respond in some way. The required response will vary on the distraction. A mobile phone ringing, for example, may require the user to answer to the phone whereas a lamp post will require the user to navigate round it.

Interfering Distractions are a third class of distraction - which may be passive or active - and interfere with the user's interaction with the mobile device. If, for example, the application provides audio feedback to the user a noise such as a passing car or ongoing conversation would interfere with that interaction.

4 Lab Implementation

The lab has been built in the basement of NRC-IIT's Fredericton site. The lab consists of a large empty space (8.65m x 17.3 m) which can be sub-divided using portable partitions. This space enables the experimental participants to be mobile during evaluations. There are two observation offices along one side of the lab, each of which has one way glass.

Visual distractions can be produced using the six roof mounted projectors in the lab. The projectors are mounted on a central rail so that they can be moved through the lab and rotated to project on the appropriate wall. The lab walls and doors are painted white and there are white blinds which can be pulled over the observation windows meaning the projectors can be used all around the lab.

Audio distractions are provided using a 7.1 surround sound system. As with the projectors, the speakers are ceiling mounted meaning the lab's floor space is obstruction free. An indoor location system [7] is used so that the presentation of the distractions can be done only when the user is in the appropriate location.

5 Future Work

The lab will enable us to develop a series of guidelines regarding the effectiveness of different forms of distraction. This will enable future evaluations of mobile applications to be performed under the most appropriate conditions.

These first of these evaluations will require the participants to monitor distractions projected onto the laboratory walls while performing their primary task. Participants will be required to acknowledge certain distractions by pressing a button on the user interface of their mobile device. The effect of different forms of distraction - and the absence of distraction – on task performance will be compared. Two different forms of distraction will be evaluated. The first form - single characters - will require the participants to look for '*'s being projected. The second form, photographs of street scenes, will require the participants to look for hazards – in this case moving vehicles. This research will indicate whether abstract distractions are as effective as realistic distractions. Future evaluations will investigate the effect of audio distractions and the different types of distraction as described in Section 3.

6 Conclusions

This paper outlines the design of a mobile HCI lab which allows evaluators to distract users in a controlled fashion. This is significant as previous research has shown that mobility on its own is insufficient to accurately simulate a real-life context while undertaking evaluations in the field is often impractical.

Acknowledgements

This paper would not have been possible without the tireless efforts of Dr. Jo Lumsden who was instrumental in the design of the lab and Danny D'Amours without whom the hardware would not have been sourced and installed.

References:

1. S. Brewster, "Overcoming the Lack of Screen Space on Mobile Computers," *Personal and Ubiquitous Computing*, vol. 6(3), pp. 188.205, 2002.
2. J. Kjeldskov and C. Graham, "A Review of MobileHCI Research Methods," in *Proceedings of the 5th International Mobile HCI 2003 conference*. Udine, Italy, 2003.
3. J. Kjeldskov, M. B. Skov, B. S. Als, and R. T. Høegh, "Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field," in *Mobile Human-Computer Interaction - Mobile HCI 2004*, vol. 1, S. Brewster and M. Dunlop, Eds. Glasgow, Scotland, 2004, pp. 71-73.
4. J. Kjeldskov and J. Stage, "New Techniques for Usability Evaluation of Mobile Systems," *International Journal of Human Computer Studies (IJHCS)*, vol. 60, pp. 599-620, 2004.
5. T. Mustonen, M. Olkkonen, and J. Häkkinen, "Examining Mobile Phone Text Legibility while Walking," in *Proceedings of CHI 2004*. Vienna, Austria, 2004, pp. 1243-1246.
6. A. Pirhonen, S. A. Brewster, and C. Holguin, "Gestural and Audio Metaphors as a Means of Control for Mobile Devices," in *Proceedings of ACM CHI 2002*. Minneapolis, MN, 2002, pp. 291-298.
7. C. Randell and H. Muller, "Low Cost Indoor Positioning System," in *Proceedings of Ubicomp 2001: Ubiquitous Computing*, G. D. Abowd, Ed. Atlanta, Georgia, 2001, pp. 42-48.

Bimanual Input and Patterns of User Behavior

G.E. Evreinov
Department of Computer Sciences, University of Tampere, Tampere, Finland

Abstract
The assumption that two-handed manipulation save time is not always the right way for interface design. This paper describes the results of the study of using two analog buttons in a target acquisition task illustrating behavioral principles for two-handed interfaces. The results showed that despite the apparent contradiction of the motion artifact to the one-to-one mapping principle, Y-axis changed with the left hand whereas the X-axis was controlled by the right hand, the subjects normally did not notice the artifact as such. This can be explained by the nature of the input mapping designed to accommodate the genuine structure of bimanual manipulation.

Keywords
Two-handed interfaces, integral behavioral pattern.

1 Introduction
The most of input techniques were designed to support one-hand interaction. As it was shown in [2], the preferred hand articulates its motion relative to the dynamic *frame-of-reference* determined by the non-preferred hand. This has an immediate effect upon both the behavioral strategy and the way the subjects could perform the task efficiently in two-handed interaction. Indirect pointing on a computer screen with an input device may or may not follow the patterns found in direct hand pointing at physical targets [1]. The mapping between cursor motion and input device displacement is often a complex transfer function of the control-display ratio, which may further increase the complexity of the hand eye coordination in a target acquisition task. In the case of computer mice, most of them have non-linear acceleration schemes. The complexity of two-handed cooperation grows exponentially when a pointing task has to be implemented in 3D environment [4]. Thus, there is a challenge to explore a number of research questions on the behavioral level with subsequent application of the results in the design of dynamic multimodal interfaces.

2 Method Design
To simulate spatial-distributed input in bi-manual interaction and facilitate interpreting the data gathered in a target acquisition task a special input device, analog buttons, were designed. The distribution of X-Y coordinates is the simplest way to estimate the behavioral features in two-handed coordination. As such a way can be considered as an “asymmetric dependent” task [2], and according to Kabbash [3], is more suitable and efficient way to simultaneously operate independent controls.

2.1 Analog buttons
The pointing device used in the experiment was a set of two analog buttons which were developed using of silicon tube. It provided stability of mechanical parameters whereas dynamic displacement was similar to a normal digital key (not less than 4 mm); the current amplifier compensated nonlinear characteristics of optical force-displacement transducer. An overall view and construction features of the buttons are shown in Figure 1.

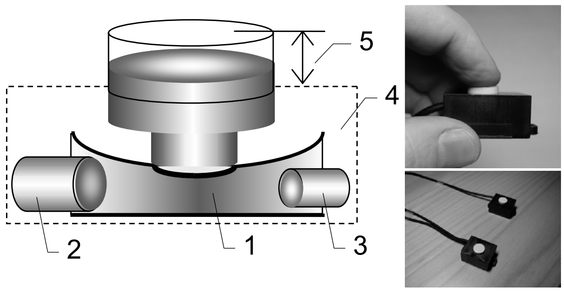


Figure 1. *Overall view and construction features of the analog buttons. 1 - silicon tube; 2 - photodiode; 3 - light emitting diode; 4 - box; 5 - nominal displacement.*

2.2 Apparatus and Procedure
The experiment was conducted on a Pentium III 800MHz desktop PC with a 19" monitor having a screen resolution of 1024 by 768 pixels. The software tools were written in Microsoft Visual Basic 6.0 SP6 under Windows 2000. The experimental software was used to present the task and capture the data from the pointing device. The pointing device comprised of two identical analog buttons connected to joystick port. The button in the left hand was used to control the Y-axis of the cursor position to move the cursor downwards, whereas the button in the right hand controlled the X-axis to move the cursor rightwards. This setup was used for all the subjects. The nominal displacement of the analog buttons was equal to 4 mm (Figure 1). It was translated to 256 pixels of the test-window and scaled by three times to fit to the monitor screen. Thus, ±5-pixels of the test-window corresponded to ±200 microns of physical button displacement. This implies that the low resolution and accuracy of button displacement could be enhanced and compensated by means of visual feedback loop.

Twenty unpaid volunteers (15 males and 5 females) took part in the test. Two participants were left-handed and two reported they used both hands equally. Other 16 participants considered themselves right-handed. All participants used computers on a daily basis. None had prior experience with the analog buttons.

The task consisted of capturing one of 49 circular targets arranged in a square grid 7 by 7. A screen distance between the centers of adjacent targets in the grid was 36 mm. Only one target was visible on the screen at a time. The targets were presented in a random order. The task was explained and demonstrated to participants and a warm-up block of trials was given. In each trial, a circular target of diameter 4 mm appeared on the display. The cursor had a shape of a cross-hair tracker. Participants moved the cross-hair tracker by manipulating with two analog buttons. The goal was to select the target by moving the tracker over it. The selection occurred only when the tracker remained within the target area for an uninterrupted period of 300 ms (dwell time). Only one trial was performed on each target, so the session consisted of 49 trials. In each trial, participants were given a limit of 10 seconds to complete selection of the target. If they did not acquire the target within the time limit, an error was recorded for the trial. An error was accompanied by a negative sound beep. The next target was presented after a delay of 1 second. Participants were instructed to accomplish the task as quickly as possible.

3 Results and Conclusion

The results showed that despite the apparent contradiction of the motion artifact with the spatial-distributed input to the one-to-one mapping principle, the subjects normally did not notice the artifact as such. This can be explained by the nature of the input mapping designed to accommodate the genuine structure of bimanual manipulation.

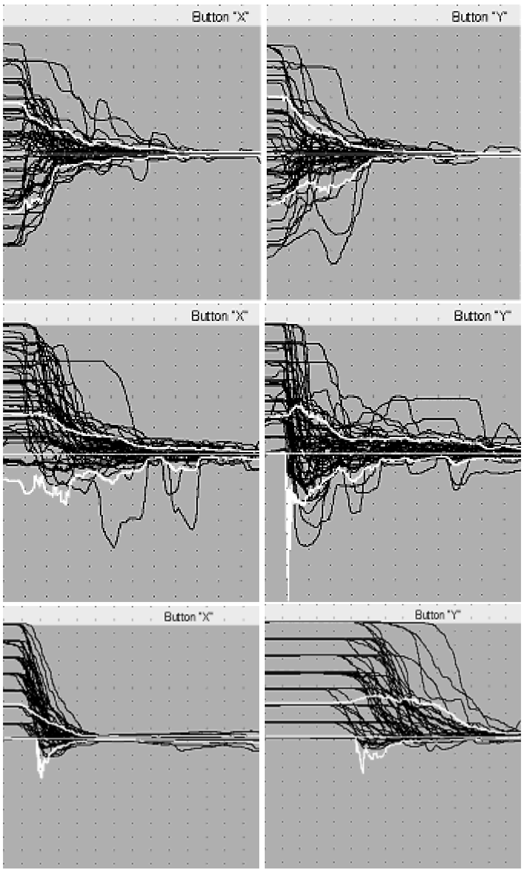


Figure 2. Plots of 49 trials and integral behavioral patterns (white lines) of the left-handed subject (middle pictures), right-handed subject (bottom pictures) and the case of ambidexter (upper pictures). The time scale (grid on X-axis) is 300 ms. The displacement scale (grid on Y-axis) is 12.8 pixels.

Our observations also revealed that the integral behavioral pattern (Figure 2) averaged on 49 trials could be divided into three phases. The first is the motor programming phase (Figure 3). It is the time span between the onset of the movement and the time, when either of the coordinates cross the aiming-on-target zone (32 pixels in the current study). The final point of the motor programming phase coincides with the end of the sudden movement of the cursor along the dimension controlled by the dominant hand. The second phase is the closed-loop phase during which the subjects manipulate the buttons to get the cursor over the target. When either of the coordinates gets as close to the target as 6 pixels away from its center, the third phase begins, the final target acquisition phase. During this phase, the other coordinate should be caught up, so that the cursor finally gets over the target.

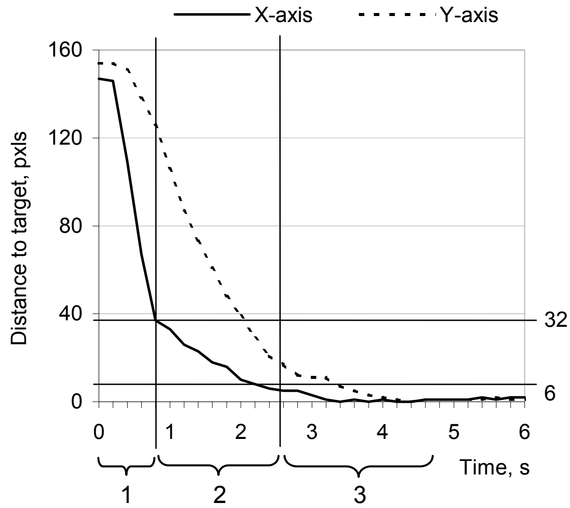


Figure 3. Generic behavioral pattern: 1–motor programming phase; 2 – aiming on target phase; 3 – target acquisition.

The most of people prefer to employ their dominant hand and the most of input techniques were designed to support one-hand interaction. This paper describes the results of the study of using two analog buttons in a target acquisition task illustrating behavioral principles for two-handed interfaces. The analog buttons developed allow the user to manipulate the cursor position across the entire area of the screen with a resolution of 256 by 256 pixels by having only 4 mm displacement. Visual feedback loop compensates both the lack of low resolution of the input device and the low accuracy in finger dexterity.

The precision of pointing in a spatial-distributed task with bi-manual input for novices is quite difficult task which may contradict or may not with previous user experience. Still, cooperation between the various motor systems might be facilitated when the input mapping would be designed to be adaptive to the personal features in bimanual interaction. The integral behavioral pattern can be recorded in a simple game-like testing procedure. The results can be used to adapt parameters of the input technique in appropriate way.

Supported by the Academy of Finland (grant 200761 and 107278).

References

1. Chatty, S. (1994) Issues And Experience In Designing Two-Handed Interaction. *CHI'94,* Boston, Massachusetts US, ACM SIGCHI, 253-254.
2. Guiard, Y. (1987) Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. *Journal of Motor Behavior*, 19 (4), 486-517.
3. Kabbash, P, Buxton, W. (1994) A. Sellen. Two-Handed Input in a Compound Task. *CHI' 94*, Boston, Massachusetts US, ACM SIGCHI, 417–423.
4. Wang, Y. and MacKenzie, C.L. (1999). Effects of orientation disparity between haptic and graphic displays of objects in virtual environments. *INTERACT '99*, Edinburgh, Scotland, 391-398.

A system for monitoring abnormal operator behaviors

Hunszu Liu
Department of Industrial Engineering & Management, Ming Hsin University of Science & Technology, Hsin Chu, Taiwany

Abstract
In this study, a PC-based computer program is developed, based on the technique of Human Error Identification System Tools (HEIST), to help industrial managers create a monitoring system for abnormal operator behaviors. The abnormal behavior is defined as the deviation from standard operation. The standard operation procedure is used as input data of this program. An interactive process is initiated to guide the users to develop the potential abnormal behavior tables, which are transformed into abnormal behavior observation checklists. Work Sampling Technique is incorporated to generate the schedule to collect the line workers' abnormal data. An alarm level of abnormal behavior is set through the Control Chart Technique. An industry case is carried out to illustrate the application of this system.

Keywords
Abnormal behavior, human errors, human reliability

1 Introduction
Abnormal operator behaviors can create product quality or system safety problems for industrial survival. These abnormal behaviors were controlled through developing appropriate measurements based on the analyzing results of corresponding outcomes. To ensure the effectiveness and efficiency of the control mechanism, this plan-do-check-action cycle is repeatedly initiated by encountered problems, which usually already cause some damage to system operations. As a result, extra costs occur.
In this study, a common PC-based computer program is developed, based on the technique of Human Error Identification System Tools (HEIST) [1,2], to help industrial managers create a monitoring system for abnormal operator behaviors. The abnormal behavior is defined as the deviation from standard operation. The standard operation procedure is used as input data of this program. An interactive process is initiated to guide the users to develop the potential abnormal behavior tables, which are transformed into abnormal behavior observation checklists. Work Sampling Technique is incorporated to generate the schedule to collect the line workers' abnormal data. An alarm level of abnormal behavior is set through the Control Chart Technique [3].

2 System architeture and functional analysis
The monitoring system comprised three modules, including the standard operation analysis module, the abnormal behavior identification module, and the monitoring plan generation module. The user first obtained the standard operation procedure through the job description and working instruction manual. It is the user's responsibility to make sure that the procedure is the most update version and conforms to operator working order. It is not uncommon that the line workers skip or change the working method. The better way is to ask senior worker to make sure the procedure is correct. Then, this standard operation procedure is input into the standard operation analysis module (Figure 1).
The operation analysis module adopts the Human Error Identification System Tool (HEIST) and generates potential abnormal behaviors table through interactive approach. This table is then input into the abnormal behavior identification module where the potential abnormal behaviors are weighted and ranked through the pair-comparison process performed by analyzer. The top ranking abnormal behaviors where the sum of their weights comprised 80% of total weights are identified as the critical abnormal behaviors. This information is input into the monitoring plan generation module. The checklists are generated based on the results of critical abnormal behaviors. An observation schedule is generated through the work sampling function. The analyzer uses theses checklists to observe the behavior performed by line workers according the scheduled time and record the nonconforming behavior [4]. These data are used to generate the U control chart through the control chart function.

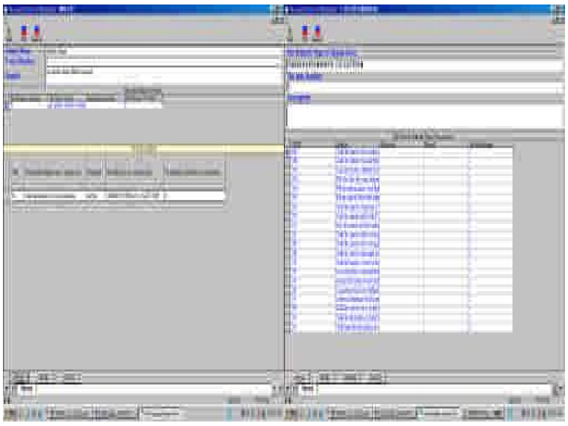


Figure 1. Standard operation analysis module

3 Illustration
An industry case was conducted in a semiconductor plant. The facility studied in this case was gas cabinet (Figure 2). Gas cabinet is a metal enclosure that is intended to provide local exhaust ventilation, protection for the gas cylinder from fire without the cabinet, and protection for the surroundings from fire within. The silane (SiH_4) gas cabinet is chosen as the study object because it has been widely installed in semiconductor plants and several accidents involved in the past.
This study investigates the 3-CYL Cabinet (2-A139 & M102FH Panels) manufactured and assembled by Air Liquid Electronic (ALE). Standard operation procedure was documented based on the on-site operation manual, observation and interview results. The observation subject is a male supervisor with five years of related experiences. The SOP can be divided into four stages: pre-purge, cylinder change, post purge and process flow stages. Pre-purge stage consists of 10 steps. Cylinder change stage consists of 13 steps. Post purge consists of 3 steps. Process flow consists of 8 steps. Sets of abnormal

behavior tables were generated by repeating the interactive process until all 34 steps were analyzed (Table 3.1).

3.1 Sample results of abnormal behaviors

Steps	Abnormal behaviors
1. Wear correct personal protection equipment	Perform too early
	Perform too late
	Forget to perform
	Incomplete
	Wrong usage
	Unable wearing
2. Proceed CYLINDER CHANGE process following instructions on control panel	Forget step
	Wrong cylinder
	Wrong step
	Unable complete step
3. Remove the empty cylinder and its in use label, install the full cylinder	Perform too early
	Perform too late
	Forget to perform
	Wrong cylinder
4. Check the cylinder valve DISS connector	Perform too early
	Perform too late
	Forget to perform
	Forget steps
5. apply specific wrench and tight the D I S S connector	Perform too early
	Perform too late
	Forget to perform
	Incomplete steps
	Forget steps
6. Perform high pressure leakage check following instructions on panel	Unable complete step
	Incomplete steps
7. Perform post purge process following instructions on panel	Unable complete step
	Incomplete steps
8. Confirm first purge process function normally	Perform too early
	Perform too late
	Forget to perform
	Neglect other related status
	Fail to consider special circumstances
	Presume system status normal
	Wong record
	Incomplete
9. Confirm the first purge cycle normal	Perform too early
	Perform too late
	Forget to perform
	Distraction
	Incapacitated
	Unable confirm normal situation
	Fail to consider special circumstances
	Presume system status normal
	Neglect other related status
10. Exchange the open tag with close tag on the cylinder	Perform too early
	Perform too late
	Forget to perform
	Wrong procedure
	Incomplete procedure
	Poor communication between ship

Figure 2. *Silane gas cabinet*

The program scheduled twenty-five observations during one month period, where total 175 abnormal behaviors were recorded and plotted in the U control chart (Figure 3). Since all points are plotted inside the control limits, it seems reasonable to conclude that there is no need to revise the control chart. This chart can be used to monitor future operations.

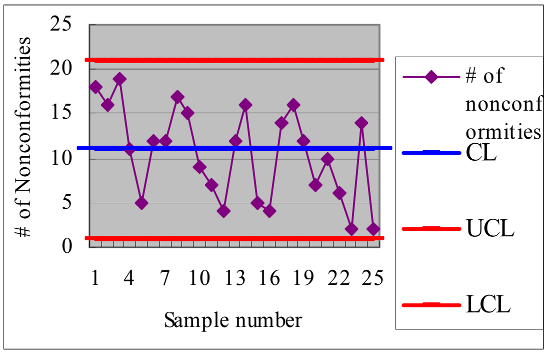


Figure 3. *Observation results within one month period*

4 Discussions

Industry managers need tools to help them establish and maintain a reliable and safe working environment. This study illustrates an approach to help plant managers understand how to get better control of their workers' performance. Company managers can also set the control limits as the target value for improving safety status. This tool can be used as the necessary approach to supplement manufacturing decision process.

Supported by Ming Hsin University of Science & Technology, Taiwan (Grant N. MUST94-IEM-012)

References

1. Kirwan, B. (1998), Human error identification techniques for risk assessment of high risk systems-Part 1：review and evaluation of techniques, Applied Ergonomics, Vol.3, pp. 157-177.
2. Kirwan, B(1994), A Guide to Practical Human Reliability Assessment, Burgess Science Press, Great Britain.
3. Montgomery, Douglas C. (2001), Introduction To Statistical Quality Control, 4th edition, John Wiley & Sons, Inc., New York.
4. Strater, & Bubb (1999), Assessment of human reliability based on evaluation of plant experience：requirements and implementation, Reliability Engineering and System Safety, Vol. 63, Issue 2, pp. 199-219.

Radiotelephony understandability verification using a program distributed via Internet

P.J. Hoogeboom and G.D.R. Zon
National Aerospace Laboratory NLR, Amsterdam, The Netherlands

Abstract

In the continuous strive for increased flight safety levels, an experiment to verify the understandability of aeronautical radiotelephony messages has been conducted. The main methodological problem for this experiment concerns obtaining a sufficiently large number of pilots with different linguistic backgrounds. Participants are scattered around the globe and have busy time schedules. To solve this dilemma, a dedicated PC program has been created which guides the participants through the experiment. The program can be downloaded from the Internet and allows participants to perform the experiment at their own pace and at a moment when they have time.
The results of the experiment are promising. Almost 300 specialists have taken part in the experiment, most with positive reactions. However it was found that it takes considerable effort and time to reach the intended audience. Best results were obtained using the existing pilot-associations.

Keywords

Aviation radiotelephony understandability,
off-line Internet experiment

1 Introduction

An experiment to verify the understandability of aeronautical radiotelephony messages has been conducted. The problem addressed concerns the pronunciation of the one hundred number either as "*one-zero-zero*" (International Civil Aviation Organization – ICAO – standard) or as "*one-hundred*". This problem has been subject to considerable debate, resulting in changes in the prescribed standard in 1987 and 1995. Current status is that some countries have not adopted the latest standard in their national laws [2]. The experiment is aimed at getting an indication of the risks of misunderstanding versus the used phraseology.
The main methodological problem encountered with this experiment concerns obtaining a sufficiently large number of pilots with different linguistic backgrounds. Potential participants are scattered around the globe and are almost always on the move. To overcome the problem with the availability of the participants, a dedicated PC program has been created which guides the participants through the experiment. The program can be downloaded from the internet and allows the pilots to perform the experiment whenever they have time and wherever they are. Results are sent back to the central location using an email attachment, removing the need to have Internet access while performing the experiment. From an experimental point of view, the advantage is that participants can conduct the experiment in parallel, reducing the overall time required to have a large group to participate.

2 The experimental set-up

2.1 General considerations

In the underlying study [3] two experiments are performed. The first experiment uses a computer test (program) that was distributed amongst pilots via Internet and CD-ROMs. The major drawback of this approach is that the experiment is less controlled (e.g. the used computer environments may differ considerably, and interruptions – like telephone calls – may occur during the execution of the experiment. Due to technological limitations only pilots' computer inputs like key-presses and mouse movements can be recorded. Recording of audio 'read-back' – which is an essential part of the normal communication scheme – is not possible with the required accuracy and result-file limitations. To accommodate this shortfall, a second experiment is performed within the study using a smaller number of participants. Emphasis of this second experiment is on the difference between clearances and the resulting read-back and values inserted in the computer program, allowing distinguishing between perception (hearing) and execution type of errors. This paper focuses on experiences gained with the first experiment.

2.2 Experiment description

The experiment consist of the following main parts:

1. Introduction text with experiment explanations
2. Biographic questionnaire
3. Domain relevant task: flying a simulated aircraft using an autopilot) in which the participant has to react on the audio messages. After the training session, two or four test-sessions are to be completed (number of sessions is based on pilot preference). Each test-session consist of five 'runs' and uses either the 'one-zero-zero' or the 'one-hundred' phraseology. Each run uses the recorded voice of one specific controller, each with a different linguistic background. The order in which the controllers' voices are presented is assigned randomly. To ensure a medium to high workload level, a secondary task based on the monitoring task of the Multi-Attribute Task Battery (MATB [1]) was added to the set-up, see also figure 1.
4. Final questionnaire with options to provide additional textual feedback. In total nine questions were asked to probe the correctness of the experiment and the pilots' preference for the phraseology.

The overall experiment took about two hours to complete (four sessions). Besides recording the pilots' responses in the experiment, the pilots were asked to provide their subjective perception of their own mental workload (a.o. RSME & NASA TLX), their performance and the appreciation of the two phraseologies in general.

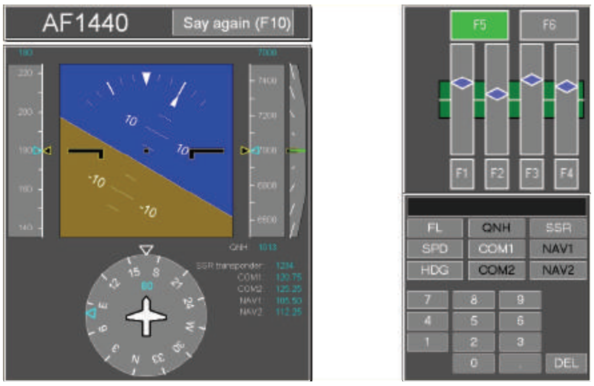


Figure 1. *Main screen of the experimental task: on the left is the 'flight display', the right upper corner shows the 'monitoring task', while the right lower corner shows the panel to control the autopilot.*

2.3 Computer program

The experiment is contained in one dedicated program, allowing only a sequential execution of the main experiment phases. The program itself is created using Borland's Delphi 5. PC requirements and installation are kept to a bare minimum. The total program including installer is smaller than 8 Mbyte and can be executed using Windows® from '98 till XP.

3 Experiences

3.1 Reaching the intended audience

It proved to be a challenge to reach the intended audience, although the program was 'advertised' in several Internet forums. Good results were obtained whenever pilot associations were involved and advertisements could be made in their local mailings. Having some enthusiastic persons, energizing relevant people, also helped tremendously. In total over 300 reactions were received, with a representative the pilot experience distribution and flown aircraft types. Also pilot reactions to the program were positive in the sense that it was regarded as domain specific, addressing a real issue, contributing to the on-going discussion regarding improving safety, interesting training tool, etc.

3.2 Audio recordings

The program required audio samples from different ATC controllers. The actual clearances given are depending on the simulated flight scenario, therefore it was necessary to split the audio samples in command sequences with the numerical values added to them. Since multiple aircraft are used, each with a different callsign (aircraft identifier), they are also recorded separately. By recording a large number of clearances, presented randomly by another computer program, it was possible to approximate the required normal 'sentence melody'. For the playback, the samples like callsign, instruction and numerical value are combined as needed. In total, this scheme worked fine.

3.3 User interface experiences

To assure that reaction times between different users and conditions would be comparable, the program had to be controlled only by mouse selections. The alternative, forcing only keyboard selections (e.g. the use of the function keys for the monitoring task), is implemented but was regarded as less optimal.

The screen layout adapts itself to the user selected screen resolution in order to provide all users with a similar layout and approximately similar distances between control surfaces. Therefore dynamic scaling is applied to all graphics and sensitive screen areas. To maximize the screen area, the windows' taskbar is disabled if the auto hide function is not turned on. However, one flaw remained. Modern laptops do not always adhere to the 4-3 relation between width and height. Therefore the scaling routine for future releases should take this into account (e.g. by reducing the visible area to a rectangle adhering to the required or allowed dimensions).

3.4 Randomization of experimental conditions

In the experiment it is essential to randomize the order of the controller voices, the order of the 1-0-0 versus 100 condition, the used callsign for the own aircraft and the actual flight scenario. For all settings an independent randomization routine is used. Due to the relatively large number of participants, this scheme worked out fine. Some small differences in the number of occurrences of specific conditions are found, but the difference itself is not thought to impact the statistical results.

3.5 Transmitting results using email

The intention of the program is to be able to be used off-line. Therefore the test result has to be transmitted later to the analysis database. At the start of the development the use of email was selected since this allows the participant to control what is being sent in a transparent way. Due to the ever-increasing threat of viruses, it is not possible to easily attach the results-file to an email message for the wide range of email clients expected. The file with results was unfortunately too large to include it in the main body (several clients have some limitations regarding the amount of data accepted through the HTML "mailto" command). Therefore the program provides the participant with detailed instructions to send the results. If done so, the participant gets rewarded with an email containing the personal results, and is added to the list to obtain a copy of final results. This scheme proved to be too difficult for some of the participants. It is therefore better to provide an automated means to send the results (e.g. through an automated FTP client-server solution). At least required user interaction should be minimized whilst there should be an option to provide additional comments and requests.

Supported by EUROCONTROL (Level Bust – Level Clearance project, TRS 065/4)

References:

1. Comstock, J.R. & Arnegard, R.J. (1992) The multi-attribute task battery for human operator workload and strategic behavior research. *NASA Technical memorandum 104174*.
2. Henriques, P. (2004) The role of R/T communications on level busts. *EUROCONTROL HUM/DIS*
3. Zon, G.D.R., Roerdink, M., and Hoogeboom, P.J. (2005) Does ATCo phraseology influence the likelihood of level busts? *NLR contract report.*

Using the Noldus Observer® to assess the reactivity of video recordings of medical interactions

Louis Penner[1,2], Melissa Franks[1,2,3], Terrance Albrecht[1,2], Tanina Foster[1], and John Ruckdeschel[1]
[1] *Communication and Behavioral Oncology Karmanos Cancer Institute, Wayne State University, Detroit, MI, U.S.A.*
[2] *Department of Family Medicine, Wayne State University, Detroit, MI, U.S.A.*
[3] *Institute of Gerontology, Wayne State University, Detroit, MI, U.S.A.*

Abstract

Video recording provides an objective record of the content of medical interactions. However, there has been concern that cameras may be reactive measurement devices that alter what normally transpires during the interaction. This study addressed the potential reactivity of cameras in medical interactions. Twenty interactions between patients and medical oncologists were video recorded and coded for seven camera-related behaviors (e. g., looking at, talking about, or gesturing toward the camera). Eight of 20 patients showed none of the behaviors. Among the other patients, camera-related behaviors were infrequent and, on average, constituted about 0.1% (one-tenth of one percent) of total interaction time. Camera-related behaviors occurred most often in early stages of interactions, when the physicians were absent from the room; they virtually disappeared in later stages when physicians were present. Only two physicians showed camera-related behaviors. Results suggest that video recording can provide nonreactive means of studying medical interactions.

Keywords

Reactivity, video recordings, real time data capture

Introduction

Video recording offers a powerful tool for understanding what transpires in medical interactions[1]. Combined with coding software such as The Observer Video-Pro® (Noldus Information Technology) video recordings can produce a rich, detailed record of the frequency, duration, and sequencing of key verbal and nonverbal behaviors that occur during the course of the interaction. However, there is a widely held belief that the presence of a camera is reactive and alters the behavior of the parties to the interaction[2]. The first purpose of the present study was to determine the extent to which patients and physicians displayed behaviors that would indicate awareness of and sensitivity to cameras present during an oncology consultation. The second purpose was to study whether the incidence of such behaviors among patients was moderated by the passage of time and/or the presence/absence of the physician.

Methods

Video recordings of interactions between 20 patients (11 males; 9 females; average age 58.26) and seven physicians (all male) at two large cancer centers were randomly selected from a set of recordings of 140 patients who had formally consented to be videotaped while they met with their oncologist (who had also consented). The video recording was done via remote control cameras that were encased in vinyl cylinders and placed in the examination rooms in full view of the patients. The Observer Video-Pro system was configured to allow coding of the frequency and duration of the behaviors of interest as judges watched JPEG files of the interactions. Data were exported to an SPSS spreadsheet for analysis. The seven behaviors selected to indicate awareness of and/or sensitivity to the cameras were: Looking at, talking about, gesturing towards the cameras, whispering, hiding from or obstructing the cameras' view, and self-reflective behaviors (e.g., fixing one's hair, adjusting one's clothing). These behaviors were coded in The Observer configuration as distinct behavioral classes with no modifiers.

Results

Judges were able to reliably code the frequency of the seven behaviors of interest (average Cohen's Kappa .69). Eight of the twenty patients (40%) and five of the seven physicians (71%) did not show any camera-related behaviors. Among the patients/physicians who did react to the camera, looking at the camera was the most common behavior, but no behavior consumed more than 0.2% (two-tenths of one percent) of the time these patients/physicians spent in the examination room. (See Figure 1).

Sixty-three percent of the behaviors occurred during the first quarter of the interactions, and within this quarter, the behaviors declined significantly across time, $\underline{r}$ (17) = -. 53, $\underline{p} < .05$. Significantly more of the behaviors (76%) occurred when the physician was absent from the room, $z=3.67$, $\underline{p} < .05$. During the last two quarters of the interactions, only one camera-related behavior occurred when a physician was in the room. (See Figure 2). However, the time by physician present/absent interaction was not significant.

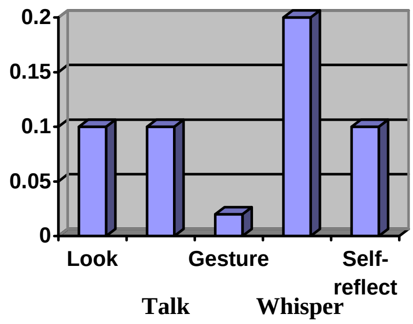


Figure 1. *Percentage of Time in Camera-Related Behaviors*

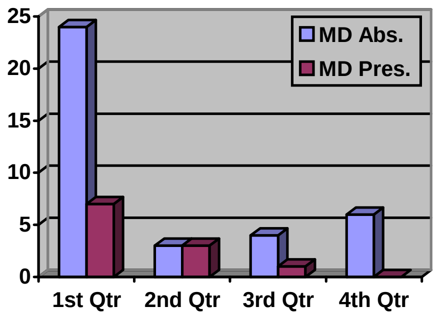


Figure 2. *Frequency of Camera-related Behaviors as a Function of Time and MD Presence/absence*

Discussion

Camera-related behaviors were relatively infrequent, lasted for relatively short periods of time, and constituted an extremely small percentage of the time both patients and physicians spent in the examination rooms. These behaviors declined over time and were infrequent when the physicians were in the room. These findings do not prove that the camera was a nonreactive stimulus, but they do show that video recordings are at most minimally reactive. It is believed that camera reactivity is minimized when patients/physicians cannot see the camera itself and are engaged in important medical interactions. Our results suggest that ecological validity may not be compromised during real time data capture of interpersonal interactions in medical settings, and demonstrate the utility of electronic coding systems such as The Observer Video-Pro. This increases their utility in settings such as medical interactions.

Supported by: National Cancer Institute Grant RO1 CA75003-01A3 (Terrance L. Albrecht Principal Investigator)

References

1. Riddle, D. L., Albrecht, T. L., Coovert, M. D., Penner, L. A., Ruckdeschel, J. C., Blanchard, C. G., Quinn, G., & Urbizu, D. (2002). Differences in audiotaped versus videotaped physician-patient interactions. *Journal of Nonverbal Behavior*, ***26***, 219-239.
2. Wakefield, J. (1985). Direct observation. In V. R. Neufeld & G. R. Norman (eds.) *Assessing clinical competence (pp. 163-182)*. New York: Springer.

Mining for meaning in driver's behavior: A tool for situated hypothesis generation and verification

E.R. Boer[1], C.A. Joyce[2], D. Forster[2], J.B. Haué[3], M. Chokshi[2], T. Mogilner[2], E. Garvey[2], J. Hollan[2].
[1] *LUEBEC, San Diego, USA*
[2] *Distributed Cognition and HCI Laboratory, Department of Cognitive Science, University of California, San Diego, USA*
[3] *Centre de Recherche et d'Appui pour la Formation et ses Technologies (CRAFT) Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

Abstract
Driving is a highly complex task. Its complexity, however is not attributed to controlling the vehicle but to the manner in which different contexts alter the set of relevant constraints that impact meaningful behavior. For driver support system (DSS) designers, it is important to understand this situated complexity in order to design systems that provide support when it makes sense to the driver to get it and not forgo support when expected.
To combat or confront driving complexity, researchers, in the interest of time, either highly constrain the conditions of study (human factors) or delve deep into the myriad of factors that influence behavior for a limited number of drivers or situations (ethnography). The primary time sinks are: i) collection of naturalistic driving data that spans a sufficient domain of driving situations to guide support system development, and ii) analysis of these data to gain constructive insight into the contextual factors that mediate drivers' behavioral expression. The CoBeX software presented in this paper addresses the latter in the context of designing and evaluating the functionalities of typical car following support systems.

Keywords
Ethnographic analysis, data visualization, data mining, instrumented vehicle, driving behavior.

1 Outlier exploration
Outliers in driving or DSS behavior are seldom the result of noise in the human-machine system; they are generally the result of a context that differs from the one that was used to define "normal" behavior; normal is defined by a model. Much of model development can be characterized as the mining for relevant input variables/factors and logical input-output transformations that predict the observed behavior best. It involves gaining a contextual understanding about the limitations in the generality of a given behavioral representation (i.e. the process to bring outliers under the model's predictive umbrella). Similarly, DSS development can be characterized as managing outliers to assure DSS operation is consistent with driver expectations.

1.1 Contextual effect on the meaning of signals
Analysis of driving behavior without context yields models that are constrained in their set of input variables (e.g. a radar-cone based car following model). Consequentially, DSS that are designed with contextually sparse data run the risk of not supporting the driver when useful and supporting when not useful; these discrepancies between driver expectations and system behavior lead to reductions in trust, satisfaction, and ultimately usage. The issue is that the signals sensed and used by the DSS have different meaning for the driver in different contexts and thus generate different behaviors.

For example, a decelerating lead vehicle that is about to turn right may not be perceived as a constraint and therefore receive no deceleration behavior from the driver. The same range and range rate condition in a different context where the lead vehicle is decelerating for a slower vehicle ahead, the driver does perceive the closing gap as a constraint and decelerates. This is an example of many similar revelations that arose from interactive CoBeX sessions.

1.2 Mining for meaning: cognitive ethnography
The necessary understanding of driving complexity for the DSS design and evaluation cycle is developed through a particular approach to ethnography motivated by the theory of Distributed Cognition [1]. This ethnographic paradigm yields a situated understanding of how meaningful behavior emerges from interactions of individuals with technology and other structure in their environment, in a complex yet natural context [1]. This lengthy process requires detailed examination of multiple data streams (i.e. video, time series, notes, annotation, verbal reports, audio recordings, etc.) from many perspectives. It molds these views into one coherent object which embodies the proposed meaning of the studied activity.

1.3 Software supported contextual exploration
Much of science aims at understanding the apparent consistency and variability in an observed process through contextual exploration. CoBeX (*Co*ntextual *Be*havioral e*X*ploration) was designed by LUEBEC & UCSD to support explorations into how context affects driver behavior and whether this differs across drivers. Through iterative data augmentation and outlier exploration of multidimensional data captured on a highly instrumented vehicle [2], we located patterns that characterize driver behavior in lane changing and car following. These were used to construct models that embody the interactions between contextual elements that produce the observed behavior (see our Petri net paper in these proceedings).

2 CoBeX analysis requirements
CoBeX's primary goal is to offer a flexible data visualization, analysis, and programming tool-set for heterogeneous data sets from instrumented vehicles.

2.1 Heterogeneous instrumented vehicle data
Ethnographic analysis relies on the ability to take multiple perspectives at the same behavior. This is facilitated with data from our UCSD instrumented Infinity Q45 vehicle [2]. The following data streams are a subset of the ones collected during each of our one-hour drives with normal drivers; each driver drove the same route to facilitate between driver contextual comparisons:

i) 8 environment and driver centered video streams including a foot camera and a 3rd-eye camera mounted on the driver's forehead, ii) GPS, iii) driver control actions such as pedal movement, steering actions, and ACC usage, iv) radar readings of distance and relative velocity to the lead vehicle, v) audio recordings of driver as well as ethnographer passenger(s) voice, vi) vehicle state variables such as speed, longitudinal acceleration, and lateral acceleration, vii) turn signal activation, viii) notes typed by the ethnographer back seat passenger, ix) traffic density estimates from a local loop-counter agency, x) heart rate (few subjects), and xi) synchronized time stamps (unixtime: seconds since Jan. 1, 1970) on all data.

From these data a number of additional time stamped data were generated through post-drive data processing:

i) lateral lane position extracted using computer vision from the forward facing camera, ii) head velocity vector from the 3rd-eye camera, iii) foot movement between gas and brake pedal from foot camera, iv) annotations of all locations based on road, junction, intersection, on-off ramp, etc. from digital maps and GPS, v) transcriptions of driver voice, vi) derived perceptual variables such as time headway (THW) and time to collision (TTC) to lead vehicle from the radar, vii) hand positions from driver view camera (few subjects), and viii) transcriptions from post-drive reflective interview with drivers to elicit more information about a number of driving periods while looking at their video(s).

CoBeX enables analysts to overlay any of these data streams (including speech transcriptions) based on a specified time or location window for any number of available subjects (see Figure 1 for example snapshots of different functionalities). When video data is played, a time marker advances in the corresponding time or location data windows.

CoBeX also allows for any two variables to be set against each other in a scatter plot. Each point in a scatter plot can be selected and used to see the corresponding video and time/location series around that point's spatiotemporal coordinates or to create a mosaic of various data views around that event. This real time cross-view referencing is the key to interactive and iterative outlier exploration (i.e. meaning making).

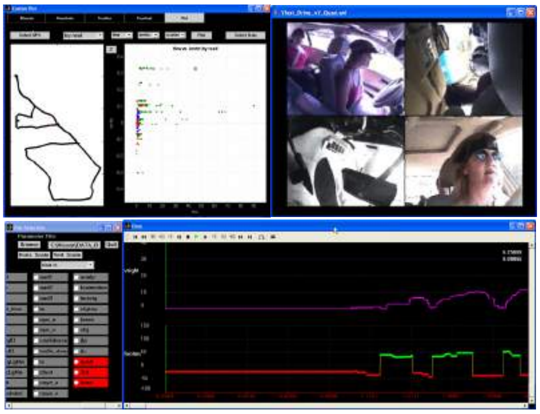


Figure 1. *Screen shots of CoBeX's user-configurable windows.*

2.2 Outliers as a focus for functionality

The visualization capabilities and data assimilation capabilities of CoBeX (i.e. generate custom data sets based on any search criteria that includes temporal, location, and variable based constraints) allow for rapid development of models whose generality can be explored by assessing those observations that fall outside the model's prediction (i.e. model outliers); may be as simple as visually judging what samples fall outside an observable trend in the data. Even though CoBeX at the moment does not include dynamic model development and identification, CoBeX is written in Matlab thus offering Matlab users the power of Matlab to augment CoBeX's functionality for their specific purpose. This is how CoBeX continues to be developed.

CoBeX facilitates exploration of many different types of outliers: i) outlier point(s) in scatter plots, ii) aberrant time series in overlays of multiple drivers or a single driver's responses to a set of events that satisfy some search criterion (e.g. TTC < 10s), iii) extraordinarily large spikes in time series, or iv) a different progression of a time series through a multi-dimensional state space (currently supported for 2 dimensions). By offering CoBeX users the means to gain insight into the factors that make meaning of these outliers, the representations and models can be contextually enriched or the DSS's operational domain contextually augmented.

3 Support system development

The logic underlying most longitudinal DSS is generally based on a radar cone perspective that provides information about range and range rate devoid of context. This means that the support generated from the range and range rate may not be appropriate for the situation at hand (e.g. lead vehicle decelerating to turn such that gap dynamics are non-critical from driver's contextually rich perspective but critical from the radar's contextually poor perspective). Drivers make meaning of the situation by looking at the turn signal and the cross road to infer that the lead vehicle is turning. If a system had similar contextual awareness, it could adapt accordingly (e.g. soften support or alter support activation criteria).

3.1 Application versus operational domain

The ethnographic analysis supported by CoBeX offers DSS designers a means to assess the types of situations where the targeted or practically feasible DSS may activate unnecessarily (i.e. false alarm) or fail activation when necessary (i.e. a miss). An example of the latter is a vehicle cutting in at close range, to which a driver would respond well before the vehicle enters a typical radar cone. The goal is to develop an understanding of the degree to which a targeted DSS's activation domain matches that of the driver and to use this understanding to modify the system to meet driver expectations.

By offering a means to accelerate the ethnographic process though CoBeX, naturalistic contextually rich data sets become a more feasible research option for increasing our understanding of driver behavior and improving the design of driver support systems.

This research was supported by Nissan Motor Co. LTD, Japan.

References

1. Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA. The MIT Press.
2. McCall, J.; Achler, O.; Trivedi, M.M.; Fastrez, P.; Forster, D.; Haue, J.B.; Hollan, J.; Boer, E. (2004). A collaborative approach for human-centered driver assistance systems, *7th IEEE Conference on Intelligent Transportation Systems*.

Behavioral entropy as a measure of human operator misunderstanding

Erwin R. Boer [1], and Michael A. Goodrich [2]
[1] *LUEBEC, 5434 Renaissance Ave., San Diego, CA 92122, USA*
[2] *Computer Science Department, Brigham Young University, Provo, UT, USA.*

Abstract
Human activity can be characterized as a process of controlling and resolving uncertainty in the quest to greater satisfaction. When satisfaction is difficult to achieve, as when human operators encounter more uncertainty, suffer from greater misunderstanding, or have reached the limits of their abilities, their activity patterns are generally more erratic and less predictable; their eye movements, hand movement, brain activity, etc. reflect a state characteristic of struggling, searching, or trying. These patterns of greater behavioral randomness are observed when operators do not know: i) what they are looking at, ii) what the relevant cues are to control a system, iii) what the dynamics are of the system they are interacting with, iv) how to solve a problem, or v) do not have the cognitive or sensory-motor abilities necessary for the demands of the interaction. The common characteristic lies in the relationship between level of "misunderstanding" and behavioral entropy (BE). Here we discuss the general theory of BE, briefly introduces the associated metrics, and points to a range of applications.

Keywords
Operator modeling, operator performance assessment, interface evaluation, learning and adaptation, cognition.

1 Learning how to interact
Norbert Wiener stated: "It is my thesis that the physical functioning of the living individual and the operation of some of the newer communication machines are precisely parallel in their analogous attempts to control entropy through feedback" [11]. Wiener's philosophy suggests a unifying approach to quantify the degree to which humans are successful in reducing entropy as reflected in the entropy of their behavior. Erwin R. Boer's thesis is that *behavioral entropy is a direct reflection of a human's understanding (i.e. the perceived entropy) of the work domain within which that behavior manifests*. In other words, once a human has recognized the cognitive structure of the work domain [5], this structure organizes behavior into a low-entropy interaction process. It is a natural human quest to resolve uncertainty through active interaction with the work domain at multiple integrating levels of perceptual, cognitive, and motor control activity. Humans feel challenged and energized by entropy as manifest in our innate interest and emotional satisfaction we feel from the gradual advancements experienced in learning a sport or partaking in scientific inquiry.

1.1 Child development
Infants interactively construct a cognitive representation of the world through sensory exploration. Their actions (high BE) integrate epistemic probing and motor program development such that future interactions with the world are more efficient and effective (i.e. low BE). The transformation from high to low BE is concomitant to the development of situated cognitive and neural structures.

1.2 Selective ignoring
Children are said to have a photographic memory during their infant years and gradually loose that skill as they mature. One explanation lies in the fact that infants have not recognized the cognitive structures of the environment and have thus not yet learned to ignore vast amounts of "information". Once children learn the spatiotemporal regularities of their surroundings (i.e. establish cognition), the amount of information that gets committed to memory (often only short term memory) decreases as does their ability to recognize every element in a scene at a later time. One important element of learning is to know what to perceptually ignore (reduced eye movement entropy) or respond to (reduced control entropy) [6]. The goal is to develop the minimal cognitive structure needed to achieve satisfactory efficient interaction (control) [5,6,10].

1.3 Epistemic probing
Adults continue their integrated cognitive and sensory-motor development through epistemic probing. Consider the following examples: i) when viewing a scene with an initially unknown structure, eye movements are unstructured until structure revealed or recognized, ii) when controlling a vehicle or other system whose dynamics are unknown, small control actions are used to identify the system's dynamics, iii) when interacting with a new superior, adults change the tone and directness of speech arrive at a mutually agreeable and efficient conversation style, and iv) when encountering a new reasoning problem, adults use cognitive searching and trying, sometimes with the help of pen and paper, until a solution emerges and the mind settles down temporarily. Each of these epistemic activities aims to arrive at a higher level of understanding that facilitates an interaction at a lower entropy level. Interestingly, perceptual and cognitive paradoxes result in a process that can be characterized as an entropic *perpetuum mobile*; a low entropy singular solution does not exist and the eyes and mind keeps switching around the paradox's solution set.

2 Measuring behavioral efficiency
Operator BE manifests in their monitoring (e.g. eye movements from one fixation to the next) and control (e.g. hand and foot movement) behaviors. These behaviors can be characterized as transitions from one behavioral state to another. These states can be temporally defined (e.g. magnitude or rate of steering wheel movement) or spatially defined (e.g. eye fixation on a particular object). A state can thus be as simple as a small magnitude range of a steering angle or as complete as a full lane change maneuver that satisfies a set of criteria. Frequently, these transitions follow characteristic patterns and thus yield to models; these models may account for various spatial or temporal spans, and may rely on both deterministic and

probabilistic elements. We can utilize such models and the discrepancy between model-predicted and observed behavior as a source for state-based behavioral assessment. In all cases, a state is characterized by the probability that it is observed within a time window of interest (time series data) and in some cases also by the probability that a behavioral state is observed after a given other state was observed (i.e. probabilistic data). Thus behavior can be characterized as a sequence of state probabilities (p_i) or as a sequence of state transition probabilities (p_{ij}) [3].

2.1 Quantifying entropy

Entropy H_{B1} of a simple time series behavior (1D characterization) is defined as [9] $H_{B1} = -\sum_i p_i \log p_i$, where p_i is defined as the frequency of behavior i. Similarly, the entropy H_{B2} of a state transition matrix representation (2D characterization) is defined as [9] $H_{B2} = -\sum_i \sum_j p_{ij} \log p_{ij}$, where p_{ij} is the probability that behavior is observed in state i after it was observed in state j. If a *log* base of two is used, the entropy definition is equivalent to the one used in information theory [9].

2.2 Attributing entropy

Behavior should not be assessed independent of the work environment within which it takes place [5,6]. Context needs to be taken into account to define the behavioral state. When human behavior is considered without context, its complexity may be falsely attributed to the human rather then to the behavioral constraints imposed by the environment (e.g. Simon's ant [10]). Once the environment's complexities are accounted for, the complexities of behavior may become more predictable rather than apparently random. This brings us to the power of using a situated model to represent our understanding of the human's behavior and use it to predict behavior. The difference between model predictions and observed behavior is then used as the sequence of observations from which states are defined and entropy calculation generated. These model prediction error states more directly reflect a lack of understanding since they explicitly encode understanding.

Understanding is now more generally defined as the understanding of all players directly involved in the process of generating and computing the BE. The three main players in the design of computer-supported control of environmental entropy are: i) the human operator, ii) the support system designer, and iii) the behavioral analyst. Thus, when a high BE is observed, it can be attributed to: i) the human operator who is not sufficiently skilled, ii) the interface designer who did not take human abilities or naturally occurring work-domain constraints into account, iii) or the analyst who observed behavior without sufficient account of context. BE thus quantifies misunderstanding of any of the human's that partake in the generation and quantification of human operator behavior.

3 Application

In nearly all disciplines where humans are part of an observed system, researchers search for meaningful ways to characterize human performance. The proposed theory of BE quantifies operators' ability to accomplish their task without imposing performance goals on the measure. This is especially important in environments where human operators need to perform multiple tasks in parallel and have flexibility in how they distribute safety and performance tolerance margins across the many processes they control. BE quantifies operators' ability to maintain their own safety and performance tolerance margins.

BE has been applied in the areas such a: i) workload quantification through steering entropy [1,2,7], ii) evaluation of driver support system effectiveness [3], iii) design and evaluation of human-robot interaction schemes [4], iv) evaluation of the effects of intoxication or sleep apnea [8], and v) a general theory for representing driver behavior. All of these applications have in common the fact that a human operator's actions are more erratic and less predictable when the operator skill is low, the interface is inefficient, the support system is counterintuitive, or the human is impaired. The fact that the measuring technique does not impose normative performance metrics attributes to the metrics high sensitivity [2] and tight correlation with subjective workload and satisfaction metrics [7]. The latter is attributed primarily to the fact that the method directly measures the effectiveness of the human's cognitive representation and control skill in maintaining satisfactory margins to self imposed and chosen constraints.

This research is supported by Nissan Motor Co., LTD., Japan.

References

1. Boer, E.R. (2000). Behavioral entropy as an index of workload. *44th Annual Meeting of the Human Factors and Ergonomics Society*, San Diego, CA.
2. Boer, E.R.; Rakauskas, M.E.; Ward, N.J.; Goodrich, M.A. (2005). Steering entropy revisited. *3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Rockport, Maine, USA.
3. Boer, E.R.; Ward, N.J. (2003). Event-based driver performance assessment. *2nd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Park City, Utah.
4. Goodrich, M.A.; Boer, E.R.; Crandall, J.W.; Ricks, R.W.; Quigley, M.L. (2004) Behavioral entropy in human-robot interaction. *Proceedings of PERMIS 2004*. Aug 24-26, Gaithersburg, Maryland.
5. Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA. The MIT Press.
6. Mulder; M.; van Paassen, M.M.; Boer, E.R. (2004). Exploring the roles of information in the manual control of vehicular locomotion: from kinematics and dynamics to cybernetics. *Presence, Vol.* 13, No. 5, pp. 549-559.
7. Nakayama, O.; Futami, T.; Nakamura, T.; Boer, E.R. (1999). Development of a steering entropy method for evaluating driver workload. *SAE Technical Paper Series: #1999-01-0892*: International Congress and Exposition, Detroit, Michigan.
8. Paul, A.; Boyle, L.N.; Boer, E.R.; Tippin, J.; Rizzo M. (2005). Steering entropy changes as a function of microsleeps. *3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Rockport, Maine, USA.
9. Pierce, J.R., (1980). *An introduction to information theory, symbols, signals, and noise*. Second Revised Edition. Dover Publications, Inc., New York.
10. Vicente, K. J. (1999). *Cognitive work analysis*. Mahwah, NJ, Lawrence Erlbaum Associates.
11. Wiener, N. (1989). *The human use of human beings*. Free Association Books, London.

Relationship between rating of perceived exertion and skin temperature changes induced by the static contraction of human muscles

Ming Yu, Kai Cui and Linyan Sun
Human Factors Lab, the School of Management, Xi'an Jiaotong University, Xi'an 710049, the People's Republic of China

Abstract
Rating of perceived exertion (RPE) and skin temperature changes (STCs) of the humerus muscles were measured by infrared thermography (IRT) as the subject took a static task of holding a dumbbell using the right arm. The main relationship between RPE and STCs related to the static contraction was presented and analyzed elementarily.

Keywords
RPE; STCs; IRT; Static contraction; Muscles

1 Introduction
Though heat production of the contracting muscles is well known, research on characteristics of heat production and proliferation is still rare. There are few studies on the relationship between rating of perceived exertion (RPE) and skin temperature changes (STCs). Creating a simple and precise model of RPE and STCs pattern related with muscular contraction to develop studies of complex locomotor thermoregulation is considerably challenging. Infrared thermography (IRT) is a useful technique in studies of human physiology [1]. This study has taken advantage of IRT to find the relationship between RPE and STCs induced by the static contraction of human muscles to unveil general characteristics of RPE and STCs.

2 Methods

2.1 Subjects
Three male and two female undergraduates of Xi'an Jiaotong University volunteered as subjects. Their ages were from 20 to 23 years. The mass was 61.2±8 kg and the height was 172.2±7 cm.

2.2 Static contraction of muscles
The upper arm of each subject was chosen to be experimented. Figure 1 contains two IRT images which respectively depict a subject's postures of the contracting and resting. The circles in the marked images show the position of the experimented muscle.

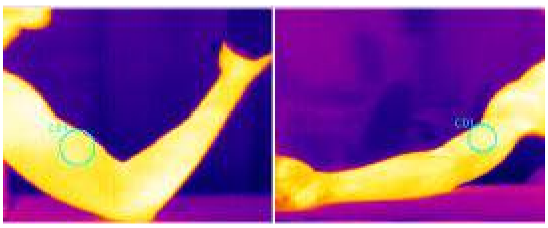


Figure 1. *A subject's IRT images*

2.3 Experimental ambient conditions controlling
Necessary measures had to be taken into account to control the impact of unexpected changes of the ambient conditions. Laboratory ambient temperature was kept in the range of 25.2±0.3℃， and air humidity 50± 5 %. Subjects were asked to make their upper arms naked and keep them as dry as possible.

2.4 Procedure
Five subjects were in a single experimental group. The experiment comprised three cycles of subjects' exercises. Five contracting tasks included in the cycle were done one by one. There was a one-hour interval between the cycles. Every subject had to participate in the three cycles and do the three contracting tasks in all. From the beginning to the end of the process, RPE and STCs data were regularly recorded once per 10 seconds.

2.5 Temperature data and subjective RPE
This study used IRT to obtain the STCs data to analyze thermal patterns of the static contraction of muscles. Initial data came from those infrared photos which had been taken by the infrared camera during contracting and resting processes. The SAT-REPORT software attached with the infrared camera was used in the data processing. Subjective RPE values were obtained at 10 seconds intervals during experimental processes using the 10-point Borg scale [2]. The scale ranges from 0 to 10. 0 means "no exertion at all", and 10 means "maximum exertion".

2.6 Statistical analysis
After thermal data of every subject in the three experimental cycles were obtained, a data averaging operation was processed. Actually, these statistical data were average values of maximum STCs of all subjects. The modified RPE and general STCs curve were verified by means of statistical regression. The model analysis was carried out using MATLAB 6.5. The statistical significance was accepted at the $p < 0.05$ level.

3 Results
A fit function having four unknown parameters was taken to fit the relationship of the mean value group. The fit function is as follows:

$$f(x) = a_0 + a_1 \times \cos(x \times w) + b_1 \times \sin(x \times w)$$

The discrete RPE and STCs values describe the pattern of RPE and STCs of each subject. At the same time, the fit curves were obtained from the discrete points, and the four unknown parameters corresponding to each subject were estimated respectively. The two figures that presented fit curves of RPE and STCs of each subject were given out in pairs to analyze the patterns and compare the characteristics of the two fit curves (Figure 2, 3, 4, 5, 6).

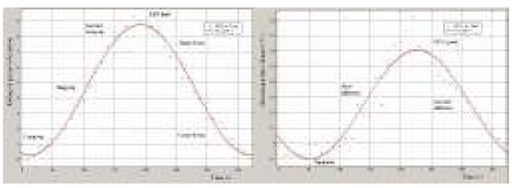


Figure 2. *The fit curves of RPE and STCs of subject 1*

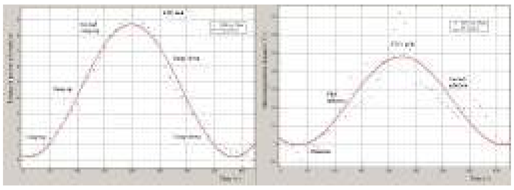


Figure 3. *The fit curves of RPE and STCs of subject 2*

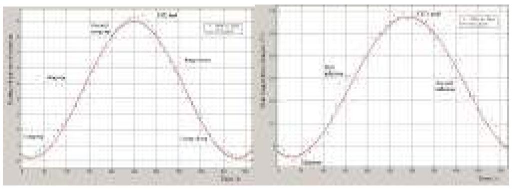


Figure 4. *The fit curves of RPE and STCs of subject 3*

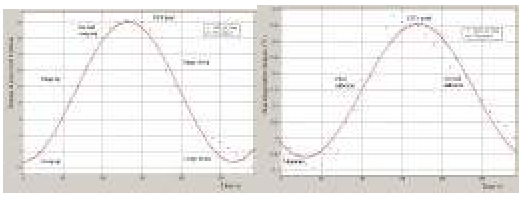


Figure 5. *The fit curves of RPE and STCs of subject 4*

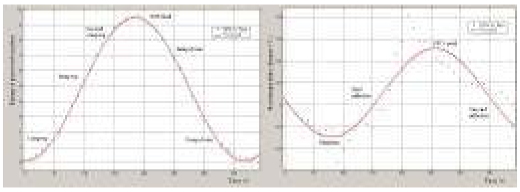


Figure 6. *The fit curves of RPE and STCs of subject 5*

The RPE curve visualizes the changing tendency of RPE obtained in the experimental process by means of feeling. The curve comprises five parts which correspond to the relevant phases in the process. The first part is a creep-up phase. In this phase, there is a mild exertion perceive, i.e. a moderate ascending of the RPE. The second part is a jump-up phase. In this phase, there is a strong exertion perceive, i.e. a sharp ascending of RPE. At this ascending phase, everyone can obviously perceive the physiological and psychological effect. A subject's heart and breathing rates begin to rise obviously. After a speedy ascending, RPE turns to ascend moderately. At the inflexion of the curve, the third part which is called a second creep-up phase begins. It gradually ascends to the RPE limit. The limit value of RPE in this phase is the maximum of the modified Borg scale. Though it can be infinitely close to the limit, the curve cannot overrun it ultimately. The fourth part is a jump-down phase. In this phase, there is a quick recovery of RPE. Generally, everyone can quickly recover from the uncomfortable status of exertion limit after moving away the load. The last part is a creep-down phase. It is assumed that the impact induced by the load can be felt slighter and slighter as the recovery goes on. That is to say that RPE will inevitably come back to the original level after a relatively long recovery.

STCs dropped down and became negative instantly after the commencement of the contracting exercises. Then, the curve makes a turning at the minimum point and keeps on rising till the skin temperature reaches the peak. The inflexion is so crucial that certain phases of RPE and STCs curve correspond to each other accordingly. After the STCs reaches the peak, the curve makes another turning at the peak point and keeps on dropping down till STCs return to the original level, zero. There is also another inflexion in the dropping period. Analogously, the STCs curve is also made up of five parts which are corresponding to the phases of the RPE curve. Four key points in the curve are marked, which differentiate the relevant five periods.

4 Disscussion

4.1 Loss of heat

According to the findings, different fluid balance of subjects can make a comprehensive impact on heat production and proliferation of contracting muscles. This impact will be more significant for a subject who is easier to sweat. Subjects in this experiment were asked to sit down and relax. Most parts of the whole body, except the right upper arm, were relaxed. Furthermore, the static contraction of local muscles in the right upper arm brought few influences on the whole body. This helped the subjects do their best to calm down and decrease the sweat rate. Less sweat brought less heat loss. Thereby, data from the experiment gave more accurate results than otherwise.

4.2 RPE and STCs curves of different subjects

Actually, subjects had different exerting abilities to perform the task. Their fatigue thresholds were not certainly the same. Different subject spent different times on each task. Therefore, it was difficult to describe the comparison of the RPE curves of different subjects. As a result, this situation could induce a deeper discussion. However, it would be easy to compare the different RPE curves if only one considered the entire changing pattern. The wave patterns of RPE curves fitting by the Fourier functions presented a considerable similarity, though the task duration of each subject was different. The same amplitude with the different frequency presented the most conspicuous waving characteristic. STCs curves had the same wave pattern, but the peaks of different curve were not the same value. Therefore, the amplitude of each cyclic curve was different. It was evident that the variable amplitude with the different frequency presented the same wave pattern.

4.3 Limitation

In this paper, there were two crucial problems unresolved. First, the small sample number (five subjects) inevitably inflicted an influence on curve fit results. Second, the subject did only one task in a cycle without doing sequential tasks. Though the subject did three tasks in all, they did that discontinuously. Thereby the only swing in the first cycle of the curve was tested, and the whole swing patterns of RPE and STCs wave including more cycles were not verified perfectly.

Supported by the 211 education project of China.

References

1. Jones, B.F.; Plassmann, P. (2002). Digital infrared thermal imaging of human skin. *IEEE. Eng. Med. Biol.,* **21 (6)**, 41-48.
2. Borg, G. (1982). Psychological bases of physical exertion. *Med. Sci. Sports Exerc.,* **14,** 377-381.

VIP-lab: A virtual lab for ICT experience prototyping

K. Coninx [1], M. Haesen [1], J. Bierhoff [2]
[1] *Expertise Centre for Digital Media, Hasselt University, Diepenbeek, Belgium*
[2] *European Centre for Digital Communication, Heerlen, The Netherlands*

Abstract

In this paper, we present a "Virtual Lab for ICT Experience Prototyping" as a joint organization of a number of research groups in Belgium and The Netherlands. The main goal of the project is to provide a multidisciplinary, profiled, accessible service, organizing all of the prototyping steps for ICT in enterprises.

To develop user-friendly prototypes, our approach will be based on ICT Experience Prototyping. We will describe a case study to illustrate the methodology, with special attention to observations of target users. The end product of this case study will be an evaluated prototype of a flexible user interface for the mobile journalist.

Keywords

User-centered design, user experience, multidisciplinary, prototyping.

1 Introduction

Nowadays developers can make an appeal to a wide range of traditional software engineering methods to develop applications. Usually these methods focus on functional requirements. Software engineering methods guarantee to some extent that the system adheres to quality attributes such as robustness and correctness, but it is not always sure that the resulting system is usable. After all it is the end-user who has to use and appreciate the system or service. Because of that reason, User-Centered Design techniques are emerging.

The term "User-Centered Design" is introduced by Norman [4]. He defines this as "*a philosophy based on the needs and interests of the user, with an emphasis on making products usable and understandable*". Based on this definition, many User-Centered Design methods are drawn up, such as GUIDE [5], ISO 13407 [2] and Contextual Design [1]. Their approaches vary, but in general they all recommend to involve the target user in the whole design cycle of an interactive system. In practice, this can be done a.o. by observing the end-users, by giving them an active role in the design team and by letting them evaluate the prototype.

The region we are targeting is made up of five provinces in the North-East Belgium and the South-East of the Netherlands. Enterprises in general rarely pay attention to the usability of innovative products. This can be explained by the fact that only a few companies and institutes provide User-Centered services to enterprises.

With this in mind, a project is set up to provide a profiled, accessible service, organizing all of the prototyping steps for enterprises. The project will be presented in this paper. First, the aims and the organization of the project are introduced. Afterwards we will describe a case which illustrates our general approach based on User-Centered Design. Finally, section 4 discusses the conclusions.

2 VIP-lab

The project presented in this paper is named VIP-lab. On the one hand this name refers to the privileged treatment of the end-users. On the other hand, this is an acronym for 'Virtual ICT experience Prototyping lab', which will be explained below.

2.1 ICT Experience Prototyping

Our approach of ICT Experience Prototyping will broaden early approaches for User-Centered Design methods. It is not only the intention to guarantee usability, but also to target the whole user experience, including accessibility, utility, functionality, likeability, sociability, playability. Based on this experience, a prototype will be developed in an incremental and iterative way.

2.2 Goals

In the first place, the aim is to make companies in the region aware of the advantages of ICT Experience Prototyping. Secondly, by centralizing their know-how, the partners of the consortium want to improve their methodologies for User-Centered Design.

In order to reach these goals the Virtual lab will during the next three years disseminate the ideas of experience prototyping to local companies and will realize a number of pilot projects, which will investigate how prototyping can improve current practices in industry.

2.3 A multidisciplinary consortium

As the project concerns a virtual lab, the emphasis is not on one physical lab facility, but the involved research groups will integrate their know-how with respect to designing and measuring the user's experience. This will be achieved by a multidisciplinary approach, with expertise varying from computer science to social and communication sciences. In particular, the consortium is composed by four research groups and one company, spread over the region.

3 A flexible user interface for the backpack journalist

Within VIP-lab, five pilot projects within several application domains will be carried out. During these pilots we wish to develop a prototype by following ICT Experience Prototyping methodologies. This can be done by involving the end user from the beginning of the design cycle.

Before the start of each pilot, local companies, involved in the corresponding application domain, can attend seminars. They will be informed about ICT Experience Prototyping, and can participate in a pilot.

3.1 Concept

The first pilot will take place in the field of journalism. This pilot is split up into two parts. The first part is about a flexible user interface for reporters, the second concerns communities and reader-driven web sites. This paper will only focus on the first part.

The concept involves a tested prototype of an application which will assist journalists in reporting. Nowadays, reporters take along a backpack, which contains several advanced newsgathering tools [3]. The problem with this equipment is that it causes extra pressure for the journalist. Next to bringing the news, he or she also has to know how to make these hi-tech tools work together. From that point of view, the idea for the prototype is grown. The aim is to find a suitable hard- and software combination and to develop an interface which guides the reporters in their work and integrates the different applications. This implies that the technical aspects of the system have to be as transparent as possible and that journalists do not have to bother about how the network connection is set up, how applications are started, how synchronization of multimedia is achieved etc. Which approach will be followed to achieve these goals, is exemplified below.

3.2 Procedure

Right after the start of the pilot project, the concept and the procedure of ICT experience prototyping in this pilot are discussed in meetings with the research groups and participating companies.

The first task is to get to know the target users, their tasks, their needs and their environment. This will be realized through a user and task analysis. The basic idea of the analysis is to get an understanding of how the tasks have to be done. In practice, this can be achieved by observing how journalists perform their current tasks. Sometimes it is hard to find the right end-users, but in this case, we can make an appeal to the interested media enterprises and their journalists.

This analysis leads to a conceptual design. In this stage the first mockups of the prototype are drawn up by the research groups. The mockups will also be evaluated by companies and target users.

After the setup of the test plan, the prototype can be developed. Even though the technical research groups are responsible for these activities, the feedback to human aspects of the user interface will be guaranteed by the incremental and iterative approach. Once a reasonable prototype is developed, it will be evaluated through usability tests. The test results will be taken into account for the further development of the prototype. How many times the alternation of the development of the prototype and usability tests will take place, will be decided in the test plan. To find out if the prototype fulfils the expectations, field tests will measure the whole user experience.

3.3 Observing target users

The global procedure as described above, contains three different stages where observing the behavior of end-users can be very important. In this section, the methods for these observations are discussed.

First, target users are observed during the user and task analysis. This can give designers an understanding of how tasks have to be carried out, and how the tasks currently are accomplished. In practice, this will be achieved by application of Contextual Inquiry [1]. During this inquiry, users will be observed and interviewed while they carry out their everyday job. They are also asked to explain the actions they are doing. The observation of a user ends with a survey to ask some additional information about his or her job. In the first pilot project several journalists with different specialties, work environments and technological knowledge will be followed while they report about the news.

The second, and most common observations take place during tests in a usability lab, and will give feedback about the prototype in the short term. These tests will be carried out in a traditionally equipped usability lab, with several observation tools such as an eye tracker, The Observer and Screen Capture Module (Noldus Information Technology, Wageningen, The Netherlands).

Field tests are comparable to traditional usability tests, but they are carried out in the end-user's work environment. Just like in the user and task analysis, journalists will be followed, but this time they perform their job with the help of the prototype. Extensive field tests are meant to test the prototype in more realistic situations. On the other hand, the realistic environment can obstruct the observations: it is harder to measure the user experience, to capture screens, to record the actions of the subject on video etc. To bring these field tests to a good end, the software of the prototype itself will log several measurements, but also a portable usability lab, with the minimal equipment for usability observations or a Pocket Observer are available for these observations.

4 Conclusion

One of the major challenges for VIP-lab is to find suitable methodologies to apply ICT experience prototyping in enterprises. This can be achieved by joining the know-how of four research groups with various scientific backgrounds. The technical knowledge is needed to determine the feasibility of the application and to develop the prototypes, while the human knowledge is advantageous to guarantee a positive user experience.

Even though the results of the first pilot project are not known yet, the general procedure is defined to some extent. It is certain that measurement of human behavior is an important part of the methodology. Because of that reason, observation of end-users will take place during several stages in the development cycle and will contribute to a user-friendly prototype.

5 Acknowledgements

VIP-lab is financed by the "Interreg Benelux-Middengebied" authoroties and is co-financed by Province of Limburg (B), Province of Limburg (NL), Ministry of Economic Affairs (NL) and Ministry of Flemish Government/Economic Affairs (B).

We acknowledge the contribution of our consortium partners: De Vlijt (B), Mediacentrum (B) and TU/e (NL).

Part of the research at EDM is funded by EFRO (European Fund for Regional Development), the Flemish Government and the Flemish Interdisciplinary institute for Broadband technology (IBBT).

References

1. Beyer, H.; Holtzblatt, K. (1998). *Contextual Design: Defining Customer Centered Systems.* Morgan Kaufmann.
2. International Standard (1999). ISO 13407:1999(E). – Human-centered design processes for interactive systems.
3. Newsgear, Ifra Newsplex, (2003). http://www.newsplex.org/knowledgebase/ntr4.shtml.
4. Norman, D. (1988). *The Design of Everyday Things*. Basic Books.
5. Redmond-Pyle, D.; Moore, A. (1995). *Graphical User Interface Design and Evaluation*. Prentice Hall.

Evaluation of an IT-Artifact – Case study at an art gallery

H. Gottlieb, H. Simonsson
Visions for Museums/Interactive Institute, Stockholm, Sweden

Abstract

This paper discusses the measuring of visitors' behavior after they had tried a multimodal installation called *Access In Mind*, an "introduction station" to contemporary art exhibited 2004 at Liljevalch's Art Gallery, Stockholm. The target group was visitors between the ages of 13 and 18 years. Sitting in *AIM* the visitor hears three voices sharing their views on art from different vantage points. *AIM* emphasizes the importance of being open-minded. It also provides the inexperienced art gallery visitor with guidelines for how to talk about contemporary art. The visitor's behavior was measured using a combination of video recordings, questionnaires and observations — a combination of methods which proved to be fruitful. Our key findings demonstrated that visitors who had used AIM showed greater interest for contemporary art, spent more time in the gallery space and used a greater variety of words to describe their aesthetic experience.

Keywords

Art education, Visitor studies, IT-artifact, Mixed reality, Multimodal interfaces.

1 Introduction

Access In Mind is a prototype developed 2002 at Visions for Museums, Interactive Institute in Stockholm. It is an audio-visual installation aimed for art galleries and it uses a multimodal approach to encourage young people (between the ages of 13 and 18 years) to explore contemporary art. Sitting in *AIM* the visitor hears three voices engaged in a kind of ongoing discussion. The voices represent fictitious characters who each share their views on art from different vantage points. *AIM* emphasizes the importance of being open-minded and wants the inexperienced art gallery visitor to feel free in his/her interpretation of art but at the same time it aims at providing the visitor with guidelines for how to talk about contemporary art.

The design consists of a spherically-shaped chair equipped with speakers and 3D sound. The sound system also produces vibrations in the chair through a basic unit in the seat. Light is emitted through a number of Plexiglas plates combined with diodes that randomly emit different colors. The chair is covered with foamed polythene plastic, and is placed on a wooden podium.

1.1 Method

The study was designed to evaluate the installation *AIM* in a real world context, in a situation where young visitors experience an art exhibition. In a socio-cultural perspective, learning can occur as a consequence of the social interplay between individuals. Therefore we decided to investigate the interaction between the visitors in pairs, which could be described as a social context.

The study was designed as a case study, comparing the respondents who had experienced *AIM* with those who had not experienced it before visiting an exhibition of contemporary art. We wanted to study both the young visitors' behavior in the art gallery and their written reflections after their gallery tour.

1.2 Data format

The collected data consisted of 1) two hours of video recorded material 2) observation notes 3) questionnaires (with open questions). In the video material and the observation notes it was possibly to detect patterns in the visitors' movements and actions in the room. We could make conclusions about the following areas: time spent in the exhibition room, visual interaction between the young visitors in each pair, gesticulation, and interest/disinterest of the pieces of art (studied as amount of stops). The respondents who participated in the video recordings were informed about the ongoing study and participated out of free will. When the observations were made the researcher was anonymously standing among other visitors in the exhibition room (only the teachers were informed about the ongoing study).

The visitors who participated in the video recordings were asked to fill out questionnaires. In these the respondents were asked to comment on his/her experience of art galleries/ museums, previous knowledge of history of art, habit of engaging in artistic activities, educational background, special memories or impressions of the art objects in the exhibition, themes for their conversation with their friend while they looked at the art, their personal definition of contemporary art and their experience of *AIM*.

Our findings from the video recordings, the observation notes and from the answers of the questionnaires were transferred to a spreadsheet. This allowed us to discover and analyze patterns and connections easily.

2 Video recordings

2.1 The Time Factor

The most easily measured feature in the video recordings was time. The time spent in the room of study increased among those young visitors that had been introduced to contemporary art by *AIM* compared to those that had been introduced by a guide, or the visitors who hadn't been introduced to contemporary art at all.

When studying the recordings and comparing them with the questionnaires we noticed how the average time the visitors stayed in the room increased with their age and with their background knowledge of art. There was a correspondence between their former experience of art and their interest of the exhibition. This tells us something about the importance of the visitors' identity, previous experiences and knowledge; the visitor looks at the exhibition through the filter that consists of his or her experience.

2.2 Visual Interaction between the Visitors

In the video recordings, we could observe two main ways that visitors would interact in pairs. If one of them behaved like the leader of the two, it was usually that person who initiated a conversation about the art, decided where and when to stop and when to leave and so on. The other pupil then imitated the movements of the "leader" and followed that person and behavior but never took an initiative of their own. This way of interaction gave the

impression of being mainly a one-way communication, with aspects of power to consider.
When the pupils were close friends, the learning process seemed to be taking place between two peers - if one of them made a comment and pointed at a piece of art the other one answered back and immediately pointed at another work. They took turns in taking initiatives. Pointing and a commenting about a piece of art were always followed by a similar behavior from the friend. This way of interaction worked like a dialogue.

2.3 Gesticulation

When studying the video recordings we could see how some of the visiting pupils went rather quickly from one piece of art to another while others stayed quite a long time in front of each piece of art. It was often impossible to tell what expressions the visitors had on their faces because they had their backs towards the camera. But the way the visitors moved their bodies while watching the art sometimes seemed to express indifference, disappointment, happiness or other feelings. What we could observe more easily than facial expressions on the video recordings was how much gesticulation they were making. But we noticed that some of the visitors were making a lot of gestures in front of every piece of art while some never made any gestures. Therefore the amount of gestures might possibly be more a matter of personality than any proof of interest or disinterest.

2.4 Interest/Disinterest

In the video recordings it is rather difficult to tell what facial expressions the visitors had while looking at the art since they often had their backs toward the camera. But noticeable is how some of the visitors look at the art really quickly and don't seem to be interested in it, while others are scrutinizing every detail carefully. Therefore we chose to analyze the visitors' interest using the number of pieces of art that caught their attention as an indicator of interest/disinterest. There was a small difference between category A (visitors who had been introduced to contemporary art by a guide) and category B (introduced by *AIM*). While the category A spent time in front of an average of 5 pieces of art, category B made stops in front of 6 out of 7 possible. The big difference was visible when category A and B were compared with the reference groups, category C that will be described in the following chapter.

3 Observations

In this study, the presence of the camera (even though it was discreetly placed) can not be neglected and probably affected the behavior of the visitors to some extent. When studying the recordings it was noticeable that most of the visitors neglected the piece of art that was next to the camera. The visitors didn't notice an observer as much as a cameraman and therefore the observer probably affected the visitors' behavior less.
We noticed a different pattern when observing and comparing the categories A and B with some other young visitors at Liljevalch's art gallery, here called category C (pupils between the ages of 16-18 years with a great knowledge of art according to their teachers). The visitors belonging to category C only got a short instruction from their teachers, who asked them to carefully study the pieces of art in the room. The pupils were not given any further introduction or information about contemporary art. It was noticeable how these young visitors only glanced very quickly at an average of only 2 pieces of art per student! They lost their interest rather quickly and left the room.

4 Questionnaires

4.1 Mastery and Appropriation

In our analysis of the questionnaires we used the concepts *mastery and appropriation* to reach an understanding of how *AIM* inspired the young visitors to communicate about contemporary art [1]. In a socio-cultural perspective, it is seen as reported speech when the uninitiated visitor learn from the more initiated (could be the *AIM*-installation, a friend, a teacher or a guide), and uses expressions and aesthetic judgments in a similar way. In the questionnaires, there were findings pointing at an influence from *AIM*.

4.2 Aesthetic judgement

In the questionnaires, there was a frequent use of expressions related to *aesthetic experience* [2]; both positive and negative comments about the contemporary art the respondents had seen in the exhibition. The positive responses were about as common as the negative ones.
In general, the young visitors' comments about *AIM* were positive. They described the installation in words as "exciting and very interesting" or "mysterious or enchanting". We argue that *AIM* worked as a cognitive tool for the visitor. It helped the visitor to learn aesthetic distinctions, to notice relevant aspects of the works of art and therefore gave the visitor ways to communicate around contemporary art with others. Moreover, these aesthetic judgments told us about the visitors' aesthetic experience, positive or negative, in the current exhibition.

5 Conclusion

Our conclusion from comparing the video recordings, the observation notes and the questionnaires is that *AIM,* together with the visitors' age and previous knowledge of art, seem to affect the visitors' interest in the exhibition in a positive way. The result of our study showed that young visitors spent more time in the exhibition room watching contemporary art if they first had experienced *AIM*. It also worked as a cognitive tool for the visitor and gave the visitor ways to communicate around contemporary art with others.
The combination of video recordings, observations and questionnaires was a fruitful combination when measuring the young visitors' behavior. But analyzing data is very time consuming and the possibility to use a computer program which could process and analyze data from the three different measuring methods, video recordings, observations and questionnaires, would be a very useful tool that would simplify the future studying of visitors' behavior in an art gallery.

We thank Eva Insulander who participated in the study at Liljevalch's and analyzed the questionnaires; the cameraman Jon Mårtensson; and the staff at Liljevalch's Art Gallery.

References:

1. Wertsch, J. (1998). *Mind as action*, New York, Oxford, Oxford University Press.
2. Dewey, J. (1987). *Art as experience. The later works 1925-53*, Volume 10. Carbondale, I11, Southern Illinois University Press.

The FaceReader: Online facial expression recognition

M.J. den Uyl, H. van Kuilenburg
VicarVision, Sentient Systems B.V., Amsterdam

Abstract

This paper describes our FaceReader system, which we presented at Measuring Behavior 2005. This system is able to describe facial expressions and other facial features online with a remarkable accuracy. In this paper we will describe the possibilities of such a system and will discuss the technology used to make the system work. Currently, emotional expressions can be recognized with an accuracy of 89% and it can also classify a number of other facial features.

Keywords

Facial expressions, classification, model-based, Active Appearance Model, real-time.

1 Introduction

1.1 What a face can tell

Apart from being the means to identify other members of the species, the human face provides a number of signals essential for interpersonal communication in our social life. The face houses the speech production apparatus and is used to regulate the conversation by gazing or nodding, and to interpret what has been said by lip reading. It is our direct and naturally preeminent means of communicating and understanding somebody's affective state and intentions on the basis of the shown facial expression [4]. Personality, attractiveness, age and gender can also be seen from someone's face. Thus the face is a multi-signal sender/receiver capable of tremendous flexibility and specificity. In turn, automating the analysis of facial signals would be highly beneficial for fields as diverse as security, behavioral science, medicine, communication, education, and human-machine interaction.

1.2 How can FaceReading help?

In security contexts, apart from their relevance for person spotting and identification, facial signals play a crucial role in establishing or detracting from credibility. In medicine, facial signals are the direct means to identify when specific mental processes are occurring. In education, pupils' facial expressions inform the teacher of the need to adjust the instructional message. As far as natural interfaces between humans and machines (computers, robots, cars, etc.) are concerned, facial signals provide a way to communicate basic information about needs and demands to the machine. Where the user is looking (i.e., gaze tracking) can be effectively used to free computer users from the classic keyboard and mouse. Also, certain facial signals (e.g., a wink) can be associated with certain commands (e.g., a mouse click) offering an alternative to traditional keyboard and mouse commands. The human ability to read emotions from someone's facial expressions is the basis of facial affect processing that can lead to expanding interfaces with emotional communication and, in turn, to obtaining a more flexible, adaptable, and natural interaction between humans and machines.

Figure 1. *FaceReader demonstrator as shown at Measuring Behavior 2005.*

2 Face reading technology

The core problem of face analysis is how to simultaneously account for the three major source of variance in face images: pose/orientation, expression and lighting. To counter the problems caused by these sources of variation, the FaceReader classifies a face in three consecutive steps:

2.1 Face finding

Firstly, an accurate position of a face is found using a method called the Active Template Method (similar to the implementation described in [7]). The Active Template Method displaces a deformable face template over an image, returning the most likely face position or multiple positions if we allow more than one face to be analyzed.

2.2 Face modeling

Next, we use a model-based method called the Active Appearance Model (AAM) [2] to synthesize an artificial face model, which describes both the locations of key points, as well as the texture of the face in a very low dimensionality. The AAM uses a set of annotated images to calculate the main sources of variation found in face images and uses PCA compression to reduce the model dimensionality. New face models can then be described as deviations from the mean face, using a compact vector called the "appearance vector". As an example, Figure 2 shows the effects of varying the principle component of the appearance vector between -3 and +3 standard deviation. The AAM manages to compactly model individual facial variations in addition to variations related to pose/orientation, lighting and facial expression.

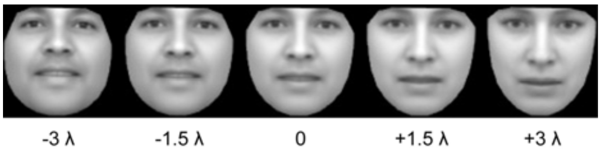


Figure 2. *Varying the first element of the appearance vector.*

Below an example is given of a synthesized face or AAM "fit" that can be automatically obtained from a face image. The AAM fit closely resembles the original face and only very little information is lost despite a very large reduction in dimensionality.

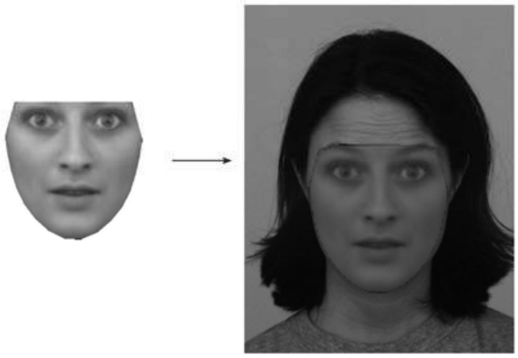


Figure 3. *Example of a generated face model.*

2.3 Face classification

The final stage in the FaceReader architecture is the actual classification of the expression or facial features we are interested in. This can be done in a very straightforward way by training an artificial neural network [1], which takes the AAM appearance vector as input. Given enough training data, we can train the network to learn to classify any facial feature as long as this feature is well modeled in the synthesized faces.

We have trained a network to classify the emotional expression shown on a face in one of the categories: happy, angry, sad, surprised, scared, disgust or neutral. These emotional categories are also known as the "basic emotions" or "universal emotions" [3]. As training material we have used the 'Karolinska Directed Emotional Faces' set [6] containing 980 high quality facial images.
Table 1 shows the performance results for this classifier. Horizontally, the actual expression shown on the presented images is shown and vertically you will see the emotion as predicted by the classifier. The total accuracy of the classification on the chosen set of emotional expressions is around 89% correct, which is among the highest performance rates reported.
Naturally, this method is not limited to the mentioned set of emotional expressions. To illustrate this, we have trained a classifier to detect 15 minimal facial actions, called "Action Units" described in the Facial Action Coding System [4]. This classifier had an average performance of 85% on the selected set of Action Units.
Besides expressions, we have also successfully trained classifiers on properties such as gender, ethnicity, age, facial hair (beard/moustache) or whether a person is wearing glasses or not. We are confident that any feature which a human observer can detect by observing a synthesized face can be learned by a classification network as well.

	happy	angry	Sad	surprised	scared	disgust	neutral	recall
happy	**138**	0	1	0	0	0	1	0.99
angry	1	**116**	2	1	3	11	0	0.87
sad	3	4	**109**	19	2	1	1	0.78
surprised	0	1	6	**128**	0	0	0	0.95
scared	0	8	5	2	**115**	5	3	0.83
disgust	1	5	3	0	3	**125**	0	0.91
neutral	0	11	2	1	1	0	**125**	0.89
precision	0.97	0.80	0.85	0.85	0.93	0.88	0.96	**0.89**

Table 1. *Confusion table for emotional expression classification.*

3 Conclusions and future work

We have managed to create a fully automatic facial classification system which is robust under varying conditions of pose, orientation and lighting using an implementation of the Active Appearance Model as core technology. Our FaceReader system can classify emotional expression with a very high accuracy and can be trained to classify almost any other facial feature.
Currently, we are working on further improving the accuracy of our system and extending the possibilities of classification, so that it will be possible in the near future to classify features which are located outside the modeled area of the face (for example the hair) or features which are poorly modeled by the AAM such as wrinkles, tattoos, piercings and birthmarks. We are also currently in the process of adding person identification to the system.

References:

1. Bishop, C.M. (1995). *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford.
2. Cootes, T.; Taylor, C. (2000). *Statistical models of appearance for computer vision.* Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering.
3. Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest*, **8**, 151–158.
4. Ekman, P.; Friesen, W.V.; Hager, J.C. (2002). *The Facial Action Coding System*. Weidenfeld & Nicolson, London.
5. Keltner, D.; Ekman, P. (2000). *Facial expression of emotion*. In: M. Lewis and J. Haviland-Jones (Eds.) Handbook of emotions, pp. 236-249, New York: Guilford Publications, Inc.
6. Lundqvist, D.; Flykt, A.; Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF, Department of Clinical Neuroscience, Psychology section, Karolinska Institute.
7. Sung, K.K.; Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(1), 39–51.

Behavioral and physiological measures of dichromatic color vision in the horse *(Equus caballus)*

C.A Hall[1], H.J. Cassaday[2], C.J. Vincent[3] and A.M. Derrington[3]

[1] *School of Animal, Rural and Environmental Sciences, Nottingham Trent University, Southwell, United Kingdom*

[2] *School of Psychology, University of Nottingham, Nottingham, United Kingdom*

[3] *Psychology Brain and Behaviour, University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom*

Abstract

In this paper we describe the correlation between the results of color discrimination training and cone excitation ratios in the horse. Six horses were trained to discriminate 15 colors from shades of gray to a learning criterion of 10 consecutive correct choices. Lightness cues were made irrelevant by randomly locating the chromatic stimuli within a chequerboard of shades of gray. The reflectance spectrum of each color was measured and the cone excitation ratios were calculated, based on spectral peak values of 429 and 545 nanometers (nm). Significant correlations between the behavioral data (accuracy, errors and latency of approach) and the cone excitation ratios were found. In addition to providing further evidence for dichromacy, the results of this study demonstrate that physiological modeling can be used to predict the ease with which the horse can discriminate colors. This approach will facilitate further investigation into color perception in the horse.

Keywords

Color vision, dichromacy, discrimination training, horse.

1 Introduction

There is evidence that the horse possesses two types of cone photopigment and is thus classed as dichromatic. Values for the spectral peaks of these photopigments have been estimated as 429nm and 545nm [6]. Behavioral evidence of the ability of the horse to discriminate between chromatic and achromatic stimuli has so far been limited to 4 colors (blue, green, red and yellow; as perceived by the human trichromat). All of the previous studies have concluded that the horse has the ability to discriminate blue from gray [4,7,8,9], but results for green, red and yellow have been less consistent.

By using the values for the spectral peaks of the two cone photopigments of the horse, spectral sensitivity curves can be derived [3]. The predicted effect of specific colors on the equine visual system can be calculated and compared with data obtained from behavioral methods.

The aim of this study was to train horses to discriminate a range of 15 colors from grays. Lightness cues were made irrelevant by mounting the colors within a chequerboard of mixed grays [10]. Cone excitation ratios were then calculated for each of the colors tested and compared with the behavioral data. Correlation between the two sets of results would demonstrate that cone excitation ratios could be used to predict the appearance of colors to the horse, and may also serve to resolve the discrepancies in the results of previous behavioral studies.

2 Methods

The study was carried out at the Brackenhurst Equestrian Centre, Nottingham Trent University.

2.1 Subjects

Six riding horses were used in the study (4 geldings and 2 mares). Their ages ranged from 5-18 years, with heights of between 1.52 and 1.65 meters (m). All of the horses had previously been trained to perform a two-choice, black/white discrimination task to a criterion of 70% accuracy on four consecutive training sessions [5].

2.2 Training Area and Apparatus

The training area was located in one half of an enclosed barn with a concrete floor. It was screened from view by sheeting to a height of 3m. A gap of 0.1m in this screening allowed the experimenter to view the subject. The training area was 5m wide and 10m long. The stimuli were placed on the floor, 6.5m from a “starting” line of masking tape. The stimuli were displayed on two identical wooden boxes, 1.5m apart. These boxes were designed to be lockable without this being visually apparent. The box displaying the positive stimulus (and left unlocked) allowed access to the carrot reward. The box displaying the negative stimulus also contained a carrot, but could not be opened by the horse.

2.3 The Stimulus Cards

The stimulus cards measured 0.375m square and were subdivided into a chequerboard pattern of eight different grays, within which was mounted a colored panel in the case of the positive stimulus, or not in the case of the negative stimulus. Eight different configurations of both the positive (chromatic) and negative (achromatic) stimuli were used.

The colors and grays were produced using the computer graphics programme Paint Shop Pro (version 5.01). The stimulus colors were printed using a Hewlett Packard laser jet 4500 color printer with a resolution of 600x600 dots per inch (dpi). The grays were produced using the grayscale option of the same computer graphics programme and a black and white laser printer (Hewlett Packard 4000 N) with the same resolution as the color printer. Fifteen different colors were produced and they were organized into a sequence according to their spectral appearance to the trichromatic human.

2.4 Colour Discrimination Training

The horses were divided into 2 groups, one starting training with the short wavelength colors, the other with the long wavelength colors. Training sessions were carried out 2-3 times/week and consisted of 16 presentations of the paired stimuli (each presentation = 1 trial). Equal numbers of left and right presentations of the positive stimulus were included in each session. Training with each color continued until 10 consecutive correct choices had been made ($p<0.001$ of reaching this criterion by chance). The next color was then introduced. Performance for each color was assessed using 4 separate measures: number of trials to criterion, accuracy (% of first time correct choices), errors (% of repeated incorrect choices) and latency to approach the stimulus (seconds).

2.5 Calculation of Cone Excitation Ratios

The spectral reflectance of each of the colored stimuli was measured under the experimental conditions using a Minolta CS1000 spectro-radiometer. Spectral radiance values for wavelengths between 380 and 720nm were recorded in 1nm steps.

The spectral sensitivity of an A_1-based photopigment was modeled using the process detailed in Govardovskii *et al.* [3] for the peak sensitivities of the S and M-L photoreceptors of the horse: 429 and 545nm [6]. The resultant cone fundamental spectral sensitivity curves were produced in Matlab, using the Psychtoolbox GovardovskiiNomogram function [1].

To assess the effect that each of the colored stimuli had on the two photopigments, the spectral energy values of the former were multiplied with the sensitivity values of the latter. In Matlab, the Psychtoolbox EnergyToCones function [1] was applied to the energy value for each wavelength in discrete steps, resulting in cone excitation values for every wavelength. These values were integrated to obtain the cone excitation value for each stimulus.

The ratio of S: M-L cone excitation was then calculated using Excel. Cone excitation ratios for the achromatic stimuli resulted in a consistent ratio of S: 0.335 M-L. The difference from this ratio was calculated for each color and was used as a measure of perceived "colorfulness". Figure 1 shows the relative cone excitation values of each of the 15 colors and thegrays. Values for the M-L wavelength colors were considered separately from the S wavelength colors. Correlation between these ratio differences and the behavioral data was then investigated.

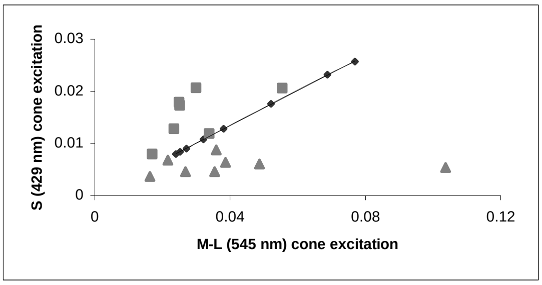


Figure 1. *Cone excitation values for each of the S (▲) and M-L (■) wavelength colors. Trend line for achromatic stimuli is shown.*

2.6 Data Analysis

The data was analyzed using SPSS 9.0 for Windows. A Kolmogrov-Smirnov (Lilliefors) test was carried out to test the normality of the data. Significant variation was found for the trials to criterion and latency data. Correlation between the cone excitation ratios and the accuracy and error scores was tested using the Pearson correlation coefficient. Correlation with the trials to criterion and latency scores was tested using the Spearman's Rank correlation coefficient.

3 Results

All of the horses reached the learning criterion for all 15 colors. For the M-L wavelength colors, a significant negative correlation was found between the predicted "colorfulness" of the stimuli and the number of trials to criterion (p=0.007). Accuracy was greater with the more "colorful" stimuli (p=0.005), fewer errors were made (p=0.004) and the selection made more quickly (p=0.019). No significant correlation was found between the number of trials to criterion for the S wavelength colors and their predicted "colorfulness", but again, accuracy was better (p=0.003), fewer errors were made (p=0,009) and the stimuli were selected more quickly (p=0.041) when the cone excitation ratios differed the most from those of the achromatic stimuli.

4 Conclusion

The results of the color discrimination training clearly demonstrate the ability of the horse to select colors from across the visible spectrum from a range of different grays. However, certain colors took longer to train than others. The ease with which colors were discriminated from grays correlated with the extent to which their effect on the visual system differed from that of achromatic stimuli. The color that proved particularly hard (blue-green) was found to have a cone excitation ratio very close to that of the grays. It is likely that this color was close to the neutral point of the horse [2].

Thus physiological modeling was found to be an effective predictor of the appearance of colors to the horse. This method could be used to make accurate predictions as to possible color confusions (e.g. red/green) likely to be made by the horse. This is a novel approach to understanding the visual consequences of dichromacy in this important domestic species.

We are grateful to the staff at the Brackenhurst Equestrian Centre (Nottingham Trent University) for providing the facilities and horses, and to Steve Westland of the Color and Imaging Institute, University of Derby, for the use of color calibration equipment.

References:

1. Brainard, D.H. (1997) The psychophysics toolbox. *Spatial Vision,* **10,** 433-436.
2. Geisbauer, G., Griebel, U., Schmid, A. and Timney, B. (2004) Brightness discrimination and neutral point testing in the horse. *Canadian Journal of Zoology,* **82,** 660-670.
3. Govardovskii, V.I., Fyhrquist, N., Reuter, T., Kuzmin, D.G. and Donner, K. (2000) In search of the visual pigment template. *Visual Neuroscience,* **17,** 509-528.
4. Grzimek, B. (1952) Versuche uber das Farbsehen von Pflanzenessern. *Zeitschrift furTierpsychologie,* **9,** 23-39.
5. Hall, C.A., Cassaday, H.J. and Derrington, A.M. (2003) The effect of stimulus height on visual discrimination in horses. *Journal of Animal Science,* **81,** 1715-1720.
6. Macuda, T. (2000) Equine colour vision. *PhD thesis. University of Western Ontario.*
7. Macuda, T. and Timney, B. (1999) Luminance and chromatic discrimination in the horse *(Equus caballus). Behavioural Processes,* **44,** 301-307.
8. Pick, D.F., Lovell, G., Brown, S. and Dail, D. (1994) Equine color perception revisited. *Applied Animal Behaviour Science,* **42,** 61-65.
9. Smith, S. and Goldman, L. (1999) Color discrimination in horses. *Applied Animal Behaviour Science,* **62,** 13-25.
10. von Frisch, K. (1950) *Bees: Their Vision, Chemical Senses and Language.* New York: Cornell University Press.

Visual discrimination learning in the water maze: an adjunct to spatially dependent cognitive tasks

D. Harbaran, L. Robinson, G. Riedel
School of Medical Sciences, University of Aberdeen, Aberdeen, Scotland, UK.

Abstract

We have developed a training procedure enabling visual acuity to be tested in the water maze [1]. Rats and mice were trained to discriminate a cue card of 1 cm wide black and white stripes from cards with 2, 3, 4, 5, and 10 cm wide striation - an acuity of 1.3 cycles per degree (c/deg). Two mg/kg scopolamine hydrobromide (scopolamine) severely impaired rats and mice in differentiating between striations of various widths, especially when discrimination was difficult, i.e. 1 vs. 2 or 1 vs. 3 cm striations. At a dose of 0.2 mg/kg, scopolamine impaired rats in a spatial reference memory paradigm in the water maze [2], but did not affect visual acuity. To demonstrate that the discrimination paradigm specifically utilises visual processing, and not on other sensory cues, rats, implanted with cannulae targeting the lateral geniculate nucleus (LGN), were infused with 0.4 ug/ul muscimol. These rats were impaired at discriminating all striations; this effect was reversible. These data suggest that it is possible to dissociate drug-induced effects on memory from changes in sensory perception, and that our task is solely dependent on visual processing.

Keywords

Visual discrimination; Water maze; Spatial learning; Scopolamine; Lateral geniculate nucleus; Rats; Mice

1 Introduction

Many learning and memory tasks using rodents rely on normal processing of visual cues. This is particularly so for water maze studies, which access spatial cognitive abilities. When investigating pharmacological compounds in this task that interfere with the cognitive abilities, it is therefore paramount to exclude drug-induced side effects on the visual system. We have previously developed a training task enabling visual acuity to be tested in the water maze [1]. Rats were trained to discriminate between two cue cards containing a pattern of vertical black and white stripes 1 and 5 cm wide; cards were presented in two adjacent quadrants of the pool separated by a barrier with the escape platform located in front of the 1 cm wide striation. The rats were trained to 80% correct responses criterion, then the 5 cm wide striped card was randomly replaced by cards with striations of 1, 2, 3, 4, 5 and 10 cm width. Both rats [1] and mice [2] learned the discrimination with acuity of 1.3 c/deg.

With such a robust visual discrimination task developed in the water maze, we have now applied the task to reveal the sensitivity of the visual system to muscarinic receptor blockade in rats and mice. Furthermore, we demonstrate that this discrimination task is dependent solely on visual processing, and not on other sensory cues.

2 Materials and methods

2.1 Subjects

Seventeen, adult male Lister-hooded rats, and six, adult male C57BL/6J mice were used, all trained on the visual discrimination task to an acuity of 1.3 c/deg [1,2].

2.2 Surgery

Eight of the seventeen well-trained rats were anaesthetised with tribromoethanol, and were implanted with bilateral cannulae into the LGN: AP –4.2; ML ±3.6; DV –4.3 [see 3 for details]. Following surgery, the rats were given one-week recovery before testing began.

2.3 Apparatus

The water maze (150 cm Ø; 50 cm height) was filled with water (21±1 ºC for mice; 25±1 ºC for rats) so that a clear Perspex platform (10 cm Ø; 34 cm height) was submerged 1 cm below the surface. White curtains surrounded the pool throughout the experiments to obscure extra-maze cues. Two cue cards (40 cm x 44.5 cm) containing vertical black and white stripes of varying widths (1, 2, 3, 4, and 10 cm) were mounted to the side of the pool. The cards were positioned in the centre of two adjacent quadrants. A clear Perspex barrier extending from the wall to the centre of the pool separated the two quadrants containing the cue cards; this prevented the animal from readily crossing from the error to the target quadrant.

Swim paths were recorded with an overhead camera and data stored both on video and online using a video tracker and PC-based software (HVS-Image; Hampton, UK.) for later analysis. Parameters recorded were correct responses, and the strategy used by the animals to perform the task, determined by measuring the time spent in different equidistant zones of the water maze: centre; middle; outer [2]. A correct responses was scored when an animal entered the quadrant containing the 1 cm wide black and white striped cue, and swam onto the submerged platform; an incorrect response was scored if the whole body of the animal entered the incorrect quadrant containing the cue of variable striation widths. For a correct response to be recorded, the animal had to enter the correct quadrant and locate the platform (on first attempt after release) without entering the incorrect quadrant.

2.4 Drugs

All drugs were purchased from Sigma U.K. Scopolamine was dissolved in 0.9% NaCl (saline) for doses of 0.2 and 2 mg/kg. Rats and mice were administered 30 and 20 min prior to testing, respectively, with a volume of 5 ml/kg intraperitoneally (ip). Saline was injected on control days.

Muscimol was prepared to 0.4 μg/μl concentration in sterile artificial cerebrospinal fluid (aCSF). Sterile aCSF was used as vehicle. Muscimol and aCSF were infused directly into the dorsal LGN over 5 min at a rate of 0.2 μl/min on alternate days 15 min prior to the start of testing. On control days, animals received ip. injections of saline.

2.5 Visual discrimination testing

All animals were well trained at the time of testing and had attained a level of 80% correct responses [1,2]. All animals were given 8 daily trials. Rats had maximum swim trial of 90 s with 30 s on the platform and an ITI of 30 s. Mice were allowed a maximum of 60 s to swim to the platform, 30 s on the platform and an 8-min ITI.

Scopolamine 0.2mg/kg and saline: Scopolamine 0.2mg/kg and saline were injected in an alternate sequence and cue cards containing gratings of 1, 2, 3, and 4 cm were randomly presented. Only two control days (with 1 vs. 5 cm striations) were given, one before testing began and the other half way through the testing period on day six.

Scopolamine 2mg/kg and saline: Scopolamine 2mg/kg and saline were injected in an alternate sequence and cue cards containing gratings of 1, 2, 3, 4 and 10cm were randomly presented. Control days of 1 vs. 5cm striations were performed after 2 experimental days.

Rats were tested with both doses of scopolamine while mice were only tested with 2 mg/kg scopolamine.

Inactivation of the LGN: Muscimol and aCSF were injected in an alternate sequence and cue cards containing gratings of 2, 4 and 10 cm were randomly presented. Control days of 1 vs. 5 cm striations were performed after 2 experimental days, with animals being injected ip. with saline.

3 Results

3.1 Effect of scopolamine on visual discrimination in rats

Figure 1A depicts the effects of scopolamine (0.2 and 2.0 mg/kg) and saline on the discrimination in rats of gratings 1, 2, 3 and 4 cm. Analysis showed a main effect of grating [$F(3,32)=61$; $p<0.0001$], which confirmed discrimination of wider striations from the 1 cm grating at the target cue. In addition, we obtained reliable effects of drug treatment [$F(2,32)=50.35$; $p<0.0001$], and an interaction [$F(6,32)=2.364$; $p=0.04$]. Significant differences in gratings 2-4cm were obtained with 2mg/kg of scopolamine relative to saline [all t's> 6.59; p's<0.01]. Discrimination of a 10cm grating, however, was not impaired [$p>0.05$].

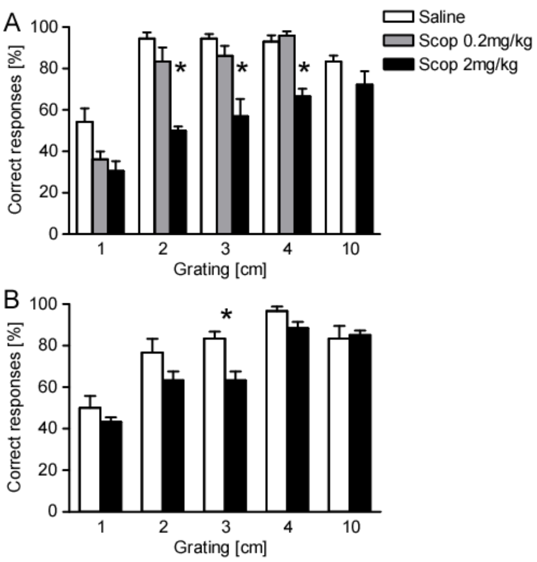


Figure 1. *Scopolamine effects on visual discrimination task. Mean + SEM. (A) Scopolamine (2 mg/kg) impaired rats in discriminating 1 cm from striations of 2 to 4 cm but not 10 cm. 0.2 mg/kg scopolamine had no effect on visual acuity. (B) 2 mg/kg scopolamine compromised visual acuity of mice up to 3 cm wide striations. * p<0.05.*

3.2 Effect of scopolamine on visual discrimination in mice

The effects of scopolamine on visual acuity in mice are summarised in Figure 1B. Mice injected with 2 mg/kg scopolamine were impaired in visual discrimination relative to saline injections [$F(1, 25)=9.29$; $p=0.0054$]. There also was a significant effect of grating [$F(4, 25)=9.29$; $p=0.0054$]. *Post-hoc* analysis revealed that visual acuity was compromised up to 1 vs. 3 cm wide striations but not 1 vs. 4 and 10 cm wide striations.

3.4 Effect of reversible LGN inactivation on visual discrimination

Figure 2 summarises the effects of muscimol infusion into LGN on visual discrimination of gratings of 2, 4 and 10 cm. There was a significant effect of treatment [$F(41\ 21)=40.14$; $p<0.0001$]. Muscimol decreased the percentage correct responses compared to aCSF infusions, with a significant reduction in correctness at 4 and 10 cm gratings. This effects of muscimol was reversible, with aCSF resulting in full recovery of visual discrimination for all gratings in this within-subject designed experiment. On control days (with saline ip. injections) performance was above 75%.

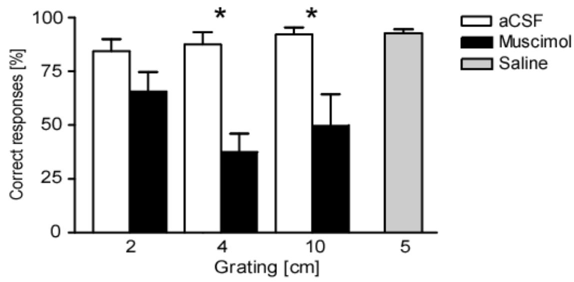


Figure 2. *Inactivation of the LGN impairs visual discrimination. Means + SEM. Muscimol impaired rats in discriminating 1 cm from striations of 2, 4 and 10 cm. Successful discrimination was regained with aCSF infusion showing that the muscimol effect was transient and reversible. Saline injections confirmed that intra-LGN infusion of muscimol was reversible, and that aCSF did not alter visual discrimination. * p<0.05.*

4 Summary

In spatial learning and memory paradigms, animals rely on proximal and distal cues in order to orient towards the goal. The visual system clearly plays a crucial role in this behaviour since visual cues enable the construction of a detailed spatial map of the surroundings and to solve the task successfully. Pharmacological agents that affect performance in such tasks can potentially interfere with either the cognitive or the visual system; it is therefore necessary to establish whether deficits are due to a compromised visual system resulting in the animals being unable to accurately discriminate between distal cues. We here show, that the visual discrimination paradigm developed for the water maze [1] can be successfully applied to this end and comparison of muscarinic receptor blockade in rats and mice established a dissociation of effects of different doses on cognition and discrimination. In both species, visual acuity was compromised by a similar dose, but a ten times lower dose already impaired spatial learning without any notable effects on visual acuity. Other drugs may reveal differences in visual processing between the rats and mice.

Moreover, it appears that the visual discrimination paradigm is only dependent on processing of visual cues, since transient inactivation of the LGN blocked discrimination over all striations suggesting an inability of visual processing. Combining both visual acuity and spatial learning procedures in the water maze may therefore strengthen the dissociation of drug-induced effects on the two paradigms.

References

1. Robinson, L.; Bridge, H; Riedel, G. (2001). Visual discrimination learning in the water maze: a novel test for visual acuity. *Behavioural Brain Research*, **119**, 77-84.
2. Robinson, L.; Harbaran, D.; Riedel, G. (2004). Visual acuity in the water maze: sensitivity to muscarinic receptor blockade in rats and mice. *Behavioural Brain Research*, **151**, 277-286.
3. Robinson, L.; Salt, T. E.; Riedel, G. (2005). Visual discrimination is compromised by transient inactivation of the lateral geniculate nucleus in rats. *Submitted*.

The Influence of Assortment Organization on Product Comparisons and Choice

E. van Herpen, H. van Trijp, and T. Kuipers
Marketing and Consumer Behavior Group, Wageningen University, Wageningen, the Netherlands

Abstract

Organization of product assortments in retail outlets affects consumer perceptions of the assortment and their purchase behavior. Previous research has largely relied on consumer self-report measures and observed purchase behavior, but the underlying mechanism is still to be unraveled. We use observational data (head movements) to examine how the organization affects consumers' "scanning" and particularly the extent to which they make direct comparisons between the items in the assortment. This paper explores how structuring assortments according to natural categories or according to a less salient attribute affects (1) consumer scanning behavior, (2) their assortment perceptions, and (3) purchase behavior from the assortment. Results show that head movements are a more precise measure for direct product comparisons than self-reported measures. Furthermore, organization of organic food products directly next to the non-organic variants leads to more direct comparisons and fewer purchases of organic products, but improved evaluations of assortment organization.

Keywords

Assortment organization, head movements, organic food products

1 Introduction

Assortment organization can influence consumer purchases from the assortment, as well as their perception of the assortment. Previous research has shown that these effects can be substantial, profoundly impacting sales and profit levels [3]. Although it is clear that assortment organization influences consumer purchases, it is not at all obvious which presentation format is best. In fact, several display improvements examined by Drèze et al. [3] actually decreased sales levels.

Items from an assortment can be grouped or structured in different ways. This structure can follow the natural categories that people define, but can also be based on a less salient product attribute. Both types of assortment organization have advantages and disadvantages, which will be examined in this paper. Our application is in the field of organic food products. We focus on two distinct ways to present the assortment: (1) according to product type (natural groups), and (2) according to the organic nature of the products (a less salient attribute).

2 Assortment organization

Products have many attributes that could potentially be used to organize them. Thus, food products could be organized based on price, flavor, size, brand, and so on. The options may seem daunting, but often there is a dominant organization, which most people use. In categorization theory this is called a natural category, and it is the level at which consumers naturally organize items [7]. For products, this is the product type level [9]. Vegetables, for instance, contain as natural categories such product types as onions, carrots, and leek. Retailers can choose to apply this natural product grouping, or to base the organization on a less salient attribute. They could organize a vegetable assortment based on local production, organic nature, or even based on flavor, depending on the attribute they want to emphasize.

2.1 Impact on assortment perception and choice

Presentation format can influence how information is acquired, and consumers have a tendency to process information according to the graphical organization of displays [2, 6, 8]. This implies that organizing assortments according to a specific attribute increases the salience of that attribute. For instance, Areni, Duhan, and Kiecker [1] show that organizing wines by region makes it easier for consumers to use wine region in their decision-making. Products which score favorably on the organizing attribute (e.g., organic nature of the products) may consequently be chosen more often.

Processing is also affected because direct comparisons between alternatives are easier when the physical distance between them is low. The number of direct product comparisons between alternatives should therefore increase with decreasing physical distance. When assortment organization is not based on the natural product types, the resulting physical distance between alternatives complicates their direct comparison. Impressions of quality and price may be affected. More specifically, consumers may not notice it when alternatives in one subgroup are priced higher or are of similar quality as similar alternatives in another subgroup.

Assortment organization can also affect the time and effort required for choice [4]. An organization that is unexpected and not in line with how consumers themselves categorize the products internally is generally liked less [5]. As most stores organize products according to product type and this is consumers' basic categorization as well, evaluations of the display should be highest for this organization and decision effort and time lowest.

In sum, we expect that organizing an assortment according to a less salient attribute (i.e., organic versus non-organic) (1) increases the purchase amount of the alternatives that are preferred on the organizing attribute (organic products), (2) decreases direct comparisons between products of the same type, (3) improves subgroup perceptions on unfavorable attributes that covary with the organization (price, as organic products are priced higher), (4) improves perceptions on favorable attributes that do not covary (quality, as organic and non-organic products in our application are identical), (5) increases decision effort, and (6) decreases display evaluation, compared to an organization according to product types.

3 The experiment

3.1 Method

Forty-four students participated in a 2-group design. Nine participants indicated that they had made the decision for or against organic products beforehand, as they bought organic products very often or never. They were excluded from further analyses, leaving 35 participants.

In a university-related farm, participants entered a 'farm store' selling carrots, leek, onions, apples, eggs, and potatoes, in both organic and non-organic form. These were organized either based on organic versus non-organic items, or based on product type with organic and non-organic variants adjacent. Organic and non-organic products were physically identical, and only varied in the information that accompanied them. Prices were provided and higher for organic than for non-organic products.

Participants were provided with a spectacles camera, and were asked to choose from the assortment, as shown in Figure 2. A fixed camera provided an overview of the layout. Advantages of observation with two cameras include that it (1) is unobtrusive, capturing the visual field of participants without restricting maneuverability, (2) provides information on direct product comparisons, and (3) allows for precise coding using both video streams.

Following a think-aloud protocol and product choices, participants reported the amount of product comparisons they made between organic and non-organic alternatives (three items, α = .88), quality perceptions (three items, α = .89), price perceptions (five items, α = .80), search effort (three items, α = 79), and evaluation of the display (three items, α = .96), all on 7-point scales. Head movements were analyzed and quantified with The Observer 5.

Figure 2. The experimental setup.

3.2 Results

As expected, participants bought more organic products when these were placed separately (denoted by *O*) than when the assortment was organized according to product type (denoted by *T*) (M_O = 2.17, M_T = 1.29; $t(33)$ = 2.19, $p < 0.05$). In addition, quality perceptions for the organic products were higher (M_O = 5.98, M_T = 5.20; $t(33)$ = 2.23, $p < 0.05$). We did not find the expected difference in price perceptions ($p > 0.05$). An organization according to organic versus non-organic products was also met with increased stated search effort (M_O = 4.31, M_T = 2.47; $t(33)$ = 4.21, $p < 0.001$), increased time spent scrutinizing the products (M_O = 121.7, M_T = 87.0; $t(28)$ = 1.99, $p < 0.05$), and decreased evaluation of the display (M_O = 3.24, M_T = 5.25; $t(33)$ = -4.27, $p < 0.001$), all as expected.

We also expected differences in product comparison processes. No such differences were found in the self-reported comparisons ($t(33)$ = 1.24, $p > 0.05$). Yet, analyses of the video data clearly supported our expectation, with many more direct comparisons between organic and non-organic variants when these were adjacent (M_O = 8.5, M_T = 28.8; $t(28)$ = 4.43, $p < 0.001$). Participants were apparently unaware of the degree to which they had made these product comparisons.

4 Discussion

This study has shown that placing organic products separately leads to fewer direct comparisons with the non-organic alternative, and increases the number of organic products that are bought. At the same time, consumers like this presentation format less. This poses a dilemma for retailers who want to promote organic products. Displaying these products separately can induce consumers to buy more of them, but can also lead to dissatisfaction with the organization, which could drive consumers away from the store.

The practical implications of this study go beyond the application of organic food products. Displaying specific items from a larger set either separately or adjacent to a direct alternative is an issue that occurs frequently. Consider the placement of health food items or specific brands (e.g., the Zonnatura brand is placed separately from related product categories in many Dutch supermarkets), but also whether to organize holiday packages according to destination or target group.

Our study also has implications for the measurement of product comparison processes. Self-reported measures appear less appropriate, whereas head movements offer a more precise measure of direct product comparisons.

We thank Noldus Information Technology for providing data recording hardware and software, and Zorgboerderij De Hoge Born (Wageningen) for providing the experiment room and food products.

References:

1. Areni, C. S., D. F. Duhan, and P. Kiecker. (1999). Point-of-Purchase Displays, Product Organization, and Brand Purchase Likelihoods. *Journal of the Academy of Marketing Science*, **27**, 428-441.
2. Bettman, J. R., and P. Kakkar. (1977). Effects of information presentation format on consumer information acquisition strategies. *Journal of Consumer Research*, **3**, 233-240.
3. Drèze, X., S. J. Hoch, and M. E. Purk. (1994). Shelf management and space elasticity. *Journal of Retailing*, **70**, 301-326.
4. Jarvenpaa, S. L. (1989). The effect of task demands and graphical format on information processing strategies. *Management Science*, **35**, 285-303.
5. Morales, A., B. E. Kahn, L. McAlister, and S. M. Broniarczyk. (2005). Perceptions of assortment variety: The effects of congruency between consumers' internal and retailers' external organization. *Journal of Retailing*, **81**, 159-169.
6. Payne, J. W., J. R. Bettman, and E. J. Johnson. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, **43**, 87-131.
7. Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382-439.
8. Schkade, D. A., and D. N. Kleinmuntz. (1994). Information displays and choice processes: differential effects of organization, form, and sequence. *Organizational Behavior and Human Decision Processes*, **57**, 319-337.
9. Sujan, M., and C. Dekleva. (1987). Product categorization and inference making: some implications for comparative advertising. *Journal of Consumer Research*, **14**, 372-378.

Extracting 3D gaze coordinates with real objects in real environments

K. Lukander, J. Virkkala
Brain Work Research Center, Finnish Institute of Occupational Health, Helsinki, Finland

Abstract

Gaze tracking offers significant advantages for studies of spatial attention and visual perception. However, the research field lacks a system for tracking gaze in natural environments. This paper describes a system for tracking 3D gaze on real world objects.

Keywords

Gaze tracking, object tracking, eye-hand coordination.

1 Introduction

We have developed a novel gaze tracking system, capable of tracking gaze on a number of tracked real world objects. The majority of current gaze tracking studies have stimuli presented on a flat screen of a desktop computer. While these can give valuable information about vision and visual attention, they lack the capability to produce data on "real world situations", eye-hand coordination, and with visual target objects at different depths.

The presented system gives accurate gaze coordinates on the surface of the objects. The application areas for the system include usability analysis of handheld devices and tracking cues on visual attention and perception while studying the form, use, and utility of different objects. This work is a continuation of earlier development [6]. The two major advancements to the previous version are gaze and eye position tracking without the use of headgear, and gaze point tracking on the surface of 3D objects.

The system shares some similarities with earlier approaches, including the work of Duchowski et al. [3] on gaze tracking in virtual environments, and the work by Epelboim et al. [4] on gaze tracking while performing natural, visually guided tasks. The approach also offers a partial solution to the general viewing problem in augmented reality (AR) environments [1]. In addition to a detailed description of the scene around the user, the AR systems need an accurate description of the location and optical properties of the viewer (here: the user's eyes) for realistic merging of virtual objects with the real scene.

2 Method

The objective of the ongoing work is to develop a gaze tracking system that

- Frees the user from wires and awkward headgear while allowing for freedom of movement
- Allows tracking the point of gaze on the surface of real world objects
- Automates data analysis

The first objective is met with recent advances in commercial eye trackers. Gaze tracking on the surface of real world objects is achieved through the use of a positional tracker to track object positions. The system integrates the measured data to build a virtual 3D model of the measurement environment. The gaze point is then tracked as the intersection point of the gaze vectors and the virtual 3D objects, corresponding to the real objects. Calculating the binocular gaze points in 3D space requires knowledge of 1) the position of the eyes, 2) the direction of gaze of each eye, and 3) the position and orientation of the gaze target object. A gaze point is then defined as the intersection point of the gaze vector and the gaze target surface.

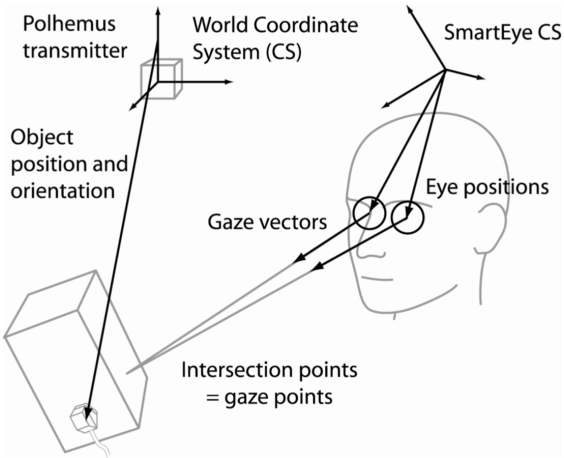


Figure 1 Definition of gaze points

The system is implemented as software running on a standard Windows platform. The current temporal resolution is restricted by the frame rate of the trackers, both currently operating at a 60 Hz frequency, and the system is capable of producing gaze coordinates with adequate accuracy.

2.1 Components

Gaze tracking Tracking the gaze includes tracking the position of the eyes and the direction of their respective gaze vectors. The current system achieves this through the use of a SmartEye 3.0 tracker [8], providing both the binocular eye position coordinates and the gaze vectors. For a well calibrated subject, the SmartEye reports eye position and gaze vector with an accuracy of about 1°.

Object tracking The objects are currently tracked with a Polhemus Fastrak magnetic positional tracker [7], providing a positional accuracy of 0.8 mm RMS and an angular accuracy of 0.15° RMS. The drawbacks of the current tracker are the wires extending from the sensors attached to the objects, the sensitivity to metallic objects and radio frequency fields, and the limited linear range of the tracker (though currently with a radius of up to 2 meters). On the other hand, the tracker provides accurate measurements with no field-of-view issues.

Software The system is implemented as software with Delphi [2], and is composed of a number of modules. Each tracker is abstracted as its own module. Visualization is generated via GLScene, an open source OpenGL library [5]. In addition, saving the data and loading it for playback and analysis is made through the File I/O module. The system includes a separate piece of software for aligning the virtual 3d model to match the real world orientation of

Figure 2 Measurement setup and the virtual model showing the eyes, their gaze vectors and resulting gaze points on the surface of the object.

the object in relation to the Polhemus sensor used in tracking the object.

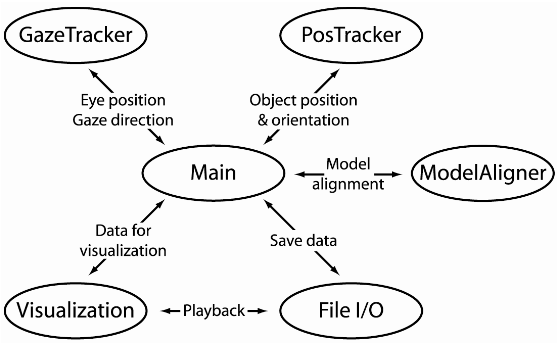


Figure 3 Software structure

2.2 System calibration

The SmartEye (SE) is calibrated for each subject by creating a personal profile that can be reused between sessions. The SE can be calibrated to return coordinates in a user defined coordinate system. This coordinate system is then measured with the Polhemus for transforming SE coordinates to world coordinates.

The system uses precisely built virtual 3D objects to represent the tracked real world objects. Aligning the models to fit to the real world objects is achieved through a custom three point alignment procedure, requiring the measurement and matching of three points on the surface of the real object and the virtual model, and aligning those points through rigid matrix transformations.

2.3 Output data

The system can be configured to output a number of parameters, including: timestamp, binocular eye positions, gaze vectors and gaze points, object position and orientation, and the name of the fixated object/surface.

3 Considerations

The accuracy of the SmartEye gaze tracker for a well calibrated subject is around one degree, corresponding to about one centimeter on an orthogonal surface at an arms length (57 cm). Compared to the accuracy of the positional tracker, this term dominates the error equation. Also, when the object surface and gaze vectors are close to parallel, the error in the estimated gaze coordinates on the object surface can increase considerably. However, as the object can be freely moved and rotated by the subject, this is not expected to be the normal case of examining an object.

The shortcomings in tracking accuracy call for the development of a method for associating gaze vectors, which pass the target object only slightly, with the object surface. This could be made by, for instance, selecting the surface, whose edge is nearest to the gaze vector, and within a threshold distance, as the gaze target, when the gaze vector does not intersect the target object.

Adding an extra dimension to the data calls for definitions for a fixation and a saccade. Two approaches are currently under consideration: 1) Using the eye rotation angles (in head or world coordinates) to classify gaze points as fixations and saccades, or 2) Classifying gaze points within a threshold area on the surface of the target object as a fixation, and movement between these fixations as saccades. The latter approach should be considered in particular because the ability to move the target objects is likely to generate smooth pursuit movements.

4 Future developments

The system can be used for instance to study user interfaces and interaction with small, handheld devices, as well as with larger displays. Another interesting application is the forming of a perception of an object. In the near future, the system will be used to reconstruct classic eye movement experiments, and study new approaches for gaze based interaction.

References

1. Azuma R., et al. (2001) Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, vol **21**, no. 6. p.34-47.
2. Borland website. Online, at *http://www.borland.com (Ref'd 19.5.2005)*
3. Duchowski A.T., Shivashankaraiah V., Rawls T. (2000) Binocular Eye Tracking in Virtual Reality for Inspection Training. *In Proceedings of ETRAF'2000.*
4. Epelboim J., et al. (1995) The function of visual search and memory in sequential looking tasks. *Vision Research*, vol **35**. No. 23/24. p. 3401-3422.
5. GLScene website. Online, at *http://glscene.org (Ref'd 19.5.2005)*
6. Lukander, K. (2005) A system for tracking gaze on handheld devices. *Behaviour Research Methods and Applications*. (Submitted).
7. Polhemus website. Online, at *http://www.polhemus.com (Ref'd 19.5.2005)*
8. SmartEye website. Online, at *http://www.smarteye.se (Ref'd 19.5.2005)*

Versatile pupil size and eyelid closure video measurement system

M. Letonsaari, J. Virkkala, J. Päällysaho
Brain Work Research Center, Finnish Institute of Occupational Health, Helsinki, Finland

Abstract
Usually pupil parameters are measured by special real time hardware. We have developed a versatile video measurement system for pupil size and eyelid closure registration using low cost consumer electronics. Pupil size variations are used in calculating Pupillary Unrest Index (PUI) for vigilance analysis. System estimates also the percentage of eyelid closure, which is an important parameter for estimating sleepiness.

Video image was recorded with an infrared video camera and with a consumer level digital video recorder. Both bright pupil and dark pupil methods can be used. Video was recorded in a laboratory and on field.

Video is analyzed offline using a computer analysis program. Using offline analysis enables more advanced signal processing making the parameter estimation less vulnerable for errors. Several image recognition algorithms are implemented in the analysis program.

Keywords
Video-oculography, vigilance, pupil, eyelid, video analysis

1 Introduction
Eye is a very useful subject for studying human behavior with non-intrusive techniques. Pupil size variation and eyelid movement tell about vigilance and sleepiness. Pupil size correlates also with emotions and cognitive load.

Increased computational power and storage capacity of desktop computers have made it possible to analyze video image. Here we present a method for recording pupil size and eyelid position using low cost consumer level electronics.

2 Methods
2.1 Subjects
The purpose of the measurements was to develop a video-oculography technique for vigilance estimation. Subjects do normal work or simulated work tasks. Work is paused and eye video is recorded with subjects in normal condition and tired. Work performance is then compared to video-oculography analysis results.

2.2 Video recording
In video pupillography there are two main methods, the bright pupil and the dark pupil imaging. In the bright pupil imaging light source is near the camera axel, while in the dark pupil imaging the light source is further away from the camera axel. Also alternating dark and bright pupil imaging can be used to enhance pupil detection [1].

In image recognition, the most important step for correct classification is to get good quality video image. Most often a good lighting is done using an infrared light. Infrared light does not affect the pupil dilation and provides good contrast between the pupil and the iris. In many applications the eye is recorded in dark conditions, which rules out using visible light.

We have used different equipment for recording the eye video. In figure 1 there is a dark pupil video image recorded in the laboratory using Hitachi KP-160 video camera with infrared pass filter and infrared LED array for lighting.

In figure 2 a bright pupil eye image recorded with Sony DCR-HC40 is presented. A built in infrared LED is used for illumination. The distance from camera to eyes is about one meter so that the angle from eye to the camera and light source is small enough to produce bright pupil effect. From this distance both eyes can be fitted into image or a larger area of movement can be allowed.

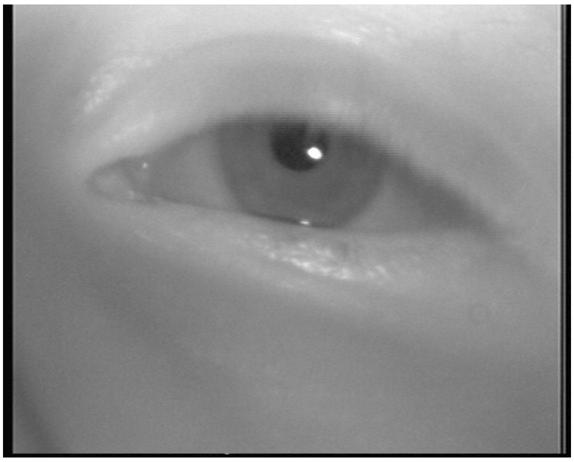


Figure 1. *Dark pupil eye image using Hitachi KP-160 and LED array.*

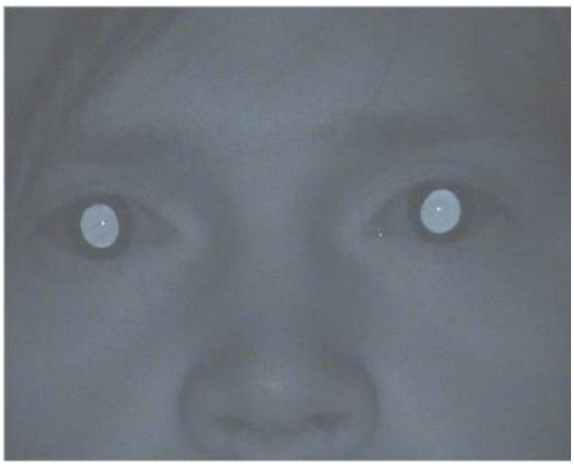


Figure 2. *Bright pupil eye image using Sony DCR-HC40 with built in infrared LED and 'Super NightShot Plus'*

2.3 Video analysis

For video analysis different analysis methods are needed for bright pupil video and for dark pupil image. A computer analysis program was developed with the basic functionality. With some modifications the same program can be used for analyzing both bright pupil and dark pupil images. Also video image with normal daylight condition was analyzed with little change in the algorithm.

Pupil detection

The first step in the pupil detection is accomplished using thresholding algorithm for finding the brightest areas for bright pupil video (and darkest areas for dark pupil video). The algorithm requires constant lighting conditions. On the other hand, in constant lighting conditions a constant threshold provides the most accurate pupil border estimation, because the pupil border is found every time in the same place without adaptation errors.

In image thresholding, dark pupil and bright pupil behave differently. In dark pupil imaging the border including pupil edge can contain eyelid shadow and reflection artifacts, but is constrained into eye. Other dark areas may include nostrils and shadows. In bright pupil imaging skin can be equally bright to the pupil and the border can include the larger part of the face. Corneal reflections must be removed from analysis. This is easier in dark pupil imaging, where corneal reflections are easily detected.

The most critical part of pupil size estimation is selecting the correct edge points for a circle or an ellipse fitting. Different methods found in literature are for example iterative circle fitting [2], iterative ellipse fitting [3], and edge curvature [4]. The previous algorithms are based on the pupil edge properties. Our novel method is based on the fact that every real pupil edge point lies between the pupil and the iris. For this task color values along the normal of a pupil edge point is examined. The iris in the infrared image is a dark homogenous area. Starting from the sides of the pupil, integrity of the pupil border is examined and only proper edge points are chosen for ellipse fitting.

The pupil parameters are calculated by fitting an ellipse to the selected pupil edge points. While circle can be a good approximation of the pupil [2] when gaze is directed to the camera, we have decided to use ellipse fitting to record the smallest changes in pupil. Using ellipse fitting also reveals if selected edge points are not good by deforming the ellipse. Fitting a circle can produce an invalid estimation without revealing the false. Fitting ellipse is done using direct least squares fitting [5].

Eyelid detection

Eyelid detection in the current analysis is an extension of the pupil detection. Pupil location is expected to be known. Iris is found around the pupil as a dark homogenous area. Furthermore the eye white is found around the iris as continuous white area. Starting from the sides, integrity of the iris border is examined. A discontinuation of the iris or the eye white is considered the eyelid. The edge of the eyelid is not as well defined in the image as for example the pupil edge. Exact eyelid position is not usually required though, but good estimate is enough.

3 Discussion

A versatile, low cost video-oculography system was presented. With consumer level electronics it is easy to build a system for infrared pupil video recording. Video analysis is possible with modern computers. There are many algorithms for pupil detection and estimation in literature. The problem is to choose the right one for the task. The idea of the algorithm presented in this paper is to use not only the pupil edge points but to classify areas of interest.

Another extension for video-oculography is to record other variables too in addition to pupil size and location. One such variable is eyelid position, which is relatively easy to measure when pupil is located. This variable is important for example for studying vigilance, where droopy eyelids may make pupil size estimation difficult and eyelid position can tell much more about the state of the subject than the pupil size variation. Eyelid position analysis may be interesting for other purposes too, like studying attention.

There exist other projects too, with the goal of easily modifiable, low cost video-oculography. Project openEyes is an active open-source open-hardware toolkit for eye tracking [6]. Another complete video-oculography project is VIPEA VOG [7], although it is a less active project. There are also instructions on how to convert standard webcam into a near infrared camera suitable for video-oculography [8].

References

1. Morimoto, C. H., Koons, D., Amir, A., Flickner, M. (2000). Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, **18**, 331-335.
2. Lüdtke, H; Wilhelm, B; Adler, M.; Schaeffel, F; Wilhelm, H. (1998). Mathematical procedures in data recording and processing of pupillatory fatigue waves. *Vision Research,* **38**, 2889-2896.
3. Li, D., Winfield, D., Parkhurst, D. J. (2005) Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. *Proceedings of the IEEE Vision for Human-Computer Interaction Workshop at CVPR*, 1-8.
4. Zhu, D., Moore, S. T., Ralphan, T. (1999). Robust pupil center detection using a cur-vature algorithm. *Computer Methods and Programs in Biomedicine*, **59**, 145-157.
5. Halir, R., Flusser, J. Numerically stable direct least squares fitting of ellipses. *Proceedings of the 6th International Conference in Central Europe on Computer Graphics and Visualization*. WSCG '98. CZ, Plzeň 1998, pp. 125-132.
6. openEyes website, Human and Computer Vision Laboratory, Iowa State University. Online, at http://hcvl.hci.iastate.edu/cgi-bin/openEyes.cgi (Ref'd 31.5.2005)
7. VIPEA VOG Project, Department of Neurology, Ludwig-Maximilians-University. Online, at http://www.nefo.med.uni-muenchen.de/~vog/ (Ref'd 31.5.2005)
8. Infra red webcam website, Geoff Johnson. Online, at http://homepage.ntlworld.com/geoff.johnson2/IR/ (Ref'd 31.5.2005)

Comparing two fixation algorithms on eye movement data from a multitask experiment

K. Hirvonen, K. Lukander, J. Virkkala
Brain Work Research Center, Finnish Institute of Occupational Health, Helsinki, Finland

Abstract
The purpose of this study was to investigate the effect of applying different fixation algorithms to the same raw eye movement data, measured while subjects (N=3) performed a five minute multitask with three different workloads. Eye movements were recorded with EyeLink I (SR Research) and the raw data from the left eye has been analyzed with EyeLink and iView (SMI, 3.01.25) software.

The data produced with different analysis tools differ by the number of fixations and fixation times. The resulting spatial data is quite similar. However, different conclusions on cognitive performance and strategies can be made depending on which of the analysis method is used.

Keywords
Eye movement, fixation analysis, multitask, workload

1 Introduction
Identifying fixations from raw eye movement data is an essential part of eye movement data analysis, and can have a dramatic impact on higher level analysis and conclusions on cognitive processes behind those movements [2].

In a recent study we have compared different eye movement measuring methods and devices [1]. Measurements were made with three different measuring methods in time synchrony, during different types of tasks. In this measurement we found out that although there are some differences between measurement methods, they are quite minimal when the analysis has been made with the same analysis tool.

The multitask is used to simulate typical work life situations, where employees have to divide their attention between several different, cognitively demanding functional tasks.

2 Materials and Methods
Eye movements were recorded with EyeLink I (SR Research) while subjects performed a five minute multitask session. Three subjects (one female), aged between 19 and 25 years, performed the multitask with three different workloads.

2.1 Multitask
Our in-house developed Brain@Work -multitask software contains four machine paced tasks: an arithmetic and a memory task, plus a visual and an auditory vigilance task. The layout of the Brain@Work -multitask is presented in Figure 1. Arithmetic, memory and visual vigilance tasks are controlled with a mouse (with the dominant hand) and the responses for the auditory vigilance task were given by the non-dominant hand from the keyboard.

Subject performance was evaluated by task specific scores and the total score. While correct answers increased the score, false or missed answers decreased the score. Three different workloads were used: medium, low, and high. The medium workload was defined individually as a workload where the subject achieved 70% of the maximum score. The low and high workloads were implemented by decreasing (high) and increasing (low) the task specific stimulus intervals by 20%. The duration of the measurement session was five minutes for each workload. Each workload was measured twice in a counter-balanced order.

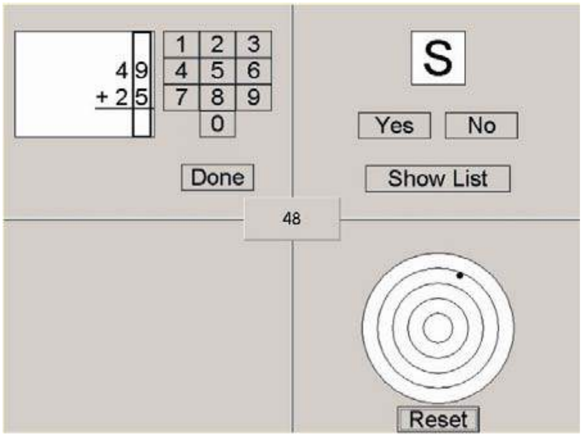


Figure 1. *The layout of the Brain@Work -multitask.*

2.2 Eye movement recordings
Binocular eye movement data was recorded with EyeLink (SR Research) with a sampling rate of 250 Hz. Before each measurement the EyeLink was calibrated with a nine-point calibration. Head movement correction was enabled during the measurements. The monitor was placed 57 cm from the subject's eye. The size of the multitask was 24 x 18 degrees of visual angle.

2.3 Eye movement analysis
The measured eye movement data from the left eye was analyzed with both the EyeLink and the iView analysis tools. EyeLink data file (.edf) is easily converted to iView data format (.dat) with the iView program.

The EyeLink analysis algorithm uses velocity thresholding. Raw data points are identified as a saccade if either the eye velocity or acceleration values are above a certain threshold value (manufacturer's default threshold values were used: acceleration $9500^{\circ}/s^2$, velocity $30^{\circ}/s$).

The iView analysis algorithm uses a spatial window. This algorithm identifies consecutive raw data points as a fixation if they are within the defined spatial window area. The defined fixation is rejected, if the fixation time is shorter than the minimum fixation time threshold value.

iView threshold values were optimized, so that the achieved results were as similar as possible to the EyeLink analysis. We used 10 and 20 pixels as threshold values for the spatial window, and 50 and 80 ms as threshold values for a minimum fixation time.

3 Results

The fixation data produced with the EyeLink analysis tool, and with the iView analysis tool, applying different combinations of threshold values, differ by the number of fixations and fixation durations. However, the spatial fixation data was quite similar between the different methods.

The standard deviation between Eyelink data and iView data with threshold values of 20 pixels and 80 ms was 2 pixels, both horizontal and vertical. For iView data with threshold values of 10 pixels and 50 ms, the deviation was 37 pixels horizontal and 12 pixels vertical. This difference was generated by the iView's discovery of extra fixations.

Longer fixations are chopped up to shorter fixations with the iView analysis tool. However, with threshold parameters of 20 pixels and 80 ms, the fixations are quite similar to the fixations produced by the EyeLink analysis tool (See figures 2 and 3). The iView analysis threshold values of 20 pixels and 80 ms gave the best visual correlation with the EyeLink data (See figures 2 and 3).

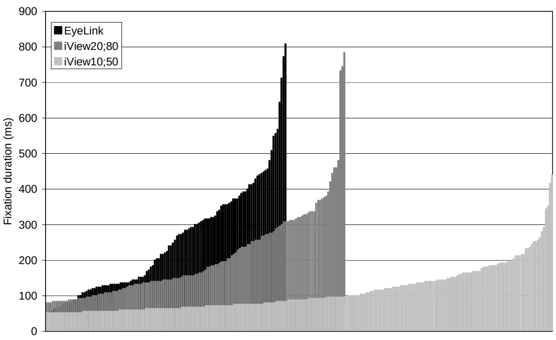


Figure 2. *Fixation durations sorted by length from shortest to longest. EyeLink fixations are presented with black, iView fixations with threshold parameters of 20 pixels and 80 ms are presented with darker gray and threshold parameters of 10 pixels and 50 ms with lighter gray.*

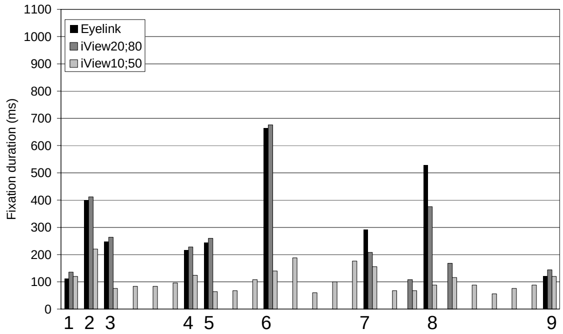


Figure 3. *The same part of the raw eye movement data has been analyzed with EyeLink and iView threshold values of 20 pixels and 80 ms, and 10 pixels and 50 ms (the colors are as in the previous figure).*

4 Discussion

Fixations produced by the iView analysis tool are shorter, because the maximum fixation threshold area cuts fixations with drift to a number of smaller fixations. The EyeLink analysis tool categorizes data points to fixations and saccades based on the eye's angular acceleration and velocity values, and tolerates small amounts of movement within a fixation.

Different conclusions on cognitive performance and strategies can be made depending on which analysis tool has been used. Figures 2 and 3 present fixation profiles analyzed with different tools and parameters.

As a conclusion, each fixation analysis tool has a characteristic way of accentuating specific features from the raw data; therefore it is essential to pick the correct analysis tool for any single study.

References:

1. Hirvonen K, Virkkala J, Näsänen R. (2001). Simultaneous Electro-oculography and Video-Oculography Recordings. *11th European Conference on Eye Movements*, Turku, Finland. August 22-25.
2. Salvucci D, and Goldberg J. (2000). Identifying Fixations and Saccades in Eye-Tracking Protocols. *Eye Tracking Research and Applications Symposium*, Palm Beach Gardens, FL. November 6-8.

Behavioral and heart rate responses to novel stimuli in Arabian mares: effects of age and sire

M. Budzynska, J. Kamieniak, W. Krupa, M. Sapula
Department of Ethology and Technological Basis of Animal Production, Agricultural University of Lublin, Lublin, Poland

Abstract
The reactivity of 54 horses to optic and acoustic stimuli was assessed using an excitability test. Equine behavior in this test is characterized by an index of response to respective stimuli. The response was evaluated using a point scale. Moreover, the heart rate (HR) of every horse was recorded using a telemetric system. Based on the results we assume that age and sire origin have an influence on behavioral and HR responses to novel stimuli in horses. Differences in reactivity according to sire origin should be considered in breeding plans and is especially important for breeding Arabian horses which are relatively low in number and genetically consolidated.

Keywords
Behavior, heart rate, stimulus, horse.

1 Introduction
Horses could survive in open grasslands owing to their flight behavior. Domestication has modified this natural behavior but there is still a wide variation in stimulus threshold and in intensity and kind of reaction of these animals that should be considered in breeding plans [1].
The aim of the study was to assess behavioral reactivity and HR response to novel stimuli in Arabian mares in relation to their age and sire origin.

2 Material and methods
Fifty-four Arabian mares, aged between 2 and 15 years, were subjected to an excitability test by Budzynski's method [2] consisting of three sessions. The test was conducted at the arena where testing equipment was located. Testing area was limited by 2 boards and stimuli generating devices were located just behind them. Testing procedure started/ended when a horse passed the boards being led by a stableman.

2.1 First session
Every horse was introduced to a moving optic stimulus – a horse was walked by a stableman along a path between 2 rotating (40 rotations per minute) black-white squares (1m x 1m). The optic objects were hidden by vertical boards to prevent seeing them by a horse before testing procedure.

2.2 Second session
Every horse was introduced to an acoustic stimulus that was generated when the horse was passing the boards mentioned earlier. The stimulus was generated with a frequency of 80 tones per minute and an intensity of 90 dB by an tonometer.

2.3 Third session
Every horse was introduced to both mentioned stimuli simultaneously.

2.4 Rating of equine behavior
Horse behavior in every session was assessed according to a 10-point scale (1-2=panic, jumping, fearful, not passing next to respective stimuli; 3-4=skittish, but passing; 5-6=careful, but passing; 7-8=fearless, with attention; 9-10=cool, unimpressed, no attention to stimuli. The behavior was rated when the horse was passing next to respective stimuli. Every horse was tested two times in every session, so the response was rated also two times. As a total score for every session we calculated the sum of the two ratings. Based on this total score individuals evaluated from 1 to 8 points (a fearful horse does not want to pass next to the source of stimuli) were classified as excitable ones, from 9 to 15 points (a horse is slightly excited by the generated stimuli but there is no problem with its handling) as medium-excitable and from 16 to 20 points (a fearless horse is not excited by the stimuli, easy to handle) as quiet ones.

2.5 HR measurement
Every horse was equipped with a Hippocard Polar Sport Tester PEH 4000 to determine HR level (heart beats/min).

2.6 Statistics
Significance of differences between mean scores was assessed by the Student t-test.

3 Results
Based on the results of the response to the moving optic stimulus, 24% of studied mares were classified as quiet ones, 57% as medium-excitable horses and 19% as excitable ones.

Age of mares	**Points**	**Heart beats/min**
2-5 years old n=21	10.7±3.0 a	136.9±35.9 b
6-9 years old n=26	13.0±3.7 a	111.7±34.3 b
10-15 years old n=7	11.4±3.3	130.4±24.3

Table 1. *Means (±SD) of behavioral and HR responses to the optic stimulus in mares in relation to their age (means with the same letter in the column differ significantly at P≤0.05).*

The most violent response (behavioral and HR) to the optic stimulus was shown by the youngest mares and a significantly lower response in 6-9 years old mares (Table 1).
According to pedigree analysis the mares were divided into 6 sire lines. Significant differences in horse behavior to the optic ($P \le 0.05$) and acoustic ($P \le 0.01$) stimulus were

found between mares of stallion 'Arbil' line and mares of stallion 'Eldon'. There were significant differences in HR response to the optic stimulus between particular sire groups. There were no significant differences in behavioral and HR responses of mares to simultaneously generated (both optic and acoustic) stimuli in relation to the mares' age and sire line.

4 Conclusions

We conclude that age and sire origin had an impact on behavioral and HR responses to novel stimuli in horses. Futhermore, differences in reactivity according to sire origin should be considered in breeding plans. This is especially important for breeding Arabian horses which are relatively low in number and genetically consolidated.

References

1. Beuing, R.; Matthe, A.; Pracht, P.; Erhardt, G. (1998). Validation and application of a test for self-confidence in Warmblood horses. *49th Meeting of EAAP*, Warsaw, Poland, 24-27 August.
2. Budzynski, M. (1984). Excitability test used to evaluate the nervous balance level in horses (in Polish). *Med. Wet.*, **40**, 156-158.

Refinements in telemetry procedures: a new resource

P. Hawkins [1], on behalf of the BVA(AWF)/FRAME/RSPCA/UFAW[2] Joint Working Group on Refinement (JWGR)
[1] *Research Animals Department, RSPCA, Southwater, UK*
[2] *British Veterinary Association Animal Welfare Foundation, Fund for the Replacement of Animals in Medical Experiments, Royal Society for the Prevention of Cruelty to Animals, Universities Federation for Animal Welfare*

Abstract
Telemetry can benefit science and animal welfare by reducing stress caused to animals (*e.g.* by restraint), which leads to better quality data, and by enabling reductions in animal numbers. It also provides indicators of animal wellbeing that can be used to refine both husbandry and procedures, in order to reduce suffering and improve welfare.

However, telemetry has an impact on individual animals, which needs to be taken into account and minimised. Animals can be affected by surgical implantation or attachment procedures; the physical impact (mass and volume) of the device; or distress caused by technical constraints on providing group housing and environmental enrichment.

A group of experts from industry, academia and welfare organisations convened by the JWGR has published two reports setting out best practice for refining telemetry procedures and for improving the husbandry of animals used in telemetry studies.

Keywords
Analgesia, animal welfare, husbandry, refinement, telemetry

1 Introduction
Telemetry is defined as the use of internally or externally mounted devices for transmitting or storing (logging) physiological data from experimental animals in the laboratory and in the field. The technique is widely viewed as benefiting science and animal welfare because it can reduce stress caused to animals, enable reductions in animal numbers, and provide indicators of animal wellbeing to help implement humane endpoints.

However, the use of telemetry does not represent a true refinement – for either welfare or science – unless every effort has been made to reduce any discomfort or distress associated with implanting or attaching the device, reduce the physical impact of the device itself and ensure that animal housing and husbandry promotes good welfare.

This paper will provide an outline of a new resource which aims to help scientists, animal technicians, veterinarians and members of ethics or animal care and use committees to refine all aspects of telemetry projects [3,5].

2 Refining telemetry procedures
There are two principle sources of harm associated with telemetry procedures that can be refined to reduce suffering and distress; (i) the surgical implantation or attachment procedures and (ii) the physical impact of the device on the animal once it has been fitted.

2.1 Refining surgery
Appropriate surgical skill and empathetic postoperative care are absolutely essential if telemetry is to be conducted humanely. A sound basic training in experimental surgery and good working knowledge of the device to be implanted are essential before attempting to conduct implantation surgery. Training materials produced by commercial companies should be used with caution as they may portray complex surgery that requires intensive training in practice. It may be preferable to delegate telemetry surgery to a specialised surgical team; continuous assessment of surgical performance and monitoring of success rates are both essential regardless of who conducts telemetry surgery.

Strict aseptic technique is essential where a foreign body is chronically implanted into an animal. Most infections originate from the skin, so special care needs to be taken when sterilising implants and the skin at the operation site. Infections usually occur within the first 8 hours after surgery, so antibiotics should be administered before surgery begins.

Low solubility, volatile anaesthetics such as isoflurane are the agents of choice for telemetry studies because they permit effective control of depth of anaesthesia, are rapidly eliminated and are not (or are only minimally) metabolised [1].

Surgical procedures cause discomfort and pain, even with appropriate analgesia. Pain causes suffering and distress, slows recovery, can reduce food and water consumption, interferes with respiration and can slow healing. Appropriate pain relief following all surgical procedures on all species - including rodents - is therefore essential. Despite this, analgesia is still sometimes not provided following surgery. Reasons for this and current thinking on pain management are summarised in the table below [2,4].

Historical reasons for withholding analgesia	**Current thinking**
Pain has a protective function - if it is relieved, animals will become too active and damage incision sites.	Analgesics rarely completely relieve pain - sufficient protective function will remain to prevent damage. Alleviating pain will speed recovery of normal physiological functions.
Analgesics may have undesirable side effects.	Side effects can almost always be avoided or managed by choosing appropriate agents and doses or by temporarily modifying husbandry protocols.
There is a lack of guidance on safe doses.	More information is now available on a range of agents and doses, due to increased concern about pain in animals.

Two especially important factors in acute pain control are the timing of analgesic administration and the way in which different classes of analgesics are combined. Post-surgical pain can be largely prevented by administering pre-emptive analgesia *before surgery begins*, in conjunction with post-surgery top ups, instead of providing post-surgery pain relief only. Multimodal pain therapy, using two or more different agents that act on different parts of the pain system (*e.g.* an opioid and a non steroidal anti-inflammatory drug), relieves discomfort and pain far more effectively [2].

Good post-operative husbandry and care is critically important for speeding recovery following telemetry surgery. A stimulating but comfortable environment will help to shift animals' attention from any discomfort and social stimulation is also important. Groups should be re-formed as soon as possible after surgery (see 3.1 below).

2.2 Reducing device impact

The physical impact of a device on an individual animal will depend on its size and mass, its shape, the material that it is made from and where it is placed on or in the animal. Adding extra mass to an animal's body can have significant adverse effects and cause discomfort and distress, especially in small species such as rodents. The device mass that mice, in particular, are expected to tolerate is disproportionately large in comparison with other species. It is essential to keep up with technological developments to ensure that the smallest device possible is always deployed and to explore the possibility of using alternative techniques such as passive transponders or non-invasive methods.

Even where the mass of a device may not have a significant effect on an animal, welfare may still be compromised if the device is not coated with high quality, biocompatible materials or is an inappropriate size or shape. For example, a 120 mg temperature transponder measuring 2.2. by 14 mm is only 1 % of the mass of a 3 week old inbred mouse, but would be likely to cause discomfort and interfere with normal posture and locomotion. Device length is especially important for species that curl up when sleeping and/or recovering from surgery, which includes most rodents.

It is also important to "balance" the device within the animal as far as possible and consider how it will affect and be affected by behaviours such as locomotion, social interactions and grooming. Subcutaneous implants are generally well tolerated by rodents, but large subcutaneous implants in any species can lead to necrosis of the overlying skin. Careful observation of animals, a literature search and consultation with colleagues may all be necessary to ensure that devices are situated in the best possible place for the animal.

3 Refining husbandry for animals used in telemetry procedures

Animals used in studies involving telemetry are sometimes individually housed, which causes significant distress in social animals. There are two main reasons for this; (i) concerns that animals will damage devices with exterior components and (ii) the fact that most commercially available devices currently transmit at the same frequency. Both of these issues can usually be addressed and so social housing should be the routine system for animals used in telemetry studies unless there are sound scientific or welfare grounds for single housing.

3.1 Reducing the risk of damage from conspecifics

Groups of animals should be re-established as soon as possible after device implantation or attachment. If implantation surgery has taken place, it is important to ensure that animals have fully recovered from the anaesthetic (a 24-hour period of single housing is recommended) and that wound closure has been refined, *e.g.* by using subcutaneous sutures. This, coupled with adequate pain relief, minimises the risk that animals will interfere with their own or each other's wounds.

In the case of devices that have partly or entirely external components, successful group housing can be facilitated by making sure that animals are habituated to harnesses and jackets and that these have been properly fitted. In all cases, the welfare and scientific benefits of group housing usually outweigh the perceived risks.

3.2 Interference between devices

Single housing should not be regarded as an acceptable solution to devices that all transmit at the same frequency. Potential alternatives are the use of the "buddy" system (where animals are pair-housed and one individual is implanted) or the use of devices that can be switched on and off *in situ* and used one at a time in pair- or group-housed animals. In addition, many manufacturers are aware of the demand for implantable devices that transmit at different frequencies, so it is essential regularly to review what is available in this respect.

4 Telemetry refinement resources

In addition to the two published reports that form the basis for this paper [3,5], there is an accompanying online resource that lists the contents of both reports and sets out

- legal issues associated with biotelemetry studies;
- an example of good practice in writing a materials and methods section for publication that includes details of refinements;
- guidance notes for producing project proposals involving biotelemetry.

All of these can be downloaded at **http://www.lal.org.uk/telemetry/index.html** - other resources produced by by the JWGR that describe good practice for laboratory animal care and use can be found at **http://www.lal.org.uk/workp.html**

References:

1. Flecknell, P.A. (1996). *Laboratory Animal Anaesthesia*, 2nd edn. London, Harcourt Brace & Co.
2. Flecknell, P.A. & Waterman-Pearson, A. (eds). (2000). *Pain Management in Animals*. London, W.B. Saunders.
3. Hawkins, P. *et al.* (2004). Husbandry refinements for rats, mice, dogs and non-human primates used in telemetry procedures. *Laboratory Animals* **38**, 1-10.
4. Hellebrekers, L. (ed.). *Animal Pain*. Utrecht, Van Der Wees.
5. Morton, D.B. *et al.* (2003). Refinements in telemetry procedures. *Laboratory Animals* **37**, 261-2.

Reduced alertness while driving: Towards a composite behavioural/psychophysiological measure

N. González[1], I. Kalyakin[2], A. Kauppi[3], D. Ivanov[2] and H. Lyytinen[3]
[1]*AGORA Human Technology Center, University of Jyväskylä, Jyväskylä, Finland*
[2]*Department of Mathematics Information Technology, University of Jyväskylä, Jyväskylä, Finland*
[3]*Department of Psychology, University of Jyväskylä, Jyväskylä, Finland*

Abstract
Drivers' performance is greatly impaired under reduced alertness states, which can lastly result in fatal accidents. This research seeks a composite measure of the psychophysiological and behavioral precursors of sub optimal alertness states drivers might experience while driving for long periods of time. In this study, 7 sleep-deprived voluntaries performed a monotonous simulated driving task (~3 h) concurrently to an embedded vigilance task. Simultaneously, an unattended oddball stream of auditory stimuli was presented through headphones. Different psychophysiological and behavioral measures were recorded. Results showed that performance fluctuated widely within and between participants. Overall, alertness decreased as a function of time on task (TOT). This was indicated by increased RTs and omitted responses in the vigilance task throughout the session. Significant correlations resulted between RT and other psychophysiological variables. The results suggest that a composite behavioral/psychophysiological measure of sub optimal alertness can be more reliable than single ones.

Keywords
Alertness, drowsiness, driver performance, reaction time, psychophysiology

1 Introduction
Many road fatalities are attributable to fatigue, drowsiness, or sleepiness experienced by the drivers. Reduced alertness or readiness states are characteristic of these circumstances. In order to palliate this growing problem, attempts are being made to develop alertness monitoring systems [2,3]. Whereas some authors advocate the use of multiple measures to monitor alertness others concentrate on single measures [3]. This research focuses on a composite measure of the psychophysiological and behavioral precursors of sub optimal alertness states the drivers might face while driving.

2 Method

2.1 Participants
Seven voluntary university students (3 male, 4 female) with mean age 24.3 years (range 22-27 years) were recruited through advertisement and were compensated for their participation. They were all healthy without sleep problems tested through the Basic Nordic Sleep Questionnaire (BNSQ) or any other condition that could have prevented their participation. Six of the participants were right-handed and one was left-handed. They had normal or corrected-to-normal vision. On average participants slept 8.5 h during regular working days and their normal bedtime was at midnight. Each of them held a valid driving license. Six participants drove less than 5000 km per year while one drove less than 15000 km per year.

2.2 Electrophysiological recording
An ElectroCap International (10/20 system) was used to place the EEG electrodes on the scalp. The electrodes were Ag-AgCl filled with ElectroGel. Disposable Ag-AgCl Neuroline electrodes filled with ElectroGel were used to place two EOG electrodes and both mastoids. The montage was monopolar using the tip of the nose as the reference. The EEG scalp locations were F3, Fz, F4, C3, Cz, C4, Oz, and left and right mastoids. The EOG electrodes were placed one centimeter above and one centimeter from the outer *canthus* of the left eye. Amplifier settings were 50 K gain with 0.3 s time constant and analog band pass filter of 0.53-35 Hz. The EEG data was converted, i.e., analog to digital (AD), with a 12-bit, 16-channel Tecmar's Labmaster, sampled at 200 Hz and recorded on hard disc by the DSAMP software. Electrode impedances were kept below 5 kΩ. Other physiological measures such as heart rate, respiration, and skin potential were also recorded though these are not reported here.

2.3 Driving task and stimulus presentation
A personal computer (PC) ran the NASCAR 2002 season game under Windows 2000.

***Figure 1. Simultaneous video recording** of drivers' performance, psychophysiological, and the embedded vigilance task. The red light (white spot on a gray-scale printout) at the bottom-left corner is on in this frame.*

The driving scenery was back projected on a screen (1080 x 924 mm) in front of the participant who sat 1250 mm from it. The driving scenario (Figure 1.) represented an illuminated oval racing track at nighttime. There were no other competitors on the track throughout the experiment. The simulated car was controlled with a Logitech MOMO Force steering wheel, throttle and brake pedals. These controls were managed through the WingMan Profiler software. The steering wheel provided force feedback.

Concurrently to driving, a vigilance task was administered by E-Prime running on a different PC. It consisted in a visual choice RT task. Turning on the external LEDs (2 cm diameter), receiving the responses from the buttons on the steering wheel, and sending transistor-transistor logic (TTL) pulses to flag the onset (light on) and offset (response) of the visual stimuli was done through the parallel port. The TTL pulses were stored by DSAMP on one channel synchronized with the physiological data.

Another E-Prime procedure presented auditory stimuli. The purpose of this oddball paradigm was to elicit event related potentials (ERPs) such as the mismatch negativity (MMN). However, these data are not reported here. E-Prime also produced the trigger signals from which DSAMP initiated the recording epochs. The analog output from the sound controller (Sound Blaster) was also fed into one of the 16 channels available on the AD converter and recorded by DSAMP. This was done to ensure accurate time-locked auditory ERP analyses. The digitized data was thus recorded as sequential epochs of 31235 ms.

3 Procedure

The participants were requested to stay awake during the day and night before the experimental session in order to increase the probability of occurrence of reduced alertness episodes during the experiment. Not each of the participants complied with this instruction. Their sleeping time ranged from 0 to 3.5 h. The experimental procedure was explained to the participants and they signed their informed consent upon arrival at the laboratory. The BNSQ and the Epworth Sleepiness Scale were administered before the EEG electrodes and other physiological sensors were applied. Starting between 2300 and 2400 h, they performed a monotonous simulated driving task for approximately 3 h concurrently to the vigilance task. In the latter, participants responded to the onset of a red or green LED light by pressing a corresponding red or green button on the steering wheel with their right thumb finger. The light was presented either at the bottom-left (Figure 1.) or bottom-right sides of the driving scene. Light color, side of presentation, and inter stimulus interval, i.e., 8, 18, or 28 s, were randomized to reduce stimulus predictability. Simultaneously, the unattended auditory stimuli were presented through headphones. Before the experiment proper, familiarization with the procedure was provided for 15 minutes. The participants were instructed to drive at a constant speed of 70 mph, and avoid sharp maneuvers, collisions against the right-side wall and driving off the road. With respect to the vigilance task they were instructed to respond to the onset of the lights as fast and as accurately as possible. Additionally, they were advised to ignore the auditory stimuli.

4 Data processing

The electrophysiological data were first reduced to one score per each 15.5 s time windows. Power spectrum in the frequency bands θ (4.0–8.0 Hz), α (8.0–12.0 Hz), and β (14–30 Hz) were calculated through a fast-Fourier transform algorithm for each time window. The driving incident index per minute (DIX) was scored through inspection of the video on a scale from 0 (absence of incidents) to 8.5 (very severe incident). A trained observer carried out the scoring of 15.5 s windows based on criteria defined a priori. The frequency of eye-blinks per minute (EBM) was scored automatically from the recorded vertical EOG data. The RTs were averaged across blocks of 32 trials. The duration of these blocks was approximately 10 minutes. For the statistical analyses the data were further synchronized and averaged for each 10-minute block, i.e., TOT, resulting in 16 consecutive data points or blocks. The first block was excluded from the analyses allowing for further stabilization of responses.

5 Results and conclusion

Consistent with the literature, performance fluctuated widely within and between participants. Overall, alertness decreased as a function of TOT [1]. In a repeated measures ANOVA corrected with the Greenhouse and Geisser method, TOT showed a significant effect on RT ($p < .05$), responses to the vigilance task were slower as TOT progressed (Figure 2.). Omitted RT responses were infrequent and occurred more often during the second half of the drive.

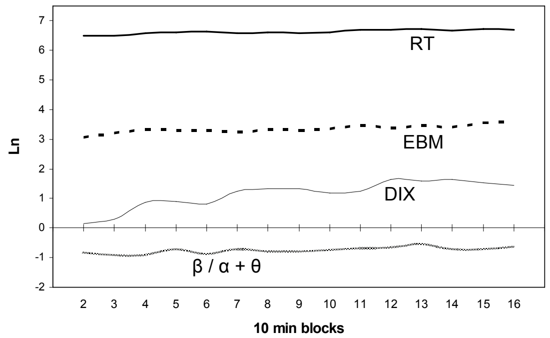


Figure 2*. Averages of RT, EBM or eye-blinks per minute, DIX or driving incident index per minute, and the β / α + θ ratio as functions of TOT. The data have been log transformed for presentation purposes.*

The averages across subjects demonstrated that RT, DIX, EBM, and the EEG ratio $\beta / \alpha + \theta$ tended to increase with prolonged TOT despite some troughs and plateaus (Figure 2.), e.g., blocks 5-9 for the DIX average. Only four of the participants showed increases in DIX towards the end of the session. Moderate but significant correlations ($p < .05$) resulted between RT and TOT ($r = .41$), DIX ($r = .34$), EBM ($r = .27$), and the ratio $\beta / \alpha + \theta$ ($r = .27$), respectively. Despite the caution taken in the interpretation of the results of this exploratory study, we conclude that a composite behavioural/psychophysiological measure might be used to index the sub optimal alertness of the drivers more reliably than single measures.

Supported by the Academy of Finland and the Finnish Henry Ford Foundation (Henry Fordin Säätiö)

References:

1. Campagne, A., Pebayle, T., & Muzet, A. (2005). Oculomotor changes due to road events during prolonged monotonous simulated driving. *Biological Psychology*, **68**, 353-368.
2. Kircher, A., Uddman, M., & Sandin, J. (2002). *Vehicle Control and Drowsiness*. VTI Meddelande 922A. VTI Swedish National Road and Transport Research Institute.
3. Lal, S. K. L., Craig, A., Boord, P., Kirkup, L, & Nguyen, H. (2003). Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Safety Research,* **34**, 321-328.

Recording neural responses to microstimulation in behaving monkeys

J.W. Gnadt[1] and S.D. Echols[2]
[1]*Department of Neurobiology & Behavior, Stony Brook University, Stony Brook, New York, USA*
and Department of Physiology & Biophysics, College of Medicine, Howard University, Washington, DC, USA
[2]*Riverbend Instruments, Inc., Birmingham, Alabama, USA*

Abstract

A fundamental technical hurdle in behavioral neurophysiology has been to record the activity of individual neurons *in situ* while using microstimulation to manipulate neural circuits in intact, behaving subjects. Because electrical stimulation pulses corrupt neural recordings with large artifacts, action potentials can be difficult to detect.

Using an adaptive noise cancellation technique we have developed a real-time method of recording that allows continuous isolation of neuron spikes during sustained experimental microstimulation. The filter learns the corrupting waveform and cancels the spectral content at the applied pattern of stimulation, thus recapitulating the uncorrupted recording in real-time.

This on-line method offers a breakthrough in the arsenal of "circuit busting" tools for behavioral electrophysiology by allowing a cellular level of investigation that is usually available only in reduced preparations. Microstimulation can be used to activate, perturb, or bias neural circuits during active behavior while recording the neural effects.

Keywords

Microstimulation, Adaptive Filter, Noise Cancellation, Spike Detection, Signal Detection

1 Introduction

As reviewed elsewhere[1], systematic electrical stimulation has been useful for revealing the working mechanisms of the nervous system in large vertebrates for over 100 years. In addition, fine "needle" electrodes have been used to record neuronal action potentials from individual neurons *in vivo* for over 60 years. For the last 40 years, techniques of chronic neural recording and "micro"-stimulation have been perfected for the study of neural function during highly trained behaviors in non-human primates and other advanced vertebrates. However, because electrical stimulation pulses corrupt *in vivo* neural recordings with large artifacts, action potentials that fall within the time course of the artifacts are completely obscured or badly corrupted, making them difficult to detect by conventional techniques. We set out to develop a real-time technique for allowing neural recording during experimental microstimulation during active behavior.

2 Theory of operation

2.1 Analytical considerations

Corrupted neural recordings can be modeled as

$$\mathbf{r(t) = s(t) + n(t) + a(t),} \qquad (1)$$

where r(t) is the voltage recording at the electrode (in time), s(t) are the spike waveforms, n(t) is the stochastic background noise inherent in high-gain amplification in biological tissue, and a(t) is the stimulation artifact recorded at each microstimulation event. The noise is a zero-mean process that approximates "pink" noise. On the other hand, the spikes and the artifacts are periodic events that are basically recurring stationary processes. The problem for the neurophysiologist then is to recover the event s(t) from r(t) in the face of corrupting signals n(t) and a(t):

$$\mathbf{s(t) = r(t) - n(t) - a(t).} \qquad (2)$$

Since real-time discriminators are readily available to solve the r(t) - n(t) operation of identifying spikes from electrically noisy recordings, the problem to be solved by our adaptive filter method reduces to

$$\mathbf{r`(t) = r(t) - a(t),} \qquad (3)$$

where r`(t) = s(t) + n(t) represents the traditional "noisy neural recording."

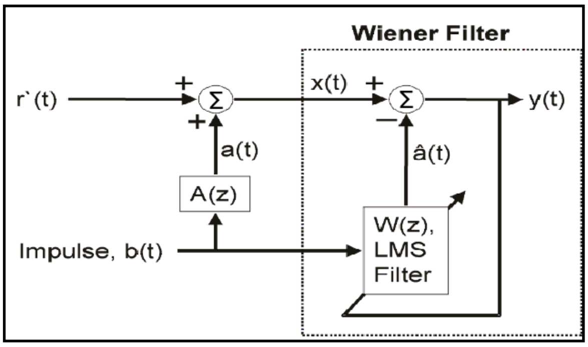


Figure 1. *Adaptive noise cancellation*

In a noise cancellation configuration that approximates a Wiener filter (Figure 1), if one uses a learning algorithm such as the Least-Mean-Squared-Error (LMS) to produce a learned approximation of the impulse response, â(t), and if the timing of the binary trigger event b(t) is known, then dynamic subtraction of â(t) should effectively remove the event a(t) from the ongoing signal r`(t):

$$\mathbf{y(t) = r`(t) + b(t)\,[a(t) - \hat{a}(t)],} \qquad (5)$$
$$\mathbf{if} \quad \mathbf{a(t) \approx \hat{a}(t),}$$
$$\mathbf{then} \quad \mathbf{y(t) \approx r`(t)}$$

Taking the Fourier transform of (5) reveals the filter's spectral cancellation in the frequency domain:

$$\mathbf{Y(f) = R`(f) + B(f)*[A(f) - \hat{A}(f)],} \qquad (6)$$
$$\mathbf{if} \quad \mathbf{\hat{A}(f) \approx A(f),}$$
$$\mathbf{then} \quad \mathbf{Y(f) \approx R`(f)}$$

where the symbol,* , indicates the convolution operation.

2.2 Adaptive noise cancellation

So that the stimulation artifacts could be eliminated in real-time, a hybrid analog/digital circuit was developed that implemented the LMS algorithm as a digital filter, with rapid cancellation by analog subtraction on the throughput (Figure 2). When the filter is activated, each trigger pulse starts an analog-to-digital conversion (A/D) of the input x(t) at 63 KHz. The output of the A/D is applied to a 252 sample 12-bit digital array x(n) with a total time period of 4.0 ms. If the learning switch is activated, a one-shot and/or LMS (Least-Mean-Squared-

Error) learning rule is used to compare the values of x(n) in synchrony with array â(n), such that â(n) is adjusted toward x(n) in user-adjustable increments. Voltage events time-locked to the trigger, b(t), will incrementally accumulate in â(n) according to the learning rule; whereas, voltage values of x(n) uncorrelated in time to the trigger, n(t), will converge toward zero. The output of the learned array, â(n), is converted back to analog (digital-to-analog conversion, D/A) and low-pass filtered thus giving â(t), a learned approximation of the artifact waveform.

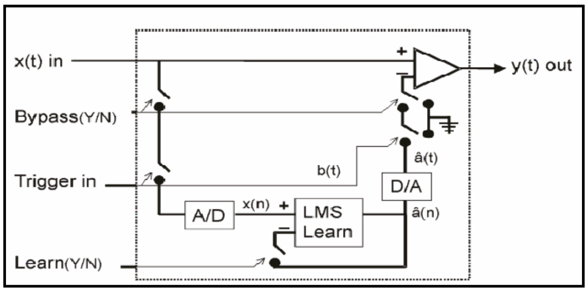


Figure 2. *Filter operation*

3 *In vivo* testing

To test the method *in vivo*, we employed standard chronic neurophysiology in behaviorally-trained rhesus monkeys as described elsewhere.[2] During eye movement behavior, we stimulated in various brain structures while recording the activity of motor neurons in the VI cranial nucleus.[1]

3.1 Artifact cancellation

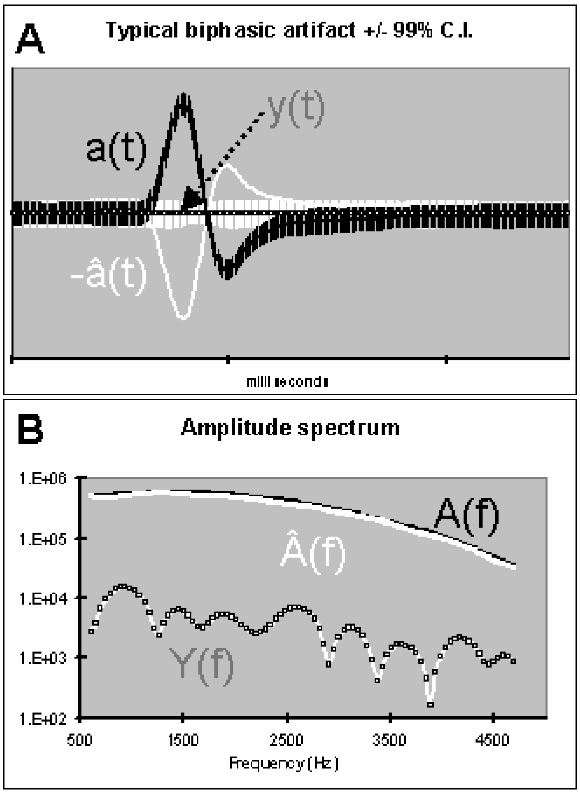
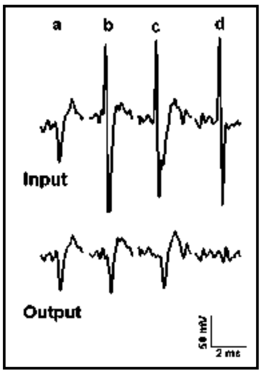


Figure 3. *Spectral cancellation*

Figure 3 shows a typical artifact at the recording electrode, the learned image and the artifact cancellation in the time-domain (top panel) and frequency-domain (bottom panel). The artifact a(t) and its spectrum A(f) is shown in black (error bars = 99% confidence interval); the learned image (inverted) -â(t) and its spectrum Â(f) is displayed in white (error bars omitted for clarity); the average filtered output during cancellation, y(t), and the spectral suppression, Y(f), is plotted with open squares. The average magnitude of artifact cancellation was –31 dB.

3.2 Neuronal spike recovery

Figure 4 demonstrates the problem that is solved by the filter by plotting the input (top) and output (bottom) of the filter for four individual cases: a. shows the faithful reproduction of an uncorrupted action potential., b. and c. demonstrate recovery of action potentials corrupted by artifacts, d. illustrates artifact cancellation, leaving the baseline neural record uncorrupted.



A signal detection analysis was used to test reliability of spike recovery for both types of detection errors: missing action potentials due to corruption and misidentifying stimulation events as neural spikes.[3] The reliability of the spike detection under real recording conditions was better than 1 error in 10^4 (mean d' = 4.36).

Figure 4. *Neuronal spike recovery*

4 Conclusions

The main findings are that: 1) The filter can adaptively learn any arbitrary waveform that is time-locked to a triggered input. 2) The technique removes the spectral content of the corrupting artifact at the applied pattern of stimulation, thus recapitulating the uncorrupted recording in real-time. 3) The filter accurately recovers action potentials from stimulation-corrupted records, regardless of the size, shape, or timing between spikes and artifacts.

This should allow use of the technique with high confidence to recover causal relationships in neural responses to stimulation during active behavior, including understanding dynamic interactions or plasticity changes among neural networks in intact circuits, looking at neural responses during stimulation-induced psychophysical manipulations and *in vivo* investigation of reflex or sensorimotor pathways, or CPGs.

Supported by US National Institutes of Health (EY08217)

References:

1. Gnadt; Echols, Yildirim, Zhang and Paul (2003). Spectral cancellation of microstimulation artifact for simultaneous neural recording *in situ*. *IEEE Bio Med Eng*, **50**, 1129-1135.
2. Ramcharan; Gnadt, Sherman (2003) Single unit recording in the lateral geniculate nucleus of the awake behaving monkey, *Methods, Recording in awake and freely moving animals*, **30**, 142-151.
3. Paul; Gnadt (2003) Reliable real-time spike discrimination during microstimulation. *J Neurosci Meth*, **128**, 191-195.

Dominance status assessed using a food competition test influences plasma cortisol levels and cell-mediated immunity in growing pigs

B.M. Hjarvard[1], O.N. Larsen[2] H.R.Juul-Madsen[1] & K.H. Jensen[1]
[1] *Department of Animal Health, Welfare and Nutrition, Danish Institute of Agricultural Sciences, Research Centre Foulum, Box 50, DK-8830 Tjele, Denmark.*
[2] *Institute of Biology, University of Southern Denmark, DK-5230 Odense M, Denmark.*

Abstract

This paper describes an experiment comparing changes in immune parameters; porcine MHC class II (SLAII), Toll-like receptor type 4 (TLR4), lymphocytes subtypes ($CD4^+$, $CD8^+$) in peripheral blood cells (PBC) besides basal levels of cortisol in plasma, in female growing pigs assigned social ranks using a feeding competition test. Pigs were housed in stable groups of four pigs each for a 6-week period. The immunological variables might be important parameters in an overall "immunocompetence". Plasma cortisol was selected as it relates to social behavior. The duration of gaining access to the feeder in a feeding competition test was used to rank the animals. High-ranking pigs had significantly higher SLAII expression than did middle- and low-ranking pigs. The ratio of $CD4^+$ to $CD8^+$ lymphocytes was lowest in low-ranking pigs. Plasma cortisol was the highest in low-ranking pigs. The results indicate that social ranks in a stable social setting are correlated with immune-endocrine functions.

Keywords: social rank, feeding competition test, plasma cortisol, immune parameters, pig.

Introduction

Psycosocial factors have been shown to affect a number of classical measures of stress besides cell-mediated immune function in growing pigs. Much of the work so far has focused on manipulations to the social environment, especially the effects of fairly severe disruption to the subjects´ social milieu, such as mixing of unfamiliar individuals [3, 6]. The relationship between dominance status and cellular immune function are however less well investigated upon in stable undisturbed social settings. Dominance influences not just specific behaviors in animals, but also physiological processes in the body [1]. This fact might explain that intensity and range of physiological changes induced by various stressors depend on an individuals' social position [4]. Consequently, social rank might offer a natural stress model.

Materials and Methods

Animals and housing

The experiment used 48 female growing pigs entering the experiment at app. 20.5 ± 4.4 kg (58±4 days old). The pigs were housed in groups of four pigs each, hence occupying 12 pens. Groups were composed so that every pig upon entering the experiment were mixed with three unfamiliar pigs. Pigs were fed a standard pig pelleted food, and supplied fresh straw every morning. Food and water were supplied *ad libitum*. Pigs were routine weighted every week during the experiment.

Feeding competition test

After 3 weeks in the experimental groups, at 78±4 days old (28.8 ± 7.3 kg), the rank order of each pig within its group was determined by a feeding competition test conducted in the home pen. The test was adapted from Lawrence et al. (1991) [5], lasted 15 min and was performed in 3 sessions of 5 min each. Pigs were food-deprived before the test by letting their feeders run empty at 15.30 the day before testing. At the start of each session, 150 g of the pig's standard food was poured in the feeder's skid leading to the hole opposite the pen partitioning. It was not possible for all 4 pigs to gain access to the feeder at the same time during the test, and using just the one skid, the access to the feeder was reduced allowing only one pig to feed at a time.
Behavior was recorded on video in order to record the total duration a pig had precedence to the feeder. Each pig's social status was then assigned based on the outcome of summing the total duration of having precedence of the feeder [2]. Pigs having access to the feeder the longest time were determined high ranking (HR), pigs having access to the feeder the least time were determined low ranking (LR). Pigs having access to the feeder an intermediate amount of time were determined middle ranking (MR)
Aggressive acts were scored by direct observation. The scoring included biting (bites directed to all parts of the body), knocking (rapid thrusting the head or snout upwards or sideways to any part of the body), displacing (sideways or backwards ramming or levering a pig away from the through) and non-damaging threats (aggressive interactions, not involving physical contact). For each animal in a group were then calculated an aggression index [8]: X = ½ x (D-R+N+1), where D is the number of other animals to which more aggressive acts were directed than received from, R the number of other animal from which more aggressive acts were received than given to, and N the group size. HR pigs were the ones having the highest score, LR pigs the ones having the lowest score and again MR pigs the ones with scores in between.

Blood collection and analysis

Blood were collected by vein puncture twice; once, just before entering the experiment and second, after 5½ weeks in the experiment. Blood for cortisol assay were drawn only the latter time. PBC were assayed using a whole blood staining procedure and a FACScan flow cytometer (Beckman Coulter). Different monoclonal antibodies were used to bind different types of cells.
Plasma cortisol was measured by radioimmunoassay using the method supplied with the Coat-A-Count® Cortisol (DPC, LA).

Statistical analysis

Initially, Spearman rank correlation analysis was performed to describe the relationship between the variables from the feeding competition test, growth and the two methods of assigning rank. Dominance relationships estimated using a feeding competition test and an intruder test has showed dominance to correlate with aggression in barren environments, and correlated with weight in enriched environments [7]. Following, data on immune parameters and plasma cortisol were analyzed by analysis of covariance. On immune parameters were used the measure of the relevant parameter before entering the experiment as co-variable, while plasma levels of cortisol used growth as co-variable. All analyses were conducted using the SAS system 8e for windows (SAS Institute). Effects were considered significant if P<0,05.

Results

Rank assed using precedence of the feeder were negatively correlated with growth (r_s=-0,29; p<0,05), but showed no connection with aggression. Rank assed using the aggression index showed no relationship with growth or duration of precedence of the feeder. Data are shown in table 1. There was no agreement between the two methods of determining rank.

Table 1 Variables (mean ± SD) extracted from the feeding competition test and total growth across the entire experiment

		Precedence of feeder, s	**Aggression index**	**Growth, kg**
Rank, precedence of feeder	**HR**	594 ± 149	2.6 ± 1.1	26.4 ± 7.5
	MR	136 ± 80	2.5 ± 0.9	25.1 ± 5.7
	LR	11 ± 28	2.4 ± 0.9	20.0 ± 10.4
Rank, aggression Index	**HR**	347 ±313	3.8 ± 0.4	23.2 ± 7.3
	MR	155 ± 202	2.5 ± 0.6	24.0 ± 8.8
	LR	240 ± 253	1.4 ± 0.2	24.4 ± 7.6

These findings led us to consider the data with respect the rank as estimated using precedence with the feeder as the determining variable for further analysis.

Plasma levels of cortisol varied with to rank (p<0,05). Both LR and HR pigs had higher levels than MR pigs. The percentage of SLAII expressing cells varied with rank (p<0,05) and was highest in HR pigs. The percentage of $CD4^+$ and $CD8^+$ lymphocytes decreased over the experimental period. The combined effect of these changes was a decrease in CD4:CD8 ratio that tended to relate to rank (p<0,1). The decrease was most pronounced in low ranking (LR) pigs. The percentage of TLR4 marked cells showed no relationship with rank.

Discussion

The rank-related inverse relationship between percentage of SLAII expression and difference in CD4:CD8 ratio suggests an interconnection between variables affecting the immune system. Rank seems to affect immune and endocrine variables differently. The relationship between psychosocial factors and cellular immune system may have important implications for disease progression and for the management, treatment and selection of pig subjects for studies in which immunological variables are of interest.

The correlation with growth found by using duration of precedence to the feeder as the rank-determining variable made this the determining variable. The lack of correlation with aggression scores indicates that rank is not necessarily demonstrated by force when in a stable social environment. The results also showed, that careful consideration is needed with respect to the amount of food offered per session and the number of sessions performed. HR pigs showed precedence's of app 600 s, corresponding to app 2 sessions. This could be the time and the amount of food a pig of a size used in this experiment needs to fulfil its immediate hunger and hence willing to risk its status on. Considering precedence of the feeder the determining variable, establishing rank in a group of pigs could possibly be automated. Placing a receiver on the feeding though, and equipping the pig with an electronic chip in its earmark could thus read the ID of that exact pig and provide the measure wanted.

References

1. Blanchard, D.C.; Sakari, R.R.; McEwan, B.; Weiss, S.M; Blanchard, R.J. (1993). Subordination stress: behavioural, brain, and neuroendocrine correlates. *Behavioural Brain Research*, **58**, 113-121.
2. Craig, JV. (1986). Measuring Social Behaviour. *Journal of Animal Science*, **62**, 1120-1129.
3. de Groot, J.; Ruis, M.A.W.; Scholten, J.W.; Koolhaas, J.M.; Boersma, W.J.A. (2001). Long-term effects of social stress on antiviral immunity in pigs. *Physiology & Behaviour*, **73**, 145- 158.
4. Hicks, T.A.; McGlone, J.J.; Whisnant, C.S.; Kattesh, H.G.; Norman, R.L. (1998). Behavioral, endocrine, immune, and performance measures for pigs exposed to acute stress. *Journal of Animal Science*, **76**, 474-483.
5. Lawrence, A.B.; Terlouw, E.M.C.; Illius, A.W. (1991). Individual differences in behavioural responses of pigs exposed to non-social and social challenges. *Applied Animal Behaviour Science*, **30**, 73-86.
6. McGlone, J.J.; Salak, J.L.; Lumpkin, E.A.; Nicolson, R.I.; Gibson, M..; Norman, R.L. (1993). Shipping Stress and Social Status Effects on Pig Performance, Plasma Cortisol, Natural Killer Cell Activity, and Lekocyte Numbers. *Journal of Animal Science*, **71**, 888-896.
7. O'Connell, N.E.; Beattie, V.E. (1999). Influence of environmental enrichment on aggressive behaviour and dominance relationships in growing pigs. *Animal Welfare*, **8**, 269-279.
8. Ruis, M.A.W.; te Brake, J.H.A.; van de Burgwal, J.A.; de Jong, I.C.; Blokhuis, H.J.; Koolhaas, J.M. (2000). Personalities in female domestic pigs: behavioural and physiological indications. *Applied Animal Behaviour Science*, **66**, 31-47.

Comparison of telemetry, video scoring and tracking to measure stimulant-induced changes in locomotor activity

Rachael Tanner, Ann Bond, Sally Loomis, Alexander Kulla, Gary Gilmour
Eli Lilly and Company Ltd, Lilly Research Centre, Erl Wood Manor, Windlesham, UK

Abstract
In this paper, we describe techniques that allow the assessment of wake promoting effects induced by stimulants such as amphetamine, caffeine or modafinil in rats. Whereas sleep EEG in telemetrized animals delivers a high resolution of the distribution of sleep and wake parameters, observational scoring of video recordings has the advantage of being able to qualify and quantify behavior. Assessment of activity via video tracking is highly automated but unable to distinguish small movements from resting or sleeping. Restrictions of each system have to be taken into account when classifying compounds.

Keywords
EEG, Telemetry, Video Tracking, Behavioral Scoring, Stimulants

Introduction
Measuring behavior is crucial in the assessment and interpretation of drug-induced changes in animals. The wake-promotion of stimulants such as amphetamine, caffeine or modafinil can be quantified in freely moving rats by measuring sleep EEG [1] with telemetric devices [2]. The assessment of EEG allows a classification into wake, nonREM and REM sleep. Current systems allow accompanying measurements of heart rate, body temperature and unspecific activity. However, the mere assessment of vigilance states with this electrophysiological technique leaves the nature of the wakefulness induced undetected. Scoring of behavior by observation in real time or from video recordings allows its classification and quantification according to a catalog. This pure observational assessment produces clear categorized counts but is laborious and inappropriate for the unambiguous measurement of resting or wake/sleep transitions. The monitoring of locomotor activity via video tracking in turn is highly automated but does not qualify behavior sufficiently. Here, we present an exemplary description of the different methods mentioned. All experiments were performed under the UK Animals (Scientific Procedures) Act 1986.

Methods and Result

Rat sleep EEG
Male Lister hooded rats (Harlan, NL) weighing 350- 400g at the time of surgery were implanted with radio-transmitters sending EEG, EMG and ECG or temperature signals (Data Sciences International, US, TL50-EEE or TL50-EET). Briefly, under general anesthesia, stainless steel screws were placed in the scull as frontal-occipital montage to record the EEG. For the EMG, transmitter leads were inserted bilaterally into the nuchal muscle and ECG electrodes were placed in the left neck area and right gluteus muscle, respectively. Rats were allowed at least 4 weeks of recovery before undergoing experiments.
EEG, EMG and ECG were continuously recorded at a sample rate of 500Hz by the acquisition hard- and software (Data Sciences International, US, ART 3.0) for the duration of the experiment.
Unspecific activity counts were automatically derived by the acquisition software and ECG signals were online-converted into heart rate counts. Individual absolute values for activity, temperature and heart rate were extracted as hourly averages and normalized.
EEG and EMG waveform data was loaded into sleep EEG analysis software (Sleepsign, Kissei Comtech, US) and automatically scored every 10 sec according to tested and adjusted algorithms. Hourly averages of the distribution of the vigilant state were extracted for each individual rat.
Data was then grouped according to treatment and displayed as mean, standard error of the mean (s.e.m) and n-number of subjects. Differences of hourly means between vehicle and treatment were assessed by using the student's t-test. (* indicating $p<0.05$).
A group of 15 rats received saline or 15mg/kg of caffeine (ip, 1ml/kg) in a 2x2 crossover design (each animal received each treatment on 2 consecutive weeks, injection time: middle of light-phase). Digital video of each individual rat was recorded from the time-point of injection on.
Here, data for unspecific locomotor activity (figure 1a) and percentage of time per hour spent in wake (figure 1b) are shown. 15 mg/kg of caffeine significantly increased unspecific activity for the hours 1-5 and 8-10 post-injection in comparison to saline treated controls. During the hours 1-4 post-injection, rats spent significantly more time in wake as vehicle controls. This caused a sleep deficit, which in part seemed to have been compensated for during hours 21-23 post-injection.

Video Scoring
A behavioral catalog was created, differentiating between foraging, ambulation, eating, drinking, rearing, grooming and resting. Video recordings (bw CCTV camera, 320x280 pixel, Samsung, Japan) made in conjunction with the sleep EEG from the time of drug injection on were digitized (Geovision G800, Taiwan) and stored on a hard drive. Each 60 sec, the behavior of an animal (male Lister hooded rat) was scored according to the catalog. Absolute values of occurrence per defined time-interval were averaged and grouped according to treatment and displayed as mean, standard error of the mean and n-numbers.
The examples shown here (figure 2) depict the differences in resting (above) and eating (below) between 5 Lister hooded rats receiving vehicle (0.25% methyl cellulose/ saline) and either 50 or 200mg/kg modafinil (ip, 1ml/kg), averaged over the first 120 minutes post-injection. A dose dependent but not statistically significant trend can be observed for these parameters: Modafinil slightly reduces the occurrence of resting behavior but increases the occurrence of eating episodes.

Video Tracking
Locomotor activity was measured in clear Perspex boxes (40x40x30cm) based on infrared fields.

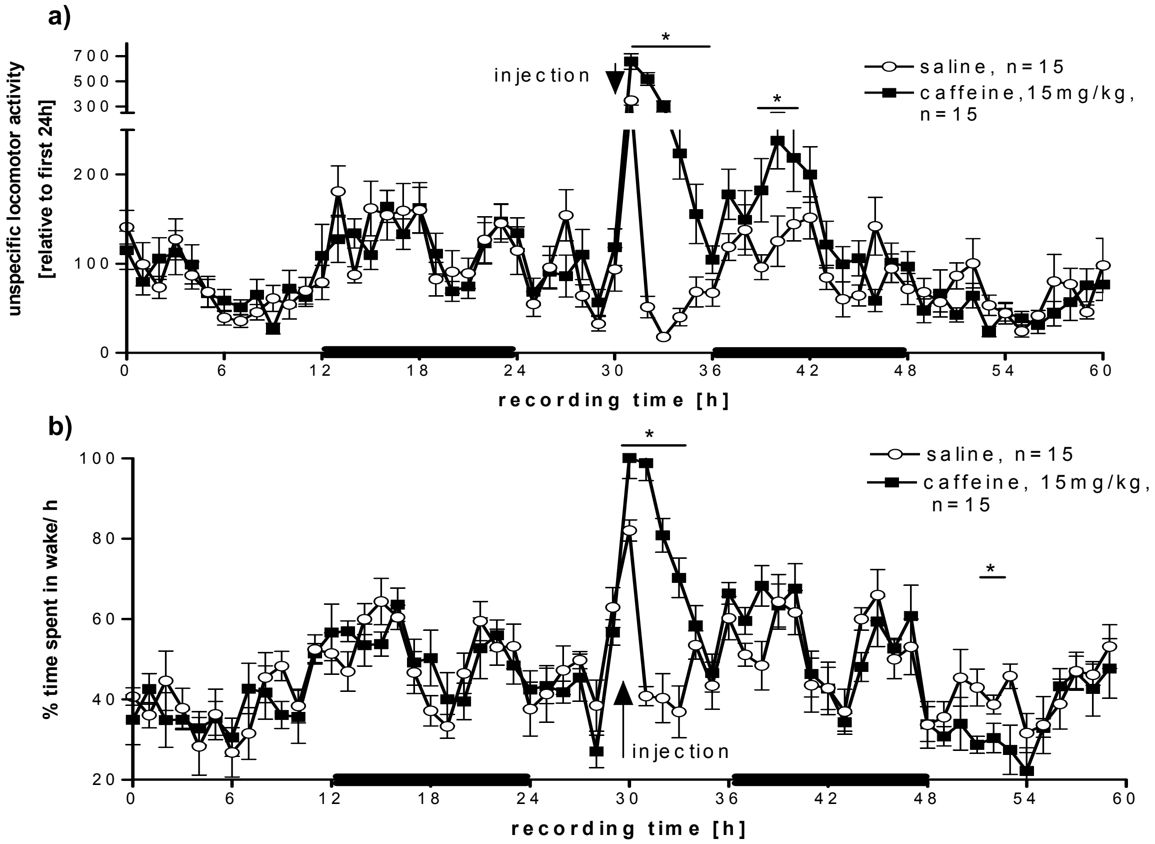


Figure 1: 60 hours of telemetric recordings. Caffeine or saline have been administered (ip) after 30 hours. a) Hourly means of normalized unspecific activity counts. b) Hourly means of percentage of time spent in wakefulness, assessed with sleep EEG measurements. The dark shaded area of the time-axis indicate the dark phases.

Four boxes were placed on each field, which were monitored using overhead infrared cameras. The cameras fed into a Quad compressor unit, which in turn fed a PC running the image analysis application EthoVision (Noldus, NL). EthoVision digitizes the path made by animals and uses this to calculate different parameters of activity.

Here (figure 3), data is shown (mean and standard error of the mean) for 120 min from male Lister hooded rats receiving vehicle (5% glucose/ purified water), 1, 2 and 4 mg/kg amphetamine (8 animals per treatment group) via the subcutaneous route (1ml/kg), 30 min after habituation to the recording boxes. Amphetamine induces a clear dose-dependent hyperlocomotion in the animals.

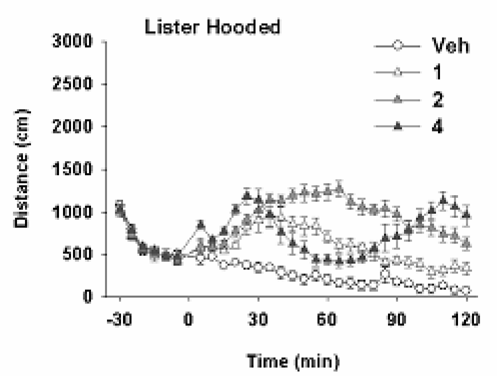


Figure 3. *Total distance moved measured by video tracking of Lister hooded rats receiving vehicle, 1, 2 or 4 mg/kg of amphetamine.*

Conclusion

Each of these briefly described methods for the assessment of drug-induced wakefulness is useful but a full and unambiguous characterization of compounds is only possible with the combined use of these techniques.

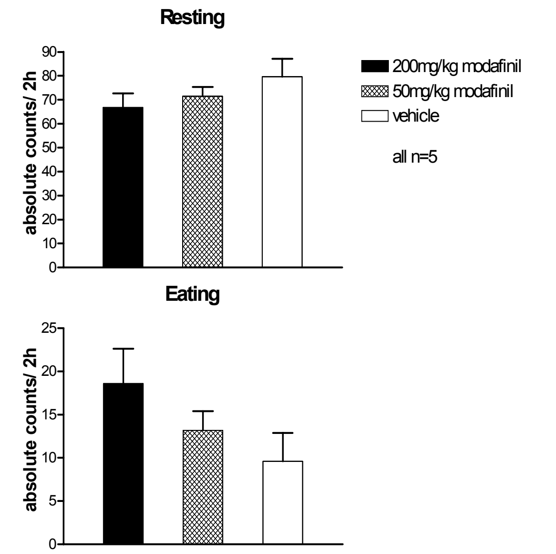


Figure 2. *Example of absolute values of counted occurrences of resting behavior (top) and eating (bottom) within 120 minutes after vehicle, 50 or 200 mg/kg of modafinil assessed by observation of video recordings.*

References:

1. Edgar, D.M.; Seidel, W.F. (1997). Modafinil induces wakefulness without intensifying motor activity or subsequent rebound hypersomnolence. *J Pharmacol Exp Ther*. **283(2),** 757-69.
2. Fitzgerald, A.L.; Juneau, P.; Cain, C.; Southwick, K. (2003). Development of a quantitative method for the evaluation of the EEG of rats by using radiotelemetry. *Contemp Top Lab Anim Sci*. **42 (1)**, 40-5.

Analysis of the human multimodal emotional signals to an interactive computer

R. Ciceri, S. Balzarotti, P. Colombo
Laboratory of Communication Psychology, Catholic University of Sacred Heart, Milan, Italy

Abstract
The present study tries to examine what kind of emotional responses human subjects express while interacting with an artificial agent if they believe that this agent is able to understand their emotional states. For this purpose different kind of computer games were programmed to elicit specific emotional appraisals. 30 subjects were asked to use the computer where an avatar provided a simulated intelligent feedback (experimental condition), or it only guided the subject across the different tasks (control). Multimodal synchronized data were captured: physiological data, posture, facial expressions, vocal behavior. All observational data were codified frame by frame using The Observer 5.0. The Theme software was used for the detection of recurrent patterns. This is an in-progress study: the paper includes initial results of the analysis of non-verbal communicative signals.

Keywords
Non Verbal, Multimodal Communication, Human-Machine Interaction.

1. Introduction
Researchers from several disciplines have turned their attention to emotion and its place in the interaction between human and artificial agents, trying to understand and re-create emotion in the user interface [5] [8].
A particular interest has concerned the implementing and design of interfaces able to recognize the user's emotional state from real-time capturing and processing of sensory modalities input via various media: physiological, facial and vocal signals [6].
This study presents an attempt to define some significant features that can be extracted from the user's expressive behaviour and a model able to describe the dynamic changing of the user's emotions during the computer interaction. We assume that a crucial factor for the interaction with an emotional artificial agent and its effectiveness is the presence of a «communicative agreement». If the user is aware of the artificial agent's ability to understand his emotional signals, he can decide to participate in the interaction exhibiting proper communicative signals.

2. Method

2.1 Partecipants
A total of 30 university students (20-23 years old) were recruited.

2.2 Stimuli Contruction
Three different kinds of computer games were projected to elicit specific emotional reactions. Systematic manipulation of appraisal dimensions (SECS, [9]) was used through the selection of types of game events that were assumed to produce specific appraisals. Specifically, game events were supposed to support four emotional evaluation checks: 1. novelty; 2. hedonic value; 3. goal conduciveness; 4. coping. All games were previously tested on 20 subjects to assess their efficacy.

2.3 Tools
An enabling system was set up to implement experimental sessions. Different kinds of devices were used to record the subject's behaviour (Figure1): 1. two high resolution web cameras: one in front of the subject and one behind him; 2. a high quality microphone to record vocal reports; the BIOPAC System [1] to record physiological signals. All instruments were synchronized.

Figure 1. Experimental situation.

2.4 Procedure
Subjects were asked to use the computer where an avatar guided them across the three different kinds of computer games. All sessions started with 2 minutes of free exploration of a web site for the baseline measure of physiological signals. Total duration was of about 20 minutes.
They were divided into two different groups, according to the kind of information received by the avatar. In the experimental condition, the subjects were exposed to a simulated emotional-intelligent computer, where the avatar provided a simulated intelligent feedback to the user to decode his emotional state and to adapt the tasks accordingly. The simulated emotional-intelligent computer was actually controlled by an out of sight experimenter. In the control condition the avatar only guided the subjects in the activities. All subjects were alone in the room with the computer.
At the end of the computer session, subjects were asked to answer to questionnaire and the Coping Inventory for Stressful Situations [4] [7]. In the questionnaire subjects were asked to assess their own abilities at using a computer; and to judge the efficacy of the interaction with the computer.

2.5 Data analysis
Different kinds of synchronized data were recorded: 1. Physiological signals (ECG, Respiration, Galvanic Skin Response, Skin Temperature); 2. Non Verbal Signals (Posture, Gaze Direction, Facial Movements); 3. Vocal Signals; 4. Verbal Signals.
This is an in-progress study and data are still under elaboration. In this study we'll start to present the analysis of non-verbal communicative signals.

At a first level of analysis, a Behavioural Coding System (BCS) was developed to code the subject's interaction. 30 behavioural units subdivided in four categories were mapped (Table 1). Some units were identified by using specific action units (AU) defined by the Facial Action Coding System [3]. Other units for posture, body movements and vocal behaviour were taken from grids used in previous studies on non-verbal behaviour [10]. All video tapes were codified *frame by frame* (25 fps) using The Observer 5.0 software and Theme 4.0 software for the recurrent pattern analysis.

Facial movements	Upper and lower face units were selected. For each unit intensity was rated (Low, Medium, High).
Gaze direction	To look at the screen, at the keyboard, around, etc.
Posture	Behavioural units of moving near to /far from the screen were considered
Vocal behaviour	It was recorded when the subject speaks or uses other kind of vocalizations (grumbling, no-words, etc)

Table 1: Non verbal behavioural units

At a second level of analysis, we used the Multidimensional Emotional Appraisal Semantic space (MEAS) to detect the user's emotional state starting from the analysis of patterns of multimodal signals [2]. MEAS, derived from SECS, analyses the intersection of the four dimensional axes (Figure 2) scored on a 5 point rating scale from -2 to +2 by a group of judges. Inter-judge agreement is calculated (k Cohen).

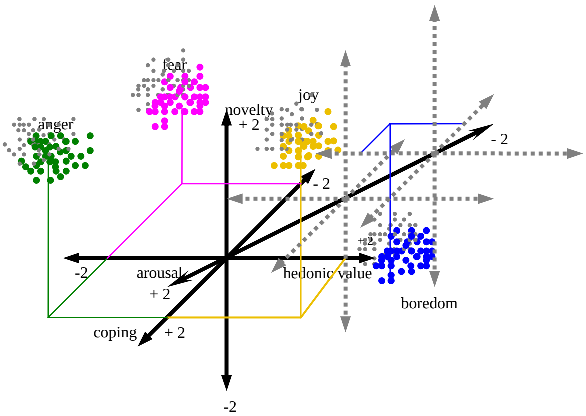


Figure 2: Dimensional axes.

3 Results

This paper includes only initial results (corresponding to the analysis of one subject, 22 minutes and 27 seconds for a total of 26948 frames) which are indicative of the methodology that has been applied.

Results from the analysis of durations show that during the interaction with the computer, the subject exhibits communicative non verbal behaviour. In particular, facial behavioural units (mean number of occurrences, n = 47.5; mean of total duration, td = 33,85 sec) are used more frequently than vocal ones (n = 10.75; td = 10.54 sec) and within these facial movements there is a higher exhibition of lower face units (n = 51.6; td = 41.42 sec). Non verbal behavioural units are exhibited with different frequencies and durations, hence they seem to have different functions and relevance.

The analysis of the most recurrent configurations of the emotional expressions show (Theme analysis: number of occurrences: 10; p<.001) the presence of patterns of different levels of complexity that involve facial movements, posture, etc. (Figure 3).

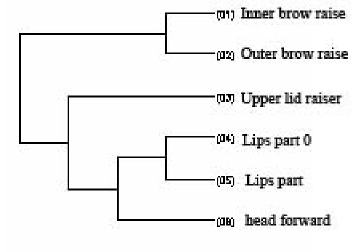


Figure 3: T-pattern of expressive behaviours.

Behavioural units are linked one to another in more complex pattern and expressive configurations. Hence the subject exhibits dynamical facial responses based on the continuous monitoring of the task and performance rather than fixed patterns corresponding to an emotional label.

This research has been conducted in collaboration with STMicroelectronics, the Dipartimento di Scienze dell'Informazione of the Università degli Studi and the Bioengeneering Department of the Politecnico in Milan.

References:

1. Biopac systems inc. http://www.biopac.com.
2. Ciceri, R.; Balzarotti S.; Colombo, P. (2005) Analysis of the human physiological responses and multimodal emotional signals to an interactive computer. *Proceedings of AISB 2005 Annual Conference*, Hatfield, 12-15 April.
3. Ekman, P.; Friesen W.V.; Hager, J. (2002). *Facial Action Coding System. The Manual*. Research Nexus Division of Network Information Research Corporation, Salt Lake City .
4. Endler, N.S; Parker, D.A (1990) Multi dimensional assessment of coping: concepts, issues and measurement. *Fith European Conference on Personality*, Ariccia, Italy.
5. Gratch, J.; S. Marcella, S. (2004). Evaluating the modelling and use of emotion in virtual humans. *Proceedings of the 3rd International joint Conference on Autonomous Agents and Multi agent Systems.*
6. Lisetti, C. L.; Nasoz, F.; LeRouge, C.; Ozyer, O., AlverezK. (2003) Intelligent Affective Interfaces: A Patient-Modelling Assessment for Tele-Home Health Care. *International Journal of Human-Computer Studies*, **59**, 245-255.
7. Pedrabissi, L; Santinello, M (1994) Coping Inventory for Stressful Situations: revision of validity and psychometric properties. *Ricerche di Psicologia*, **4 (18)**, 49-63.
8. Picard, R. W. (1997) *Affective Computing*. MIT Press, Cambridge, UK.
9. Scherer, K.R. (2001) Appraisal considered as a process of multilevel sequential checking. In: K. Scherer, A. Schorr, T. Johnstone, (Editors) *Appraisal processes in emotion: Theory, methods, research. Series in affective science*, London, Oxford University Press, xiv, 478 pp.
10. Wallbott, H. (1998) Bodily expression of emotion. *European Journal of Social Psychology*, **28**, 879-96.

Suprasegmental and verbal features of repeated questions in interview during the process of secret disclosure

R. Ciceri, F.Biassoni
Laboratory of Communication Psychology, Catholic University of Sacred Heart, Largo Gemelli, 1 20143 Milan, Italy

Abstract
The study seeks to investigate the interview process aimed to solicit from the subject the disclosure of his/her secret in order to find out and objectively measure effective strategies to lead and hold the subject during the process of disclosure. attention is focused on the interviewer's questions, both in verbal content and in vocal nonverbal characteristics. Data obtained from the analysis of five interviews show the effectiveness of a strategy called "repeated questions". The repetition of the same question (from the verbal content point of view) brings the interviewer nearer to the subject's secret; acoustic digitalized analysis, then, put into evidence the role of the suprasegmental dimension in this strategy. Three parameters have been considered in acoustic analysis: frequency, intensity and rhythm. Data shows that vocal nonverbal features convey a variety of meanings connected with cooperation and empathy between the subject and the interviewer, so taking up both semantic and relational function.

Keywords
Question, vocal nonverbal, secret, empathy.

1 Introduction
The interview process in secret disclosure contexts has been often analyzed from a "clinical" point of view in order to find out good interview strategies so as not to be suggestive (Cederborg et al., 2000; Davey and Hill, 1999; Coulborn Faller, 1996; Cantlon, Payne and Erbaugh, 1996). The present work takes into account the interview process and the question features from a communicative point of view. In particular, the question is looked upon as complex and a multilayered linguistic act. For this reason, the analysis of disclosure goaled interviews includes both questions content analysis and vocal nonverbal profile analysis.

2 Objectives
The present work has three main objectives:
- to pinpoint questioning strategies;
- to analyze a particular questioning strategy, called "repeated questions", by finding out a possible relationship between verbal expression and effectiveness of the question;
- to analyze the role of the vocal nonverbal dimension in supporting self-disclosure through both furthering comprehension and conveying empathy.

3 Method

3.1 Matherials and sample
Five interviews about sexual abuse have been analyzed; the interviewer was a psychotherapist. All the interviews were written out and all the psychotherapist's questions were drawn out. The first-level sample ("psychotherapist questions") is made up of 667 questions.

3.2 Questions categorization
The 667 therapist questions were categorized depending on three criteria:
- the effectiveness (the power to get an answer);
- the content (about the subject's life, the abuse experience, feelings and thoughts about abuse experience);
- the level of information sharing between therapist and subject pre-existing the interview (Derlega, Metts, Petronio e Margulis, 1993; Cantlon, Payne e Erbaugh, 1996).

3.3 Selection of the "repeated questions"
Questions categorized in level 2 of the criterion "effectiveness" (questions gaining answer after one ore more repetitions) and in level 4 of the criterion "content" (questions about details of the abuse experience) turned out to be the most numerous category (173 questions). They were labelled "repeated questions" (Jakobi, Blanchet, Bromberg, 1988) and moreover analyzed.

3.4 Verbal streak analysis
The sample of 173 repeated questions were initially analyzed from the verbal point of view. Four repetition strategies were pin pointed:
1. "identical repetition": the question is repeated just alike in the verbal streak;
2. "repetition with focus reverse": the same words are used in the repetition, but with different order in the sentence; the result is the change of position of the emphatic stress;
3. "repetition with focusing question": the repetition uses the "funnel technique";
4. "subject-focused repetition": while the first question is general and objective, the repetition is subject focused (e.g. "What was the reason…?", "What do you think was the reason…?").

3.5 Vocal non verbal profile analysis
All the repeated questions were submitted to digitalized acoustic analysis; ten acoustic parameters were quarried:
- four about rhythm (measured in seconds): sentence length, pauses length, speech rhythm and articulation rhythm (rhythm of speech purified of pauses);
- three about pitch (measured in Hz): average, standard deviation and range (lowest and higher value);
- three about energy (measured in dB): average, standard deviation and range (lowest and higher value).

3.6 Instruments
Digital acoustic analysis was performed through CSL (Computerized Speech Lab, version 4300B, Kay Elemetrics Corps, Pine Brook, NJ 07058-2025 USA).

4 Results

4.1 Content analysis

The whole sample of the therapists questions was categorized depending on the three criteria (effectiveness, question content and level of information sharing). The results were submitted to statistical analysis; variance analysis showed that all frequency differences were statistically significant (content: ANOVA: df=4, χ^2= 51,78, p<0,05; effectiveness: ANOVA: df.=2, χ^2= 24,22, p<0,05; information sharing: ANOVA: df=2, χ^2= 25,69, p<0,05).

4.2 Verbal strategies

The 173 "repeated questions" were categorized depending on the four verbal repetition strategies by three judges. Here is the distribution:

Strategy or strategies combine	Frequency and percentage
1	7 (11,7%)
2	3 (5%)
3	31 (51,7%)
4	3 (5%)
1+3	9 (15%)
2+3	2 (3,3%)
3+4	3 (5%)
1+2+3	2 (3,3%)
Total	60

The judges' assessments were submitted to Pearson r test to find out correlation coefficient; all correlations except one turned out to be statistically significant.

Strategy	Correlation coefficient J1-J2	Correlation coefficient J1-J3	Correlation coefficient J2-J3
1	0,632** sig. 0,000	0,540** sig. 0,000	0,544** sig. 0,000
2	0,302* sig. 0,019	0,284* sig. 0,028	0,457** sig. 0,000
3	0,685** sig. 0,000	0,717** sig. 0,000	0,536** sig. 0,000
4	0,704** sig. 0,000	0,416** sig. 0,001	0,169 sig. 0,197

**= (p<0,05)
*= (p<0,01)

4.3 Vocal non verbal strategies

For each strategy or strategies combine an example of question repetition has been chosen and the suprasegmental dimension has been analyzed.
Here is an example of the qualitative analysis of strategy 1 (identical repetition).

Acoustic parameter		Rep.1	Rep. 2
rhythm	sentence lenght (sec)	1,97	1,98
	articulation rhythm (sill/sec)	4,06	4,04
pitch	range- (Hz)	141	92
	range+ (Hz)	294	264
	average (Hz)	220,71	207,14
energy	range- (dB)	52	51
	range+ (dB)	76	73
	average (dB)	63,81	61,86

Strategies combine has been described and the vocal non verbal profile analyzed as well. Here is an example of strategy 1-strategy 3 combine.

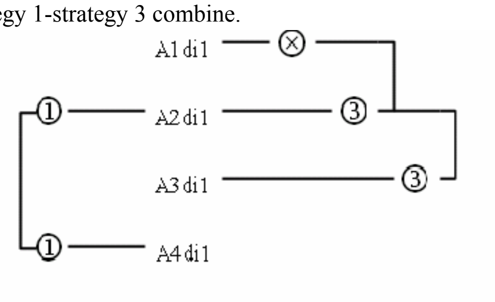


Figure 1. *Strategies combine: strategy 1+ strategy 3.*

Sonographic outlines show different speech depending on suprasegmental profile.

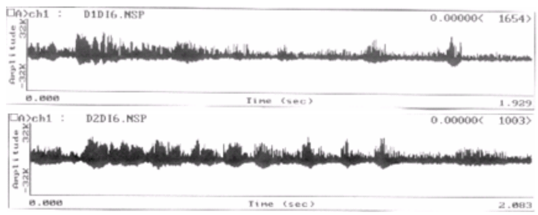


Figure 2. *Sonographic outlines.*

5 Discussion and conclusions

Each one of the repetition strategies turned out to have a distinctive suprasegmental profile. The more interesting strategies proved to be strategy 1, 2 and 3: strategy 1 shows that when the question is repeated just alike in the verbal streak, it is uttered with higher pitch and higher energy. Strategy 2 shows that when the question is repeated changing the words' order in the sentence, the first part of the sentence is uttered with lower rhythm: more time is devoted to the part of the sentence where the focus of the question is.
Strategy 3 shows that the second repetition, where the question is repeated in a narrower and focused way, is uttered with higher speech rhythm and more emphatic stresses.
Data emerging from the present work show the complex role of vocal non verbal behaviour in supporting the questioning process aimed to a secret disclosure. Suprasegmental dimension furthers cognitive comprehension by stressing some elements in the sentence (Cigada, 1989; Schegloff, 1968), so revealing the *focus* of the interviewer's question and supporting the interviewer communicative intention (Jakobi, Blanchet & Grossir-Le Nouvel, 1990). Moreover, vocal non verbal code is a powerful way to convey a variety of meanings connected with cooperation and empathy between the subject and the interviewer, so taking up both semantic and relational function (Anolli and Ciceri, 1995; 1995). Measurable aspects of verbal and non verbal behaviour, so, turn out to be central in determining the effectiveness of a specifically aimed interview.

References

1. Anolli L., Ciceri R. (a cura di) (1995), *Elementi di Psicologia della comunicazione*, Milano LED.
2. Cantlon, J., Payne, G., Erbaugh, C. (1996). Outcome-Based Practice: Disclosure Rates of Child Sexual Abuse Comparing Allegation Blind and Allegation Informed Structured Interviews. *Child Abuse and Neglect,* **20** (11), 1113-1120.

3. Cederborg, A.C., Orbach, Y., Sternberg, K.J, Lamb, M.E. (2000). Investigative Interviews of Child Witnesses in Sweden. *Child Abuse and Neglect*, **24** (10), 1355-1361.
4. Cigada, S. (1989). *Il linguaggio metafonologico*, La Scuola, Brescia.
5. Coulborn Faller, K. (1996). Interviewing children Who May Have been Abused: A Historical Perspective and Overview of Controversies. *Child Maltreatment,* **1** (2), 83-95.
6. Davey, R.I., Hill, J. (1999). The Variability of Practice in Interviews Used by Professionals to Investigate Child Sexual Abuse. *Child Abuse and Neglect,* **23** (6), 571-578.
7. Derlega, V.J., Metts, S., Petronio, S., Margulis, S.T. (1993). *Self-Disclosure.* Sage Publications, Newbury Park, California.
8. Jakobi, J.-M., Blanchet, A., Bromberg, M. (1988). Effects of prosodic focus in referential reiterations during interviews/Les effets du focus intonatif dans les réitérations référentielles en situation d'entretien de recherche., *Revue Internationale de Psychologie Sociale*. **1**(3-4), 345-363.
9. Jakobi, J.M., Blanchet, A., Grossir-le Nouvel, B., (1990). Four forms of propositional interrogation during the research interview/Quatre formes d'interrogation propositionnelle dans l'entretien de recherche. *Psychologie Francaise*. **35**(3), 207-215.
10. Schegloff, E. (1968). Sequencing in Conversational Openings. *American Anthropologist,* **70**, 1075-1095.

Audiovisual cues to finality

Pashiera Barkhuysen [1], Emiel Krahmer [1], Marc Swerts [1]
[1] *Faculty of Arts, Tilburg University, Tilburg, The Netherlands*

Abstract

This paper has two goals. First, to describe a paradigm to combine production of speech with perception. Second, to illustrate this paradigm with a case study. The question explored here is whether there exist auditory and visual cues, such as intonation and facial expressions, that speakers exploit to signal whether or not they want to continue their turn. Participants were able to predict the end of the utterance, and showed the fastest reactions in the audiovisual condition.

Keywords

Facial expressions, conversation, turn-taking, finality, intonation.

1 Introduction

1.1 Problem statement

The main research question is how one can measure *spontaneous facial expressions*, occuring in natural conditions. Different attempts have been made to develop human-observer-based coding systems, some of which are anatomically based (Cohn, Xiao, Moriyama, Ambadar, & Kanade, 2003). These methods, however, are often time-consuming. Often it is established in advance what to look for by using deliberate, sometimes posed visual expressions. Research suggests that deliberate visual expressions differ in a fundamental way from spontaneous ones, in that they are more intense, and are inferior and less symmetric in appearance, the latter due to the control by different motor pathways (Cohn et al., 2003).

1.2 Production-Perception Paradigm

Our goal, however, is to study visual expressions occuring in natural conditions, expressions which themselves can not be controlled. The conditions, on the other hand, can be controlled. We have developed a paradigm in which we evoke visual expressions in a semi-spontaneous, but controlled way, so that we know what the exact context of these expressions was (Barkhuysen, Krahmer, & Swerts, 2005; Krahmer & Swerts, in press; Swerts & Krahmer, in press). We take the role of the naive observer who does not know what to look for. By controlling the exact circumstances in which the expressions are evoked, we can at least estimate what the possible functions of the expressions are. In this way we might be able to compare function to form in a later stage. This first part of the process, designed to evoke visual expressions, will be described later in section 2.1 about the *production task*.

The next step in the design is the *perception task*. Here we extract samples of these expressions from their context, and present them in a different experimental set-up to naive observers who have to judge these samples. As they don't know what the context of these samples is, they can base their judgments solely on the cues available in the sample. The judges have to reconstruct the context in a controlled way. If the judgments match the context (i.e. the context can be reconstructed) there is reason to believe that the available cues carry information value. More about the *production-perception paradigm* can be found at the website: http://foap.uvt.nl.

It is useful to consider the distinction between measurement and judgment of stimuli. On the one hand it is possible to measure the actual facial behavior by EMG (electromyography), MRI-scans (magnetic resonance imaging) or 'objective', anatomically based coding systems (Ross, 1999; Wagner, 1997). However, the disadvantage of these methods is that they don't shed a light onto the information that is conveyed by the facial behavior. In order to address these questions judgment studies have been widely used (Scherer, 2003; Wagner, 1997). Within the field of judgment studies, different types can be distinguished such as the *forced-choice method*, i.e. to which response category (e.g. finality) does a certain stimulus (e.g. film fragment) belong? (Wagner, 1997)

1.3 Labeling and Cue Masking

Our paradigm can be regarded as an implementation of judgment studies, combined with controlled elicitations, on the basis of which we can work towards a more exact measurement of individual expressions. Once it is shown that there are features available in the chosen film fragments, and it is established in which samples (the ones that have been proven to be statistically relevant), we can start to search for individual features. This refinement can be achieved in various ways.

One way to do this is by using the inter-observer reliability (e.g. using the *kappa-statistic*) (Carletta, 1996; Fridlund, 1994; Scherer, 2003; Wagner, 1997). A number of observers look for specific features on the data. Whenever a feature is encountered it is marked on a binary or gradual scale. Their individual scores are compared statistically. If the correlation is high enough, the feature can be regarded as present. These features can then be compared with the judgment scores in the perception task (Scherer, 2003). The correlations will tell whether the present features serve the functions under investigation.

Another way to refine the judgment scores is to extract parts of the original samples which seem promising. For example, if there is possible cue value in the eyebrows, one can extract the upper part from the face from a sample and display this in a judgment experiment as well. This procedure can be regarded as a form of *cue masking* (Scherer, 2003). In this example, correlations between the *new* and *old judgment scores* will reveal whether the judgments are mainly based on the upper part of the face and thus possibly the eyebrows. In this way it is possible to refine the model.

1.4 Case Study

The next part of this article focuses upon a case study in which we apply the paradigm to audiovisual finality cues. During spoken interactions, conversation participants are able to adequately detect when their partner finishes a turn so that they can elegantly take over, without much overlap or delay. Previous research showed that speakers partly base such decisions on intonational cues (e.g. low- ending contours are reserved for turn-final position) (Swerts, Bouwhuis, & Collier, 1994). The question explored here is whether there also exist visual cues, such as facial expressions, that speakers exploit to signal whether or not they want to continue their turn.

2 Case Study: Finality

2.1 Production task

We collected utterances from 8 speakers, who were asked simple questions that elicited sequences of nouns (varying in length) during an interview in front of a camera. These questions were intended to evoke lists of words, such as questions requiring general knowledge, like "What are the colors of the Dutch flag", or questions eliciting a set of numbers, like "What are the odd numbers between five and fifteen?". The lists that were asked for varied in length, from 3 words to a sequence of 5 words.

2.2 Perception Task

We further implemented the design by taking a selection of the collected utterances in the production task, consisting of speakers' answers without the preceding question of the interviewer. We subsequently presented these to 30 participants during a perception experiment in one of the following modes: the original film containing audio and video (audio-visually or AV), only the visual part of the material (vision-only or VO) or only the auditory part while the visual channel only depicted a static black screen (audio-only or AO). The participants' task was to indicate as fast as possible when they felt a speaker's utterance had reached its end, by pressing a dedicated button at this exact moment. All subjects participated in all three conditions (and none of them had participated as a speaker in the data collection phase).

To get a baseline performance, the actual experiment was preceded by a test in which subjects had to respond to stimuli without finality cues. The aim of the baseline session was to find out how long it took subjects on average to respond to a *simple stimulus* presented in a certain modality and to control for inter-individual differences. The subjects had to press a button as soon as the end of the stimulus was reached. These stimuli consisted of (bimodal or unimodal) presentations of a video still (a single frame of one of the speakers) and/or monotonuous sounds (a stationary /m/ uttered by a male or female voice matching the sexe of the pictures), creating the impression of a speaker uttering a prolonged "mmm".

Figure 1. *The speakers SS and BB while uttering the first and middle word and just after the final word of a three word answer, such as "red, blue, white".*

3 Results

The results are displayed in Table 1. Participants were able to predict the end of the utterance, and showed the fastest reactions in the audiovisual condition. In the baseline session, however, reaction times were slower for the audiovisual stimuli.

Table 1. *Reaction time in milliseconds for the different conditions in the baseline session and the experiment.*

	Baseline	**Experiment**
AV	430,713	508,757
VO	343,817	688,216
AO	399,567	532,751

4 Discussion and conclusion

The findings of the baseline session suggest that cognitive load is higher when subjects need to focus on two modes at the same time. However, if we compare the baseline with experiment the results in the experiment are opposite, suggesting that our perceptual system might be facilitated when two different information sources (visual, auditory) are congruent in their cue value. This case study provided a good basis upon which we can exert a further, more detailed elaboration, which can be achieved via independent labelling or cue masking experiments, as is described in section 1.3. We believe that the *production - perception paradigm* can be easily implemented and provides clear and well interpretable results.

References

1. Barkhuysen, P. N.; Krahmer, E.; Swerts, M.G.J. (2005). Problem detection in human-machine interactions based on facial expressions of users. *Speech Communication*, **45**(3), 343-359.
2. Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, **22**(2), 249-254.
3. Cohn, J. F.; Xiao, J.; Moriyama, T.; Ambadar, Z.; Kanade, T. (2003). Automatic recognition of eye blinking in spontaneously occurring behavior. *Behavior Research Methods, Instruments, & Computers*, **35**(3), 420-428.
4. Fridlund, A. J. (1994). *Human facial expression : an evolutionary view.* San Diego, Calif. [etc.]: Academic Press.
5. Krahmer, E.; Swerts, M. G. J. (in press). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech.*
6. Ross, E. D. (1999). Affective prosody and the aprosodias. In M.-M. Mesulam (Ed.), *Principles of Behavioral and Cognitive Neurology*, pp. 316-331.
7. Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, **40**(1-2), 227-256.
8. Swerts, M.G.J.; Bouwhuis, D.G.; Collier, R. (1994). Melodic cues to the perceived finality of utterances. *Journal of the Acoustical Society of America*, **96**(4), 2064-2075.
9. Swerts, M.G.J.; Krahmer, E. (in press). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language.*
10. Wagner, H. L. (1997). Methods for the study of facial behavior. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression*, Cambridge, Cambridge University Press, pp. 31-54.

Child affective speech recording and its acoustic analysis in a cross-linguistic perspective

I. Grichkovtsova [1], I. Mennen [1]

[1] *Speech and Hearing Department, Queen Margaret University College, Edinburgh, United Kingdom*

Abstract

This paper presents the developed methodology for the cross-linguistic study of the production of child affective speech. Taking ethical considerations into account, only acting and light induction methods were chosen to record child affective speech. Children were playing emotions on one token utterance with the help of visual materials (16 repetitions for each emotion). English and French token utterances were selected with a similar sound and prosodic structure. This methodology enabled us to record a cross-linguistically comparable corpus of bilingual Scottish-French children and their monolingual peers. The results, based on acoustic measurements of pitch range variation, intonation patterns and timing of intonation patterns, compare the emotions across languages and across children.

Keywords

Affective speech, child recording, prosody.

1 Introduction

The importance of affective speech research has been acknowledged over the last years and received much attention. Nevertheless, this research is mainly concerned with adults, and studies of child affective speech are very rare and mostly deal with perception only [1,2]. The main reason for this is the methodological difficulties of affective speech recording, which are even more problematic regarding children, a particularly vulnerable subject group. After providing a short summary of recording methods, and existing studies of child affective productions, the developed methodology for this study will be described. The method design was also conditioned by the primary research objective to study child affective speech production in a cross-linguistic approach.

2 Affective speech recording

The present-day research uses spontaneous, acted and induced affective speech [3]. ***Spontaneous speech*** is considered to have the most authentic emotions, but it gives rise to various difficulties. This method requires the speech recording in completely natural conditions with people who do not suspect being recorded, which is ethically incorrect. Moreover, the researchers using this method encounter many problems with data analysis, like categorization, comparability, statistical validity. The method of ***induced affective speech*** gives good quality and better comparable recordings, but it raises even stronger ethical problems than spontaneous speech, as it is the researcher who puts the subject into a desirable affective state. As a result, the induced affective speech is usually very mild. ***Acting of emotions*** is the most widespread method in the affective speech research. It does not have ethical difficulties and it gives quality recordings. The nature of acting influences the affective realization: the subject aims at one emotion at a time with an intention to be understood without the additional help of context and semantic meaning. As a result, acted emotions are often criticized for their stereotypicity and exaggeration. It comes from the fact that not all vocal characteristics can be voluntarily controlled, and those that can be controlled, are often exaggerated to reinforce the intended expression. The level of control and individual preferences for different voice characteristics can vary across speakers, nevertheless, the imitation can be well recognized, if some crucial voice characteristics are successfully imitated [4].

3 Developed methodology

The starting point for this study was to develop an appropriate methodology to record child affective speech. Taking into account ethical and analysis questions, it was decided to use acting with an induction element. Acting has already been effectively used in the study of child facial expressions of emotions [5]. The induction element was introduced with the help of visual materials, which were developed in a realistic manner together with a professional illustrator and based on the research of facial expressions. They represent a child expressing four emotions (happiness, sadness, anger and fear); the child's gender and age are adapted to that of the recorded children. The randomized cards have 16 repetitions for each emotion. One token utterance was selected with a similar sound and prosodic structure for English and French: "I see a banana there./Je vois une banane, moi." The subject is instructed to say the token utterance in the same way as the child at the picture, thus the child is playing emotions through association with the drawing. Visual materials played a very important part in the recordings: they served as the reference of the expressed emotions, and an affect inducing material, as the subjects associate themselves with the picture.

Neutral was not represented as a facial expression card, neutral utterances were recorded in a separate test at the very beginning of the recording session. In this test, the child was instructed to say the token utterance (I see * there./Je vois *, moi.) with a word from a pile of cards (picture naming); 16 repetitions of the original word (banana/banane) were mixed among the others. Later all the utterances with the original token wording were selected for analysis. A small number of utterances was rejected due to noise, hesitation or incorrect wording. 8 Scottish English-French bilingual children and 16 monolingual peers, aged between 7 and 10, were recorded in total. Bilingual children were recorded in two sessions (one for each language) by different experimenters in order to control for language mode and code-switching; there was a period of about two weeks between the sessions. This methodology allowed us to record a cross-linguistically comparable corpus of bilingual and monolingual children.

4 Affective speech analysis in a cross-linguistic perspective

Recent cross-linguistic studies [6] show that there is a number of available means to express emotions in speech (pitch range, rhythm, intonation patterns, voice quality, etc), but their usage, level of importance and meaning vary in different languages. They suggest that the vocal realization of emotions is based both on universal mechanisms and social/cultural conventions. The research of affective bilingual speech is just starting. Even though bilinguals may differ in their production of affective speech from the monolinguals in each of their languages (a phenomenon, which is widely attested in bilingual studies of other phonetic aspects [7]), the research into the affective speech of simultaneous bilinguals can help to solve the problem of speaker variability and to identify important cross-linguistic tendencies.

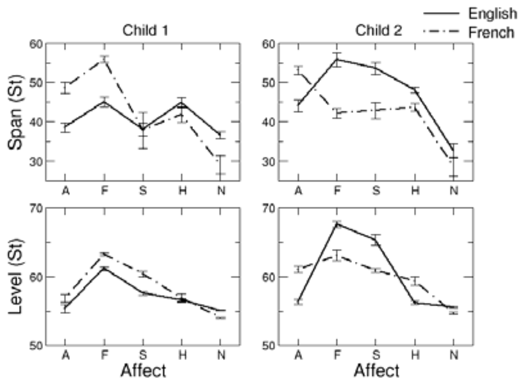


Figure 1. *Pitch range in semitones for two bilingual children with standard errors. Abbreviations for Affect: A (Anger), F (Fear), S (Sadness), H (Happiness), N (Neutral).*

4.1 Acoustic analysis

Using *wavesurfer* software, acoustic measurements of pitch range variation and peak alignment (timing of intonation patterns) were taken. These measurements were shown important for cross-linguistic and second language acquisition studies [8]. Following Ladd [9] we describe pitch range using two partially independent measures of variation: overall level and span. Overall level refers to the "height" of a speaker's voice, span refers to the width of pitch frequencies covered by a speaker (how big the excursions are). These measures have been shown [10] to give better and perceptually relevant results, than the more commonly used measures, such as statistical moments (mean, standard deviation, difference between maximum and minimum F0, etc.). Peak alignment is a ratio of the distance of the main peak from the onset of the main accented vowel as a function of the duration of that vowel.

All the utterances were manually labeled at the following pitch points: IF0 (initial value of F0), H1 (the peak of the first accented word), V (valley, the lowest point between two peaks), H2 (the peak of the main accented word - main accent), FL (final low - the lowest value after the peak). In cases where there was no well defined peak, an absolute F0 max value was taken near the accented syllable. Segment duration measurements were taken at the following points: the first consonant, the first vowel and the second consonant of the main accent. F0 values were then extracted automatically to data files with the help of ESPS algorithm, and again checked manually for any pitch perturbation or voice quality errors. In our measurements, FL point was taken as the overall level, and the difference of H2 and FL, as the span. The measures are expressed in semitones.

4.2 Results

The preliminary findings (2 bilingual children) show that children realize differences between emotions and between languages, and this is apparent in each of the measurements taken in this study, although it is most obvious in pitch range (Fig.1). However, there are differences in the way the emotions are realized across children and languages.

5 Conclusion

This study shows that our method for collecting a cross-linguistically comparable corpus of bilingual Scottish-French children and monolingual English and French children was appropriate. All children were able to act out emotions in both their languages. Our study also confirms that the chosen measures of pitch range and peak alignment are effective in capturing cross-linguistic differences in affective speech. The preliminary results show that bilingual children realize cross-linguistic differences, and sometimes realize emotions differently in their two languages.

References:

1. Van der Meulen S., Janssen P. and Den Os E. (1997). Prosodic abilities in children with specific language impairment. *Journal of Communication Disorders*, **30**, 155-170.
2. Clément J. (1999). Structure des représentations prosodiques. Développement normal et pathologique du traitement de la prosodie. *PhD thesis*, Université de Paris 5.
3. Campbell N. (2000). Databases of Emotional Speech. *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, Ireland.
4. Zetterholm E. (2003) Voice Imitation: A Phonetic Study of Perceptual Illusions and Acoustic Success. *Travaux de l'Institut de Linguistique de Lundt*, **44**.
5. Field, T. M.; Walden T. A. (1983). Production and discimination of facial expressions by preschool children. *Child Development*, **53**, 1299-1300.
6. Scherer K. R.; Banse R. and Wallbott H. G. (2001) Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, **32(1).**
7. K. Aoyama; J. E. Flege, S. G. Guion; R. Akahane-Yamada, and T. Yamada (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, **32**, 333–250.
8. Mennen I. (2004).Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, **32**, 543-563.
9. Ladd D. R. (1996). Intonational phonology. Cambridge University Press.
10. Patterson D. (2000). A linguistic approach to pitch range modelling. *PhD thesis*, University of Edinburgh.

Terminal radar approach control: measures of voice communication system performance

O.V. Prinzo
FAA Civil Aerospace Medical Institute, Oklahoma City, Oklahoma USA

Abstract

When technological advances lead to innovations in communications system development, these emerging systems may be evaluated against the existing system's performance parameters. This paper describes the method used to extract baseline analog voice communications system performance data using waveform and spectral analysis. The data include duration parameters and the prevalence of blocked, stepped-on, and clipped transmissions. The findings presented here are but a first step in providing objective and quantifiable communications system performance metrics that may prove valuable to communication systems developers and personnel charged with evaluating, certifying, and deploying the next generation of communications systems.

Keywords

PTT measures, Voice communications

Introduction

Communications in the National Airspace System (NAS) are an essential safety component to successful air travel. As the NAS migrates from its current ground infrastructure and voice communications system to one that encompasses both ground and airborne systems, digital data transmissions may become the principal communication medium [1].

A review of the voice communication literature revealed that existing results do not provide current information concerning the frequency of occurrence or the severity of stepped-on or blocked transmissions. This information is critical for an analysis of the operational and safety benefits that justify investing in the development of new communications systems. Unfortunately existing data, collected 8-10 years ago, may not constitute a valid basis for comparison with, or extrapolation to, the operating environment in 2010 [2].

Consequently, data on the current operational communications system were needed to establish a baseline against which the future communications system's performance can be compared. Therefore, voice tapes of operational air traffic control voice communications were obtained, transcribed, and analyzed. The results provide objective and quantifiable communications system performance metrics that may prove valuable to communications systems developers and personnel charged with the evaluation, certification, and deployment of the next generation of communications systems.

Materials

Audio Tapes. Five of America's busiest Terminal Radar Approach Control (TRACON) facilities provided approach and departure control voice communications on digital audiotapes (DAT) using the NiceLogger™ Digital Voice Recorder System (DVRS). The NiceLogger™ Digital Voice Reproducer System (DVRS) was used to copy DAT onto audiocassettes. Approximately 10 hr of busy communications were analyzed — 1 hr of approach and 1 hr of departure transmissions for each facility.

Audio Software. Adobe Audition™ (1.5), an audio editing software tool, was used to record, convert, and save the audio input as digital audio files for subsequent analysis. It was used to extract voice and Push-to-Talk (PTT) onset and offset times.

Extraction of Voice Data Analysis Points

Before transferring the audiocassettes recordings to the computer, the PC soundcard was adjusted using Record Control (In Line Volume) to record the bulk of data between ± 75% of scale, with no peaks outside ± 80%. This established the maximum dynamic range to retain small signal details.

The voice communications system performance duration measures were computed from the data points identified by listening to and inspecting individual waveforms. Their definitions and the process of identification are expounded upon next.

Push-to-Talk Onset was the first transition from static signaling that the microphone had been keyed. As shown in Figure 1, PTT onset occurred 2 min 13 s 226 ms into the tape. The dashed vertical line identifies the point of PTT onset.

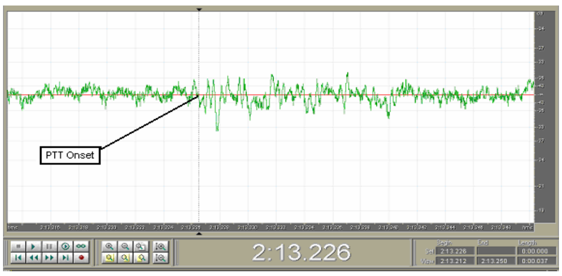


Figure 1. *Magnification of PTT onset*

Voice Onset was defined as the start of the first word in a transmission. By listening to and advancing the transmission along the timeline from PTT onset until the first word was affected aided in determining the point of voice onset. Figure 2 shows that the audio signal began to resemble human speech at 2 min 13 s 357 ms, and that point was recorded as voice onset.

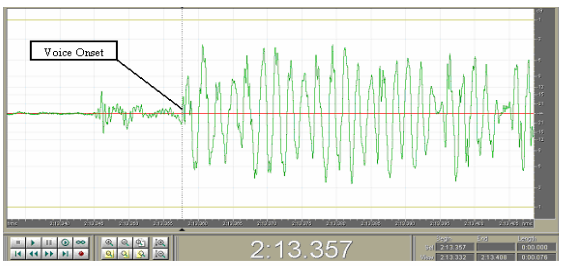


Figure 2. *Magnification of voice onset*

Voice Offset was defined as the end of the last transmitted word. Once again, by listening to and advancing the transmission along the timeline until the final word was affected aided in determining the point of voice offset. Figure 3 shows an example of the earliest point that the last word was unaffected. Continuation of the signal beyond this point was due to electronic effects from the transmitter, as evidenced by the spectral view.

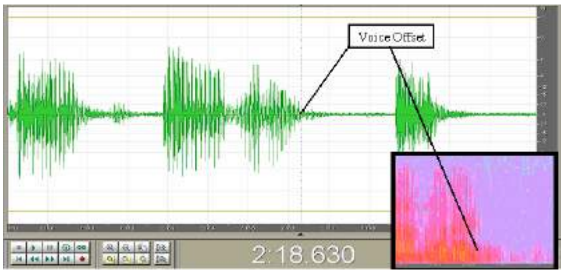


Figure 3. *Waveform and spectral view of voice offset*

Push-to-Talk Release was defined as the start of the pop back that occurred after release of the PTT button. For most ATC transmissions, like the one presented in Figure 4, it was very noticeable at 2 min 18 s 867 ms.

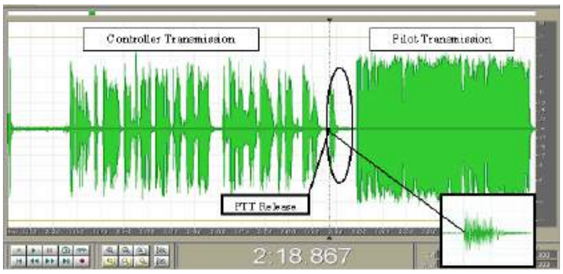


Figure 4. *PTT release with magnification*

Push-to-Talk Settle was the first return to a low signal condition following release of the PTT button. The low signal condition starts when the transmission decayed to < 90% of its peak value. As seen in Figure 5, it occurred at 2 min 18 s 990 ms.

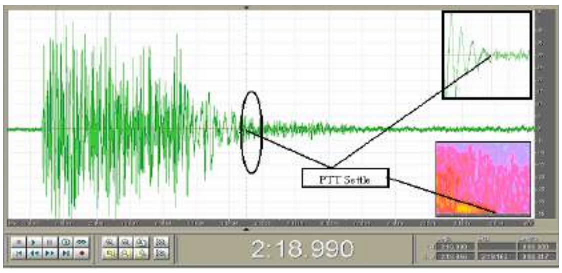


Figure 5. *PTT-settle with magnification*

Voice Communications Systems Performance

The voice communications data points, like the ones presented in Table 1, were used to compute the duration measures of voice communications system performance. Arithmetic duration (time-stamp difference) was used to derive their values.

Table 1. *Examples of Voice Communications Data Points*

		TIME STAMP (second. millisecond)				
T #	**ID**	**PTT Onset**	**Voice Onset**	**Voice Offset**	**PTT Release**	**PTT Settle**
1	ATC	00.937	00.949	03.185	03.458	03.585
2	FD	03.854	03.900	05.370	05.380	05.417

Setup delay was the momentary pause preceding voice onset. It was the difference between the point of voice onset and PTT onset. For T1, the silence preceding voice onset was 12 ms (00.949 - 00.937). *Voice-streaming time* was the difference between voice offset and voice onset. Again using T1, it was 2 s 236 ms (03.185 – 00.949). *Pause duration* was the difference between PTT release and voice offset. For T1, 273 ms of silence preceded the release of the PTT button (03.458 – 03.185). *Message propagation* represented the minimum duration for the switching mechanism to return to a resting state following PTT release. During this interval the opportunity for blocking can occur. For T1, it was the difference between PTT settle and PTT release (03.585 – 03.458 = 127 ms). *Frequency occupancy time* was the difference between PTT release and PTT onset. It represented how long the radio frequency was in use per transmission. For T1, it was 2s 521 ms (03.458 –00.937). *Lag time* was the duration of silence between consecutive transmissions. For example, lag time between PTT onset for T2 and PTT release for T1 was computed as 03.854 – 03.458 = 396 ms. *Number of transmissions per aircraft* was tallied as the number of pilot and controller messages exchanged to and from a particular aircraft.

Disruptions to Efficient Information Transfer

Particular attention was paid to the detection of any disruption in information transfer such as stepped-on, blocked, or clipped transmissions. Unintelligible messages, as well as repeated transmissions due to a lack of response or a request by the receiver (e.g., say again, who was that calling in), were also tallied. In most cases, blocked transmissions were easily identified because they were announced by the spoken word "blocked." At other times, the controller told the pilot that the transmission was blocked or stepped-on.

Results and Discussion

Nearly 8,000 transmissions were analyzed that represented the busiest air-ground communications from the five terminal radar approach control facilities with the highest number of operations in the contiguous United States. The EXCEL™ 2000 spreadsheets were imported into *SPSS*© *11.5 for Windows* for statistical analysis.

Typically, setup delays lasted 81 ms, voice streaming 2568 ms, pause duration 127 ms, and message propagation 73 ms for a total of 2849 ms per transmission. On average, transmissions were separated by 1736 ms of silence.

Disruptions to efficient information transfer can result from blocked, stepped-on, and clipped transmissions — but they are rare events and occurred in only 1.16% of the sampled transmissions. A comparison between aircraft with and without disruptions revealed that when a disruption was present, an average of 14.54 messages were transmitted compared with an average of 9.90 messages when no disruption was present. Even so, there appears to be some type of a detection mechanism in place to alert the controller to the presence of blocked transmissions. The source is of this detection system is unclear; however, systems developers may want to exploit and expand this capability to include stepped-on and clipped transmissions.

References:

1. FAA. (2005). *NAS Capital Investment Plan.* Office of System Architecture and Investment Analysis. http://nasdocs.faa.gov/nasihtml/cip/CIP_2005.pdf. Accessed 5 May 2005.
2. FAA. (July 2004). *FAA Long Range Aerospace Forecasts Fiscal Years 2020, 2025 and 2030.* http://apo.faa.gov/Forecast%20LongRange%2004/long04.pdf. Accessed 5 May 2005.

Sound analysis of chinchillas living in sociable groups

J. Bartl[1], M. Schneider[1], H.-U. Kleindienst[2] and M. Erhard[1]
[1] *Institute of Animal Welfare, Ethology, and Animal Hygiene, Faculty of Veterinary Medicine Ludwig -Maximilians – University, Munich, Germany*
[2] *Max - Planck - Institute for Ornithology, Seewiesen, Germany*

Abstract

The aim of this study was to get a survey of vocal behavior of chinchillas *(Chinchilla lanigera*) living in sociable groups. Therefore vocalizations of 15 chinchillas, housed in three-dimensional designed cages in 2 sociable groups of 3 to 7 individuals, were recorded on magnetic tape with a 4 DS Racal Recorder with frequency response of 100 Hz to 19 kHz. All vocalizations were digitalized with the Avisoft SasLabPro® software program (16 bits, sampling frequency 22050Hz) and spectro- and sonagrams were made. Syllables were sonographically analyzed and classified according to parameters like frequency, intensity, duration and number of harmonics. After that the arrangement of the syllables to entire calls was examined. While chinchillas were active their behavior was directly observed.

Keywords

Chinchillas, vocalizations, sonagrams, Avisoft SasLabPro®, sound analysis.

1 Introduction

Chinchillas are twilight- and night-active, social rodents living in familial groups [3]. The male progeny is driven away after reaching sexual maturity, whereas the females mostly remain with the group. In mixed groups leadership is often assumed by a male animal [5]. It is well known that chinchillas have the best sensitivity of their hearing range equal to that of humans at 125 Hz-16 kHz [2] and have many different kinds of vocalizations at their disposal [1, 5]. The aim of this study was to get a survey of vocal behavior of chinchillas living in sociable groups.

2 Materials and methods

Animals and their accommodation

The vocalizations of chinchillas living in family groups were recorded during one year. The group of chinchillas studied consisted of 15 animals, 5 of them male and 10 female, among them 8 young animals born during the study period. These chinchillas were kept in mixed-sex groups in 2 neighboring cages (width 1.70 m, depth 3.05 m, and height 3.50 m). The cages had a three-dimensional design. Each cage had 5 sitting boards on different levels, with one sleeping house each. Connection to the floor was provided by means of branches for climbing. The cages were separated by wire mesh, so that between the two groups hearing, sight, smell and – to a limited extent – physical contact was possible.

Sound recordings

Instruments

Sound recordings were made using 3 Elektret microphone capsules (by Conrad) with an impedance of 600 Ohm, a frequency range between 20 Hz and 18 KHz, and a sensitivity of –70 dB. The microphone capsules were built into steel tubes to protect them from bites. A universal pre-amplifier (by Conrad) for a frequency range of 20 Hz to 20 kHz was used, with an amplification factor of 40 and a maximum input voltage of 100 mV. The sounds were recorded using a store 4 DS racal recorder on three direct channels with a tape speed of 3¾ in/s, which corresponds to a frequency bandwidth of 100 Hz to 19 kHz. The recording level was 0.1 Volt. Magnetic tapes of 6.25 mm 1800-3600 feet in length x 0.5-1 mm of the companies Maxell (25-120), Scotch (3M 177) and Realistic (laboratory standard) were used.

Software

The sounds recorded on magnetic tape were digitalized using the software Avisoft Sonograph ProSasLabPro® Version 2.0. For that purpose, the soundcard was set at a sampling rate of 22,050 Hz and 16 bits. Initially, the digitalized sounds appeared in the main window as an amplitude against time (oscillogram), with a small sonagram below. For additional analyses an extra sonagram had to be created in another window. In this study, the following spectrogram parameters were set for that purpose: FFT length 256, overlap 50 %, frame 100 %, window Hamming. The digitalized data were then used to create sonagrams in the form of frequency–time diagrams.

Data collection

The sonagrams were initially used to characterize individual sounds according to the following parameters: duration of the individual sound in seconds, frequencies, relative intensities, number of harmonics, frequency with the highest amplitude, and characteristics of the appearance of the banding pattern. Then the combination of the individual sounds to form sound sequences was investigated, and the interval between the individual sounds in seconds was measured. Duration of the sound, intervals, frequencies and intensities were measured in the sonagram using the cursor of the software. The frequencies were determined using the cursor in steps with a difference of 86 Hz, and 87 Hz for every 7th step. The times were measured in steps of 6 ms and 5 ms, respectively, using the cursor. The relative intensities were determined steplessly by setting the cursor using the mouse. At the peak in the oscillogram additional logarithmic power spectrums were created (under "analysis – one-dimensional transformation – power spectrum" (logarithmic)) in order to obtain a more precise picture of the distribution of intensities onto certain frequency ranges, with the intensity being represented as the abscissa against the frequency on the ordinate. Here, the frequency was rendered steplessly, and up to 1 kHz with an accuracy of 3 decimal places. At the frequency spectrums of individual sections the maximum intensity and its associated frequency, the mean frequency and various numbers of intensity peaks as well as the respective frequency ranges were recorded using the "spectral characteristics" software function.

Observation of behavior and recordings

The fact that the chinchillas were kept in groups made it impossible to attribute each sound recorded to the animal that made it. For that purpose, the sound recordings were accompanied by simultaneous direct observation of behavior by means of behavior sampling using continuous recording [4], sometimes with a handheld video camera (Sony DCR-TRV 828E). This also served the purpose of documenting the respective behavior of the chinchilla group prior to, during and after making the sound, and attributing it to a specific behavioral pattern. Given the main activity phases of chinchillas filming was done almost exclusively at night, which necessitated the use of an infrared lamp along with the digital handheld camera.

3 Results

All sounds recorded were within the set frequency range of the software. The repertoire of sounds made by the chinchillas contained both tonal and atonal sounds. All sounds measured were within a frequency range of 0 Hz to 11 kHz. The occurrence of ultrasounds has not been investigated. Sounds which were made by the chinchillas during close contact and were correspondingly soft could not be analyzed as the microphones were too far away due to the fact that the animals were kept in groups in the cage. In terms of the acoustic criteria 10 different types of sound were identified. They were attributable to the behavioral contexts of social contact, sexual behavior, defensive and offensive aggression, exploratory behavior and predator avoidance. Two sounds were exclusively made by young animals. One contact sound directed at adult chinchillas was recorded until the 10th week of life of the young animals. The other sound was made by the young animals while sucking their mother and disappeared after the animals were weaned at the age of 4-6 weeks.

4 Discussion

Many sounds made by the chinchilla group under investigation were recorded in very good quality, some in good quality. However, the repertoire of sounds comprises other sounds as well, which could not be recorded at an adequate loudness level using this method. Sounds that are too soft do not stand out against background noise and therefore escaped analysis. For instance, sounds of actually good quality were blotted out by background noises made by the chinchillas jumping onto boards or against the wire mesh of the cage, eating food pellets, gnawing on sitting boards and climbing branches, or drinking from the water bottle. Further amplifying the microphones also increased the loudness of the background noises and therefore did not prove useful. Keeping the animals inside a cage created another problem in that the microphones used for recording the sounds made during close contact were not close enough to the animal, so that the sounds were not recorded in a quality good enough for analysis. These problems reduced the probability of obtaining sounds in a quality sufficient for digitalization. This necessitated an increased number of recordings and, accordingly, longer study periods. In order to increase the probability of obtaining usable sound recordings 3 microphones were employed simultaneously on 3 channels of the tape recorder. The 3 microphones were suspended at places popular with the animals, where social behavior was accordingly likely to occur. One microphone was suspended at the feeding place (pellets, hay and water), another at the place where a climbing branch provided a way up from the floor to the lower sitting boards, and the third close to the most popular sleeping house and the border to the neighboring cage.

Another problem is the chinchillas´ strong urge to gnaw on things. Cables as well as microphones had to be protected from the chinchillas. The cables had to run outside the reach of the chinchillas. The microphone capsules were additionally placed inside short steel tubes, their bottoms aligned so as to avoid formation of an echo.

5 Conclusion

The method used was suitable in order to gain an overview of the sound repertoire of chinchillas living in sociable groups. In order to be able to record and analyze soft sounds made by chinchillas during close contact other measuring methods will have to be employed. One possibility would be to install the protected microphones and cameras directly inside the sleeping houses as the experience of our project shows that there, many interactions between the chinchillas take place. The alternative of attaching microphones directly to the animals is not feasible as chinchillas are very clever in the use of their hand-like front paws. In addition, any manipulation means stress for the animals. Another measuring method would be to place one or more animals inside a small cage equipped with microphones. Analysis of the sounds made by chinchillas during everyday life within a sociable group, however, is not useful in such cages.

References

1. Bignami, G.; Beach, F. A. (1968). Mating behaviour in the chinchilla. *Animal Behaviour*, **16**, 45-53.
2. Heffner, R.; Heffner, H. E. (1991). Behavioral hearing range of the chinchilla. *Hearing Research*, **52(1)**, 13-16.
3. Hoefer, H. (1994). Chinchillas*. Veterinary Clinics Of North America: Small Animal Practice* **24(1)**, 103-111.
4. Martin, P.; Bateson, P. (1986). Measuring behaviour. *Cambridge University Press*, Cambridge, London, New York, New Rochelle, Melbourne, Sydney. ISBN 32368 1.
5. Schweigart, G. (1995). Chinchilla. *Heimtier und Patient*. Gustav Fischer Verlag, Jena, Stuttgart. ISBN 3-334-60957-X.

Novel equipment for monitoring Ultrasonic Vocalizations

H.J. van der Wiel [1] , C. van der Leden [1]
[1] *TNO-Prins Maurits Laboratory, Rijswijk, The Netherlands*

Abstract
Automated equipment was developed that monitors and analyses ultrasonic vocalizations (USV) in laboratory animals making use of a microphone, a Personal Computer and LabVIEW™. The measurement has potential for: early recognizing of animal discomfort, as a parameter in behavioral studies and as a monitoring tool of animal well-being during transport, acclimatization and housing conditions

Keywords
Ultrasonic, well-being, pain-relief, communication

1 Introduction
Ultrasonic vocalizations (USV) are sounds within the frequency of 20-100 kHz, a frequency range which is above the human range of hearing. However, bats, dolphins, rodents and insects produce and hear USV as part of their behavior. Evidence indicates that USV could be used as a sensitive indicator of subtle emotional and motivational changes in animals. For this reason, USV produced by animals in various conditions could be explored as a non-invasive method for registering the welfare status of an animal. Modern equipment is able to transforms the USV sounds in a for humans audible signal which is ready for analysis.

2 Objectives of USV measurement
The use of USV to measure the welfare status of an animal would be advantageous in a number of situations:
-on the basis of USV at an early stage of an experiment pain relief can be administered;
-the effectiveness of the pain relief can be measured (monitored), and if necessary adjusted
-as a parameter to determine the extent of discomfort or wellbeing;
-as a parameter in behavioral studies.
-as a monitoring system during transport and acclimatization
-as a monitoring system during normal housing conditions

3 USV Measurement
Modern equipment transforms the USV sounds in a visible and for humans audible signal ready for analysis. The very time-consuming and labor-intensive method with a bat detector combined with a tape-recorder is no longer an option for a versatile measurement. TNO-PML developed equipment that monitors USV sounds making use of a microphone, a Personal Computer and LabVIEW™, see Figure 1. Monitored USV sounds are directly analyzed, see Figure 2 which present the calls of a marmoset monkey. Ultrasonic sounds in the environment can have a negative effect on animals resulting in stress and deterioration of their wellbeing. In the last few years the importance of ultrasounds on animals has been recognized widely and even an article of law has been made to alert researchers to ultrasounds with respect to animal housing.

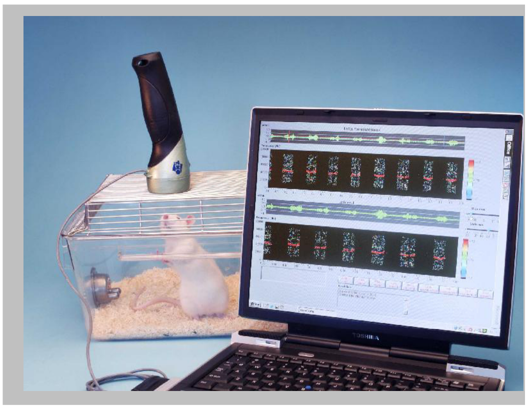


Figure 1. *Equipment to monitor USV sounds.*

4 What can be achieved
By monitoring the USV sound one can learn much about the well-being of animals. The animal research is refined and extended by an objective analysis of USV sounds.
By registering the USV of an animal species the occurrence of "positive" or "negative" sounds under various conditions. USV measurement could lead to:
- Less discomfort, improvement/monitoring of animal welfare, and better living conditions for laboratory animals.
- A decrease in the need of laboratory animals because at an earlier stage of an experiment effective countermeasures could be taken.

References:

1. Barfield, R.J.; Greyer, L.A. (1975). The ultra-sonic post-ejaculatory song of the male rat. *Science*, **176**, 1340-1350.
2. Blanchard, R.J.; Blanchard, D.C.; Takamishi, T.; Kelly, M. (1977). Attack and defensive behaviour in the albino rat. *Animal behaviour*, **25**, 622-634.
3. Blanchard, R.J.; Blanchard, D.C.; Agullana, R.; Weiss, S.M. (1991). 22kHz alarm cries to presentation of a predator, by laboratory rats living invisible burrow systems. *Physiol. Behav.*, **50,** 967-72.
4. Cuomo, V.; Salvia, de M.A.; Lacomba, C.; Tattoli, M.; Cagiano, R. (1990). A new experimental approach for detecting emotional and motivational changes produced by neuroactive compounds in rodents. *Ann. Ist. Super. Sanita*, **26**, 895.
5. Diamond, I.T. (1967). The sensory neocortex. *In: Contributions to sensory physiology. W.D. Neff (Ed.). Academic Press, New York. Vol 2.*
6. Knutson, B.; Burgdorf, J.; Panksepp, J., (1998). Anticipation of play elicits high-frequency ultra-sonic vocalizations in young rats. *J. Comp. Psychol.*, **112**, 65-73.
7. Noirot, E.; (1968). Ultrasounds in young rodents. II. Changes with age in albino rats. *Animal Behaviour*, **16**, 129-134.
8. Sales, G.D.; Pye, J.D. (1974). Ultrasonic communication by animals. *Chapman and Hall, London.*

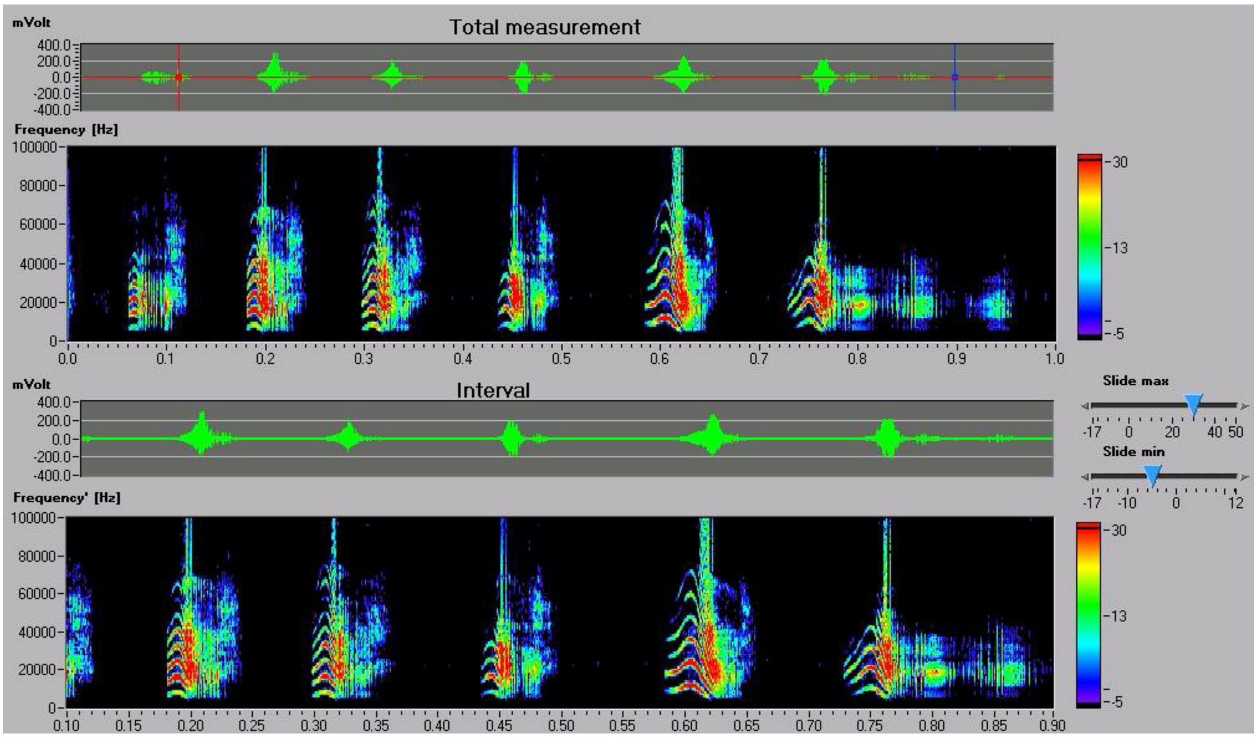


Figure 2. Screen display of the calls from the marmoset : Example of harmonic.

How to test for a ritual: the case of whispering

J. Cirillo, D. Todt
Institute of Biology, Free University, D-12163 Berlin, Germany

Abstract

This study examined a hypothesis explaining whispered (unvoiced) speech as a ritualized form of normal (phonated) speech. The methodological framework combined criteria of ritualization with the theory of honest signaling which predicts that rituals should be more costly than their originals. By comparing the number of syllables which adult subjects were able to produce by the air of one single breath, we could show that whispering is more arduous than normal speaking, and thus also more 'costly'. An additional inquiry into the ontogeny of speaking revealed that whispering requires a prior training and competence in using a phonated voice. Spontaneous interactions by whispering were observed only in children of about four years of age; i.e. when children develop their own concept of secrecy and privacy. These results confirm the cited hypothesis and additionally suggest that the applied procedure makes an expedient methodological tool of testing signals for specific ritual properties.

Keywords

Whispering, ingroup-outgroup effects, ritualization, honest signalling, speech development.

1 Introduction

Julian Huxley [1] defined ritualization as a development of signal patterns that are derived from other signals or homologous action patterns, but concurrently lose their original function and, instead, obtain a novel and often symbolic communicative role. This process of signal generation is linked with distinct changes in signal structure and performance, which include, for example, a reduction of pattern elements, an amplification of the remaining element(s), a conspicuous exhibition of the new structure and, above all, a signal performance that stresses the contrast between the novel and the aboriginal signal [2]. It is also widely accepted that ritualization should lead to honest signals only, and that such honesty requires that these signals are costly [3].

A well-known paradigm of ritualized signalling is human laughter. It evolved from the play signals of nonhuman primates and, after being shaped by ritualization, is used now as a culturally universal display [4]. Alike laughter, also whispered vocalizations are wide-spread across human cultures, and there is some evidence that they can be explained as universal rituals as well [5]. Unlike laughing, however, whispering is an exclusively human accomplishment that cannot be traced back to homologous precursors of nonhuman primates. Therefore it was suggested that whispering could have evolved only as a ritualized form of normal speech [6]. It was clear that any study designed to examine such suggestion could not lay on classical methods, but had to develop and apply a different methodological procedure instead. In this article we report an approach that basically used the following two procedures:

First, we investigated the costs of whispering, which – according to the theory of honest signalling – should exceed the costs of normal speaking [3]. Second, we studied the early ontogeny of speaking abilities and especially tested for developmental precursors of whispered vocalizations.

2 Material and Methods

2.1 Production

Participants (n=8) were students of biology with neither hearing nor articulation deficits. Subjects had to deal with six different tasks; i.e. to speak a given text as loudly and as softly as possible, and to vocalize it at an intermediate level as well. The particular challenge of this experiment was that subjects had to use one single breath to articulate as many syllables as they could and also with both a phonated and a whispered voice. To standardize the design, subjects had to speak successions of different numbers. Subjects were seated in a sound-protected test chamber with walls shielded against echo-effects. To avoid serial effects, the order of experiments and also the order of tasks within experiments were changed across subjects. Amplitude measures were taken at a speaker-microphone distance of 1m by using a Rode & Schwarz EZ GA2 precision sound level meter. To test for statistical significance of results, we applied Friedman ANOVA, χ^2-methods. Significance was accepted at a level of $p < 0.05$.

2.2 Ontogeny

Material used for the study was collected from children (n=30) between 1 to 5 years of age. Participants were contacted in a familiar environment; i.e. either in their 'kindergarten' with well-known peers and a caretaker present, or when being at home with their mother. To prepare the tests, the experimenter visited a given 'kindergarten' or family several times and engaged in playful encounters with the children. Tests began only after such familiarity seemed to be achieved. In the tests, children were invited to utter whispered vocalization. Such invitations took place during playful encounters, or by addressing them with a whispering voice. Here, the caretaker, or the mother, or J.C. acted as an initiator. Additional recordings were made when two or more children were engaged in playful interactions themselves. All vocal utterances were sampled by a Sony TC D5 recorder equipped with two Sennheiser Me80 microphones placed in a hidden position. Additionally, further data were sampled by interviewing caretakers or parents about (a) when a given child had performed his/her first whisper, (b) in which contexts the child used this vocalization, and especially (c) which attitude they themselves had about whispering. Analyses of these data followed classical procedures.

3 Results

3.1 Production

The topmost amount of syllables spoken per breath was around 160. Individuals who reached this amount also by whispering did so only by using an extremely soft voice. The number of syllables whispered loudly was smaller than those of syllables spoken otherwise. Distributions of syllable numbers measured for each performance category were similar for normal speech [Friedman ANOVA, $\chi^2 = 5.68$, N = 8, n.s.] but extremely different for whispered vocalisations [Friedman ANOVA, $\chi^2 = 13.56$, N = 8, $p < 0.001$]. Finally, the inter-individual variation of amplitude data ascertained within a given performance category was remarkably small for whispered vocalisations, but larger for that of data belonging to the measures of normal speech.

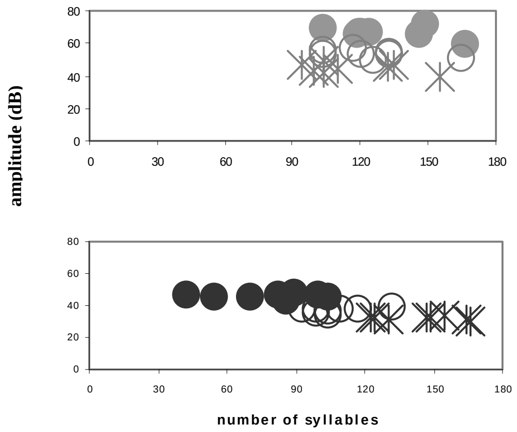


Figure 1. Amounts of syllables that subjects were able to produce per breath, here plotted against their acoustical amplitude. Top: phonated speech. Bottom: whispered speech. Filled circles: vocalizations uttered as loudly as possible. Circles: vocalizations uttered at an intermediate amplitude level. Crosses: vocalizations uttered as softly as possible.

3.2 Ontogeny

At an age of about 1 year, children (n=6) paid attention to the whispering of an adult, but did neither respond otherwise nor whisper themselves. At an age of about 2 years (n=12), six children replied to a whispered address by whispering themselves. Such responses did not persist, and the 'vocal adjustments' ended soon. At an age of about 4 to 5 years, all children (n=12) readily entered into a whispered dialogue, after having been addressed by either a familiar adult or peer beforehand. However, form and duration of their interactive engagement differed remarkably across children. Both, observations and interviews with caretakers indicated that at an age of 4 or 5 years children also began to use whispering in two ways: (a) as a kind of honouring the addressed person as a confident partner, (b) as a kind of punishment of a peer (or an adult) by actively excluding him/her from a whispered conversation. Finally, none of the children mixed whispered words with words uttered with a normal voice, and such performance mode suggested that the two kinds of verbal expression are two different signal systems.

4 Discussion

Our study provided strong support for the view that whispering is a ritualized form of normal speech. Such can be concluded from our results showing that a production of unvoiced speech requires more effort and thus more costs than speaking with a normal phonated voice, and also from our inquiry into the development of speaking. As whispered words were never mixed with words spoken by a normal voice, we conclude that whispering is 'special system' of human communication.

4.1 Ontogeny of whispering

Our study showed that first whispered vocalizations may occur only when children are around two years of age and already able to apply various phonated words and simple sentences. An interactive use of whispering emerged, however, only at a time when children are considered to have developed their own concept of secrecy and privacy [7]. Thus, children showed a performance that is typical also for adult speakers, then.

4.2 Ritualization and cost of signalling

Aside from evolutionary aspects of whispering, that have not been tested yet, there are several characteristics which distinguish this vocalization from normal speech and thereby support the view that the latter can be interpreted as a ritualized class of the former [1, 2]. First, there are clear structural differences between the two vocal patterns. For instance, whispering does not built on a fundamental frequency but just a turbulent noise, thus whispered words can clearly be distinguished from phonated ones [8]. Second, whispering is not a common substitute of normal speech. Because of a low sound pressure level it does not reach very far and requires a close contact between interacting individuals [4]. Third, whispering can be escorted by visible signals, and it shares this characteristic with some social rituals of animals, e.g. the duet displays by many mated birds [9].

As a result of our study this list can be supplemented, now by some further aspects. That is, normal speech can be regarded as an ontogenetic precursor of whispering, because the latter develops only after a good competence of speaking normally has been achieved. Then, the use of a whispering voice requires more effort and costs than that of a phonated voice, and this aspect meets the criteria of 'honest signalling' [3]. According to a recent study [4], the aspect of such costs can be extended and related also to a different sort of costs: i.e. the performance of the signalling can be *risky*. As shown, whispering individuals often run the clear risk to receive signs of discomfort or even protest, if they use this vocalization among unfamiliar people or within a public audience [5].

References

1. Huxley, J. (1966). A discussion of ritualisation of behaviour in animals and man. *Phil. Trans. Royal Society London*, **251**, 247-256.
2. Todt, D. (2003). Rituale als Spielmaterial einer denkbaren Evolution von ästhetischer Erfahrung. *Dimensionen ästhetischer Erfahrung.* Suhrkamp taschenbuch-wissenschaft.
3. Zahavi, A.; Zahavi, A. (1997). *The Handicap Principle*. University Press, Oxford, UK.
4. Vettin, J.; Todt, D. (2005). Human laughter, social play and play vocalizations: an evolutionary approach. *Behaviour*, **142**, 217-240.
5. Cirillo, J. (2004). Whispering: Communication by Unvoiced Speech. *Annals of the Brazilian Academy of Sciences*, **76**, 413-423.
6. Cirillo, J.; Todt, D. (2006). Perception and judgement of whispered vocalizations. *Behaviour*, **141**, 113-128.
7. Watson, A.J.; Valtin, R. (1997). A structural analysis of children's concepts of secrecy. *Australian Journal of Psychology*, **49**, 49-54.
8. Tsunoda, K; Otha, Y.; Niimi, S.; Hirose, H. (1997). Laryngeal adjustment in whispering resonance imaging. *Ann. Otol. Rhinol. Laryngol*., **106**, 41-43.
9. Todt, D.; Naguib, M. (2000) Vocal interactions in birds. *Advances in the Study of Behavior*, **29**, 247-296.

Verbal and non-verbal cues in small groups discussions

S. Livi, A. Battantier, F. Maricchiolo, F. Mannucci
Department of Social and Developmental Psychology, University of Rome "La Sapienza", Italy

Abstract

This study is a preliminary analysis and aims to understand the contribution of hand gesture in defining a powerful/powerless communicative turn in small group conversation in decision-making task discussions. Both turns and hand gestures are coded through systematic observation with The Observer. First results show that speech turns in which speakers use cohesive and/or self-adaptor gestures are more probably interrupted by other group members rather than turns in which deictic and object-adaptor gestures are used by speakers. Implications and future lines of research are discussed.

Keywords

Small group interaction, turn interruption, hand gestures, powerful speech, observational analysis.

1 Introduction

In the last two decades an increasing number of social psychologists are rediscovering group communication as the main path in order to deepening and widening the understanding of groups dynamics [1]. In fact, the theoretical reflection as well as the researches on this topic have demonstrated that contextual characteristics and the personal disposition preexisting at the group's constitution, are not sufficient to predict individual and group performance. In our opinion, the most interesting contribution on group processes are represented by the innovation and advances on methodology of small group research. Apart of the development on statistical analysis (Social Relations Model, Multilevel Analysis, Sequential Analysis [2]), a growing number of coding schemas of group interaction, advanced methodology and techniques on data acquisition of communication processes benefit from the increasing potentiality of personal computers.

The aim of contribution is to present the analysis of the intercrossing of two different coding schemas of the communications acts: one related to the communication content units (conversational analysis) and the second related to the actual non-verbal behavior (gestures during the interaction) applied to a small group discussion on a social psychology lab.

Traditionally, those two coding schemas were analyzed separately as they belong to different research approach even if they represent two important facets of interaction: communication includes spoken language (words and intonation marking topic and focus), but also hand gestures (handshapes, points, beats, and motions). There is a linkage between speech and gesture: hand gestures are synchronized to what is going on in conversation (linked or not to the content of speech) and are able to replace sequences of words. Gestures can also regulate the flow of conversation, search for feedback during an interaction, or express emotion. People produce hand gestures spontaneously while they speak, and such gestures support and expand information conveyed by words.

With this aim, using The Observer, we coded all conversational turns and hand gestures in order to nest the gesture expressed in different kind of turns (namely those interrupted vs. non –interrupted). Given the importance of both conversational turns and gestures we propose the preliminary analysis of the co-occurrence of turn and specific hand gesture modalities linking two methods of analysis of group communication coded through systematic observation via The Observer: in particular, analysis was aimed to see if turns interrupted by other subjects – compared to those non interrupted – present some specific pattern associated with gestures that signal turns transition and/or weakness of the speaker.

1.1 Speaking Turns in Conversational Analysis

Conversation Analysis (CA), allows to study the structures of ordinary face-to-face interaction: People use a suite of techniques for maintaining spoken conversations that are (usually) coherent and understandable. Sacks and colleagues [3] argue that spoken conversations have turn and response structures governed by a set of rules that organize how turns of talk will be exchanged between groups of people, with an orderly allocation system. Using turn-taking rules, people are able to sustain spoken conversations across a wide variety of topics where there is almost always one party talking at a time. Sacks and colleagues underline that interruptions and overlaps occur but are brief, and transitions between speakers commonly occur without gap or overlap. The model proposed further, underlines that transitions between speakers occur at the potential end of a unit type. An interruption occurs when a second speaker begins to speak before a potential transition point occurs. Once a turn is gained, the speaker may establish conversational control by engaging others in a hearer role, maintain or change the topic, or allocate the following turn to a specific other person. Speaking turns are a resource for influence and some authors underline the importance of turns obtained through interruptions in the emerging powerful style [4, 5]. If interruption is deeply analyzed through speakers interruption and how they maintain the floor control in conversations, less attention was devoted to understand why speakers are actually interrupted (or non-interrupted) by other conversants. Probably the "powerful speech" is intended as specular result of asymmetries between interactants describing participant interrupted as simply "victims" of the "powerful" speaker. We believe, though, that, in leaderless discussions, some communication features associated with speaking turns, namely hand gestures, transmit important signals to interactants such as status (see "powerless" speech) as well as some non verbal cues of conversational transition.

1.2 Hand gesture in interaction

Hand gestures can play a role in communicating semantic information [6], in facilitating speech production [7] (or in obtaining persuasion and social influence [8].

In our analysis, we define hand gesture as each movement of one/both hand/s produced by each group member during the discussion. Gestures, coded according to some hand gestures taxonomies (e.g. 9, 10, 11), are divided into two macro-categories: "discourse linked gestures" and "discourse non-linked gestures". The 'discourse linked gestures' category includes **"rhythmic"**, **"cohesive"** and **"ideational"** gestures. Rhythmic gestures (related to vocal speech but not to the actual speech content) are hand/finger movements moved along with the rhythmical pulsation of the speech. "Cohesive gestures" (repetitive hand movements accompanying discourse development, e.g. circular, forward-backward), not linked to the discourse content but to the narrative structure, give it a sense of continuity and coherence The "ideational gestures" refer to the content of speech, including **emblems** (symbolic gestures easily understandable and directly translatable into spoken language), **iconic** (formal reproduction of the object being spoken about), **metaphoric** (referred to abstract concepts) and **deictic** gestures (pointing out an object present either in the physical environment of the speaker or in the "ideal environment" of the discourse). "Non-linked to discourse" gestures include contact gestures, aimed to develop or maintain interpersonal contacts with objects/persons ("hetero **object/person** adaptor") or with parts of ones own body ("**self-adaptor**" gestures).

2 Overview and Design

Participants were 16 female psychology-students at the University of Rome, "La Sapienza" (mean age 21,4; s.d. 1.8) divided in four leaderless four-member groups discussion: each member role-played the manager of one of four corporate departments meeting to negotiate the division of a monetary prize among four candidates representing their respective sectors. To that end, each manager was provided with his/her own candidate's resume. Video recordings of the ensuing discussions were analyzed by three observers that coded conversational turns (in terms of speaking turns interrupted by other speakers *vs.* speaking turns interrupted and hand gestures using The Observer.

2.1 Procedure

Participants were asked to volunteer for a study concerned with a simulation of a group-discussion. They role-play a department manager at a meeting of an Awards Committee of their corporation. The committee's task was to consider four subordinates nominated for a merit-based monetary award. Participants were further told that the four candidates all came from departments whose heads were members of the Awards committee. Information about each candidate included a brief resume and a recommendation letter from the candidate's supervisor. During the committee's discussion each participant's task was: (1) to present valid arguments in favor of his/her candidate, and (2) concomitantly help the committee to arrive at a best decision. Their objectives and mandates thus defined, the four participants commenced their negotiations. Unbeknownst to the participants, behind a one-way mirror a video-camera recorded the entire group interaction. At the end of the session, participants were asked their permission to allow the use of the video-recordings for purposes of data-analysis. All participants consented to this request. The videos was then analyzed by two different observers that, using The Observer, coded numbers and length of both hand gestures and turns interrupted or non-interrupted turns by other participants

3 Results

The results of the preliminary analysis show that subjects use significantly more **cohesive** (f%=47.4) and **self-adaptor** (f%=16.3) gesture during interrupted rather than during non-interrupted turns, whereas **deictic** (f%=24.3) and **object-adaptor** (f%=9.6) gestures are more used during non-interrupted rather than interrupted turns (see Table 1).

Turns Gesture	Non-interrupted	Interrupted	Δ
Cohesive	43.3%	47.4%	**4.1%**
Rhythmic	3.7%	2.2%	-1.4%
Emblem	1.5%	0.7%	-0.8%
Iconic	1.1%	1.5%	0.3%
Metaphoric	5.1%	5.9%	0.9%
Deictic	24.3%	20.7%	**-3.6%**
Obj.-adaptor	9.6%	5.2%	**-4.4%**
Self-adaptor	11.4%	16.3%	**4.9%**
Total	100%	100%	

***Table 1**. Percentage of occurrence of each type of gestures during non-interrupted an interrupted turns and difference.*

4 Conclusion

Interruption is "a device for exercising power and control in conversation" because it involves "violations of speakers' turns at talk" [12]. We have considered interruptions when a second speaker "usurps" another speak7er's right to continue speaking by taking the conversational floor in the absence of any evidence that the other speaker intended to relinquish the turn. In our research lost turns, coded by interruptions, can be linked to specific hand gestures. We found that people interrupted (later losing their turn) used more cohesive and self-adaptor gestures and less deictic and object adaptor gestures, rather than people not interrupted by others. Probably, a large amount of cohesive gestures convey an overproduction of crammed full communication, informing about an already know fact: the speaker is talking. The consequence is that too much cohesive gestures could transfer a perception of excess of "communicative zeal", then stopped by an interruption. Cohesive gestures are normally useful to carry on own argumentations, but, an abuse of such gestures can convey the impression to walk with a sort of "conversational crutches" and they rather risk to communicate speaker's "handicap". Self-adaptor gestures could be perceived by listener as emotional adaptation to a specific context. Nevertheless, using too much self-adaptor gestures could also be perceived as a hidden symptom of irritation, nervousness, anxiety and, probably (but not necessary), of weakness. Further, adopting a communicative style characterized by too much self-adaptor gestures (not directly connected with concrete information) could be perceived as if current speaker would run a bit "out of topic" to the discussion, so such gesture expose to risk to reduce intelligibility. Measures issues concerning less deictic and object-adaptor gestures accompanying interrupted turns could be explained by the concreteness of such hand gestures. Indeed, deictic gestures make concreteness to the concept, giving physicalness, materiality to abstraction. Focusing attention on ideal discursive referent and pointing out virtuality, deictic gestures help to give a sense to ideal or internal objects, transmitting a sense of clearness. Object-adaptor gestures share with deictic gestures the orientation to concreteness: "by touching" objects in environment floor, people occupy physical space like speakers occupy conversational floor.
Overall, these results show the importance of analysis that links coding systems coming from different level of behavioral analysis (verbal and non verbal features). Nevertheless, we think that in order to fully understand groups dynamics of communication is important to analyze the length of the turns and gestures and, most important, the sequence of conversational exchange through appropriate lag analysis [13].

References

1. Levine, J.M., Moreland, R.L. (1990). Progress in small group research. *Annual Review of Psychology,* **41**, 585-634.
2. Kenny, D.A., Mannetti, L., Pierro, A., Livi, S., Kashy, D. (2002). "Statistical Analysis of Data from Small Group". *Journal of Personality and Social Psychology*, **83**, 126-137.
3. Sacks, H., Schegloff, E., Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, **50**, 696-735.
4. Ng, S. H., Bell, D., Brooke, M. (1993). Gaining turns and achieving high influence ranking in small conversational groups. *British Journal of Social Psychology*, **32**, 265-275.
5. Mannetti, L., Battantier, A., Pierro, A. (1996). Dominance in group conversations. *International Journal of Psycholinguistics,* **12**, 297-307.

6. Beattie, G., Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology*, **93**, 179-192.
7. Krauss, R. M., Chen, Y., Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology*, **28**, 389–450.
8. Burgoon, J. K., Birk, T., Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human Communication Research*, **17**, 140-169.
9. Ekman, P., Friesen, W. V. (1969). The repertoire of nonverbal behavior. *Semiotica*, **1**, 49-98.
10. McNeill, D. (1992). *Hand and Mind*. Chicago: The University of Chicago Press.
11. Bonaiuto, M., Gnisci, A., Maricchiolo, F. (2002). Proposta e verifica di una tassonomia per la codifica dei gesti delle mani in discussioni di piccolo gruppo. *Giornale Italiano di Psicologia*, **29**, 777-807.
12. West, C., Zimmerman, D. (1983). Small insults: A study of interruptions in cross-sex conversations between unacquainted persons. In Thorne, B., C. Kramarae, and N. Henley (Eds.). *Language, Gender and Society*. Rowley, (pp. 102-117). MA: Newbury House.
13. Dabbs, J.M. Jr., Ruback, R.B. (1987) Dimensions of group process: Amount and structure of vocal interaction. In L. Berkowitz (Ed.) *Advances in experimental social psychology* Vol. 20, (pp.123-169). San Diego: Academic Press.

Assessment of leading psychic functions development in preschool age

E. Bespanskaya
Department of social-medical work and medical psychology, Belarusian State University, Minsk, Belarus

Abstract
Questions of development of psychic functions system are interconnected to the problem of diagnostics of mental development. Therefore urgency of research is also determined by necessity of construction of such system of functional-stage age qualification diagnostics, which would differentiate not only age stages, but also could give the necessary information on development of system of child's psychic functions. The aim of the study was to develop the diagnostic technique on detection of leading psychic functions for children of preschool age (3–7 years). Methodological base of this research is the function-stage model of ontogenetic development (Yuri N. Karandashev). For each leading psychic function were designed diagnostic technique D1/D1, D2/D2, D3/D3, D4/D4. Were counted psychometric indexes criteria of validity and reliability by yardstick of internal conformity. These techniques can be used for diagnostic of leading psychic functions in preschool age.

Keywords
Development, psychic functions, diagnostic, preschool age.

1 Introduction
The problem of psychic development of children and its diagnostics are very actual in developmental psychology. Studying laws of change of psychic functions, their originalities is very important. Studying of development of mental functions is directly interconnected to a problem of their diagnostics.
Now there is a lot of a various diagnostic toolkit. In preschool psychodiagnostics many of existing diagnostic systems are focused on studying of psychological readiness of the child to school.
Existing techniques are not coordinated among themselves. The contents and structure of techniques depends on the purposes, object of diagnostics and the theoretical positions underlying techniques. Therefore it is necessary to create specific complex system of diagnostics for studying system of leading psychic functions at preschool age.

2 Theory
The system approach to studying mentality and mental development is the most effective [3].
We consider development of psychic functions within the framework of the system approach. The basic concept here is the concept "functional system". In process of psychology development the concept "system" underwent changes. It is possible to tell, that for today there is no uniform definition of the given concept. Many scientists understand system as active structure, structure with a set of transaction, processes of an exchange between her elements [1].
Methodological base of this research is the function-stage model of ontogenetic development (Yuri N. Karandashev) [2]. In this model a child psychic development is considered as a development of system of psychic functions. According to this model each stage of psychic development is characterized by functional system which will consist of base and leading psychic functions. The preschool age is differentiated on two stages. Each of stages shares on two phases. In preschool age leading psychic functions are *"comprehension"*, *"relation"*, *"understanding"* and *"reflection"*. A basis of the named functions is positionality as central growth the given age period.

3 Aims
The aim of the study was to develop the diagnostic technique on detection of leading psychic functions for children of preschool age (3–7 years).
The research problems:
1. The theoretical analysis of leading psychic functions at preschool age.
2. Creation of a technique for each psychic function.
3. Calculation of validity parameters of techniques
4. Calculation of reliability parameters of techniques.

4 Methods
For each leading psychic function were designed diagnostic technique D1/D1, D2/D2, D3/D3, D4/D4.
In the given techniques multi-coloured buttons (red, white, green, yellow, dark blue) are used. Techniques are carried out in the form of game. The purpose - to collect it is as much as possible buttons of the certain color. Color of buttons is set by the experimenter. In game participates two children of preschool age.
The child at performance of techniques can operate with two ways: 1) it is independent and 2) at participation of the adult. If the child independently does not carry out the task the experimenter passes in a mode of the help. If the child has executed the task independently the experimenter passes in a mode of a handicap.
The empirical analysis represented by statistical data processing, conducted with the help of a method by the rank correlation and component analysis.

5 Results
We carried out the procedure of standardization of the developed techniques. Were counted psychometric indexes criteria of validity (Table 1) and reliability by yardstick of internal conformity (Table 2).

Table 1

Techniques	Index of validity
D1/D1	R=0,61, p=0,001
D2/D2	R=0,29, p=0,069
D3/D3	R=0,37, p=0,019
D4/D4	R=0,59, p=0,001

Table 2

Techniques	Index of reliability
D1/D1	r=0,66
D2/D2	r=0,5
D3/D3	r=0,31
D4/D4	r=0,54

6 Conclusions

Pursuant to the obtained outcomes it is possible to draw a conclusion that techniques meet on complexity to preschool age and can be used for diagnostic of leading psychic functions in preschool age.

The given techniques allow to receive the information on a level of development of psychic functions of the child. Further on their base it is possible to construct the complex system of correction of psychic development of children in preschool age. Results of the given research can be used for effective planning process of education and training. Techniques can be used in view of individual and age features of development of the child person.

References:

1. Flammer A. (1988) *Entwicklungstheorien. Psychologische Theorien der menschlichen Entwicklung.*
2. Karandashev Y. N. (1991). *Age function-stage model of ontogenetic development of children.*
3. Vigotskiy L.S. (1984). *Children's psychology.*

Externalizing behavior problems of young criminals with different level of subjective control

Svetlana Stupish
Department of Forensic Psychology, Academy of Ministry of Internal Affairs, Minsk, Belarus

Abstract
The article is devoted to the research of the locus of control of juvenile criminal offenders with different sorts of behavioral disorders including aggression, disruptive behavior, pathological gambling, pyromania, sexual disorder and antisocial personality disorder. The purpose of this study was to examine the relationships between locus of control and adolescent behavioral disorders. Significant differences in the level of subjective control of young criminals are revealed. The results of the research can be used in the practical work of psychologists of education and penal system, for the elaboration of the strategy of intervention of behavioral disorders in adolescence.

Keywords
Internal-External locus of control, juvinile criminal offenders, behavioral disorders.

1 Introduction
Deficiencies in impulse control and executive function are prominent features of childhood-onset behavioral disorders and could entail an increased risk of criminal offending.
Existing techniques for the study of the level of subjective control are not coordinated among themselves. The contents and structure of techniques depends on the purposes, object of diagnostics and the theoretical positions underlying techniques. In our research we used modified version of "The Subjective Control's Level Scale" (Bajhyn, Golynkina, Etkind, 1990) for delinquent adolescents.

2 Theory
The problem of subjective control was studies by the psychologists in frameworks of the various theoretical concepts. Edwards (1957), Millham and Jacobson (1978), Crowne and Marlowe (1980) were engaged in research of the locus of control and its influence on the behavior of the person. The concept of I -E was first proposed by J.Rotter (1966), and it forms a relatively small part of a more extensive personality theory incorporating many of the principles established in the psychology of learning. This theory is known as *social learning theory* (Rotter, 1954; Rotter, Chance and Phares, 1972). Rotter proposed that the degree to which people believe their lives to be under their own control is an important dimension of individual variation. People who are relatively internal believe they are responsible for their destiny, whereas people who are relatively external believe that the good and bad things that happen to them are determined by luck, chance or powerful others. In Rotter's social learning theory I-E is regarded as a characteristic attitude towards the world, referred to as a generalized expectancy. The expectancy about the locus of control over rewards and punishment generated by a person's position on the I-E dimension will influence the way that person perceives most situations, and hence will partially determine how the person will behave.
Research into locus of control can be divided into three categories:
1) the study of different components of I-E,
2) the characteristics of relatively internal and relatively external people,
3) the investigations of the interplay between I-E and other determinants of behaviour.
Much of this research involves making comparisons between subjects identified by measures of I-E as tending to believe that the *locus of control is internal* and those tending to believe that the *locus of control is external*. As a shorthand these two groups will be referred to as *"internals"* and *"externals"*, but it should be remembered that I-E is a continuum and there are no internal or external types, but rather only degrees of internality and externality.
There are two ways in which narrower distinctions subsumed by the broad locus of control dimension may be made. *The first* is situational: a measure of I-E may be developed to assess locus of control beliefs pertaining to a specific subset of situations. For example, the Health Locus of Control Scale was developed to measure I-E beliefs with respect to health-related behaviour (Wallston, Wallston and DeVallis, 1978). There is also a measure of locus of control for affiliation and behaviour in social interactions (Lefcourt, von Baeyer, Ware and Cox, 1979). Internal affiliation beliefs are associated with active listening and socially skilled behaviour in social interaction (Lefcourt, Martin, Fick and Saleh, 1985). *The second* way of subdividing I-E involves the fractionating of the construct itself into different components. One such multidimensional measure of I-E assesses two components of external beliefs – control by powerful others and control by chance – in addition to measuring internal beliefs (Levenson, 1974). Another subdivision of I-E measures the extent to which people believe in internal locus of control for successful outcomes and external locus of control for failures.
The characteristics of internal and external people have been explored in a large number of studies (Lefcourt, 1982; Phares, 1976, 1978; and Strickland, 1977). Investigations reported by Phares (1978) into the relationship between I-E and demographic variables suggest that there are no sex differences in I-E, that whites tend to be more internal than blacks, and that middle-class subjects tend to be more internal than working-class subjects. Phares also reports that there is no substantial relationship between intelligence and I-E. From the research summarized by Phares (1978) relating I-E to a wide variety of behaviors. The internal person is more likely to be receptive to aspects of health care such as weight watching, giving up smoking, taking exercise and carrying out prophylactic measures. Their desire for self-determination is reflected in their greater resistance to social influence and attempted attitude change. In the area of mental health, internals are generally found to be better adjusted and less anxious than externals, and external beliefs are symptomatic of a number of psychiatric disorders such as

depression and schizophrenia. In short, the internal individual, in contrast to the external, is independent, achieving and masterful.

3 Aims

The main aim of the research was to study of correlations between level of subjective control and different sorts of behavioral disorders including aggression, disruptive behavior, pathological gambling, pyromania, sexual disorder and antisocial personality disorder among juvenile criminal offenders.

4 Methods

The Sample. The objects of the present research are 13-17-year-old Belarusian juvenile criminal offenders with different sorts of behavioral disorders.

Methods In order to reach the aim of investigation following techniques were used: modified version of "The Subjective Control's Level Scale" (Bajhyn, Golynkina, Etkind, 1990), biographical techniques, and Screening Test for Adolescent's Personality Disorders (Krasov).

The empirical data analysis was conducted by methods of the rank correlation and component analysis.

5 Results

The results show that there are correlations between locus of control (Internal-External) and forms of behavioral disorders in adolescence.

There are **significan**t **positively correlations** between:

- General Externality and disruptive behavior,
- Externality in the area of health and eating disorders,
- General Externality and pathological gambling,
- General Internality and auto-aggression,
- Internality in the area of achievements and eating disorders,
- Externality in the area of interpersonal interaction and sexual disorders.

There are **significant negatively correlations** between:

- antisocial personality disorders and general Internality
- Internality in the area of achievements and sexual disorders,
- aggression and general Internality,
- Internality in the area of family relationships and antisocial personality disorders,
- Internality in the area of achievements and Internality in the area of family relationships both correlated negatively with disruptive behavior.

6 Conclusions

Previous studies showed considerable influence of Internal-External control in determination of behavioral disorders in adolescence.

Adolescents with low scores of General Internality (low level of subjective control) demonstrate aggression, disruptive behavior, pathological gambling and antisocial personality disorders. Adolescents with high scores of General Internality (high level of subjective control) are less inclined to behavioral disorders and only in rear cases show auto-aggression and some borderline personality disorders. Therefore, we could suppose that External locus of control (low level of subjective control) is one of the risk factor in adolescent aggression, disruptive behavior, eating disorder, pathological gambling, sexual disorder and antisocial personality disorder. But at the same time we agreed, that low level of subjective control may be co-morbid with these behavioral disorders.

References:

1. Fonagy P., Target M., Steele M. Crime and Attachment: Morality, Disruptive Behavior, Borderline Personality Disorder, Crime, and Their Relationships to Security of Attachment //Attachment and Psychopathology / Ed. by L. Atkinson and K.J. Zucker. – N. Y., 1997. – P. 154 – 209.
2. Hampson S.E. The Construction of Personality. – London and New York, 1998. – 320 p.
3. Lefcourt H. Locus of Control: Current Trends in Theory and Research. N.Y., 1996. – 70 p.
4. Rotter J.B. Internal versus External Locus of Control // Psychol. Monogr., v.80, N 1 (whole N 609). – P. 56-90.

Complications of the stage of separation in teenagers with health problems

N. Volodarskaya
Laboratory of psychology of personality, Institute of psychology of Ukraine, Kiev, Ukraine

Abstract
The present article deals with the problem of separation of a child-invalid from his/her mother and family related with family dysfunctions are considered. I used psycho-correction oriented on changing relations between a child-invalid and his/her parents. Specifics of these relations consist of complete confluence. Step by step therapy of such relations positively influences rehabilitation of family functions. Parents passing through the therapy improve their mutual relations. The therapy described in this article assists in changing relations of parents with their child and promotes his/her separation.

Keywords
Separation, personal frames, relations of confluence, verbalization of feelings, dysfunction.

Introduction
A child with health problems goes through the complicated stage of separation from his/her mother in the family. This stage can take many years. An uncompleted separation can occur after marriage or the birth of children. Grandchildren play the role of children for women, and the mother prolongs her role with the guardianship of the new young family. A mother is in confluence with her child-invalid. When declaring anxiety for her child's independence, in reality she subconsciously prevents this process from taking place. Any attempt by the child to make independent decisions is interrupted by the mother, are condemned in an open or hidden form, and creates situations of "closed circle". It is the mother's initiative to ask for assistance of a psychologist in getting her child's separation and independence. Determination of internal and external borders of a child's personality is complicated job for psychologists.

Methods of psycho-correction
In the method I apply, I give the task of defining the territory of the mother and the child using a peace of cord. One end of the cord is given to the mother and the other end to the child, so they are asked to make circle representing their territory. As a rule, circles have different sizes. The definition of their territory's size gives the possibility to correct internal borders of personality. Using toys, mutual relations are defined between the mother, the wife and the children. Selection of toys characterizes a role and place in the family, and family members gradually realize these roles and their correction. It is possible to take pictures of toys arrangement at the first meeting and at the last meeting after therapy. Comparison of pictures and analyses of toys selection and arrangement is very indicative for the dynamics of mutual relations in a family. It is important in domestic psychotherapy to do it with all members of a family, not only with the mother and her sick child. The mother and her child get home tasks. I ask them to play the role of independent family members for some time, making independent decisions. The mother admits the right of her child to make mistakes.

Early therapy and its results
Early therapy helps mollify the separation process of a child with health problems. It is very is effective to use the method of family arrangement on the board of Gering. Selection of wooden figurines of family members and their arrangement on the board helps the mother and her child to understand mutual relations of confluence. The therapy used the method of the genogram, showing that the patterns of behavior of parents of invalids frequently pass from generation to generation. If parents had relatives-invalids in their family, then their behavior with respect to their child with health problems changes, they have definite experience of relations with the child-invalid, and emotional tension in the family is not so big. For parents who had invalids in their family it is easier to stand emotional trauma after child-invalid birth, matrimonial relations become stronger and their family consolidates. The aim of assisting their child in his/her socialization is his/her entry in society. If such aim of parents translates into hypertrophied anxiety for child guardianship, then it results in confluence, codependence, and feeling of solitude of the child with the mother or father. In these cases it is necessary psycho-correction of mutual relations in the family.

Psycho-correcting work with parents of children-invalids includes such phases as realization of their feelings, their verbalization and training to express feelings directly by all members of a family.

As a rule in families with a child-invalid we note distortion of family functions. In such families attention of all members is riveted on the problem of the child-invalid. It is distorted educational function and is prolonged after child's majority. Parents are trying to fulfill this function even after their child-invalids creates its own family. The economic and social function of the family has not been created, the child-invalid does not have formed experience of housekeeping. It is distorted emotional function of the family – the members do not have satisfied their needs in liking, respect, admission, emotional support and psychological protection. These distortions lead to emotional instability and breach of interpersonal communications with other people out of the family.

Due to confusion about personal limits among family members, they appropriate emotional reactions of other family members. Also the child-invalid is out of emotional contact. Nobody knows what he feels.

The function of social control is frequently concentrated in the hands of one of the family members, as a rule, in the hands of the mother. She fulfills the function of initial social control. This function is completely distorted. All family members control each other.

With time, significance of these functions changes. Some functions are lost, other change accordingly with change in social conditions. The therapy is oriented to recognition and re-comprehension of all family functions, since many of them have been distorted and disturb the action of the family.

Depending on the way of interruption of contact with the environment, members of a family with children-invalids are divided in confluent and retro-flexible. More frequently parents are in confluence with a child-invalid. They cannot and do not know how to express their feelings and they suppress them. The family is characterized with rigid external frames and weak principles of sub-systems.

Individual frames inside the family are weak, there is no territory that belongs to any individual, adults permanently interfere in affairs of children (even grown-up), control their actions, time and plans.

The family members are indifferent, they easily get in emotional confluence with each other. They are not able to separate their emotions from emotions of other people. Emotional involvement is so high, that family members are always sure that they know thoughts and opinions of each other. Such families are characterized with breach of psychological frames. Family members are trying to interfere in personal life of their relatives.

Rigid frames do not allow the family to interact with the external world. Therapeutic interventions are necessary in such case, they are oriented on weakening frames, providing more deep contact and adaptation. So we can get positive changes in putting right family functions. It is necessary to direct therapy to changing rules and roles in the family and establishing personal frames of each family member.

Ambulatory recording of cardio-respiratory signals in freely moving human subjects

E.J.C. de Geus
Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands

The Vrije Universiteit Ambulatory Monitoring System (VU-AMS) is used for the ambulatory 24-hour recording of peripheral physiology in human subjects. Although originally developed to serve local research goals, the VU-AMS is now used world-wide by over 40 research groups (and increasing) to study stress and emotion in naturalistic settings. Main variables measured are Heart Period, Heart Period Variability, Pre-Ejection Period, Respiration Rate, Respiration Depth, Stroke Volume and Cardiac Output. This SIG meeting intends to review the accumulated evidence on the reliability and validity of ambulatory recording with the VU-AMS, but to also draw clear boundaries of what can be validly measured at the current state of the technology. Houtveen presents extensive cross-instrument comparisons of respiration rate and depth measurements, Goedhart on the temporal stability of impedance derived ambulatory PEP and Stroke Volume, and Kupper on the genetic contribution to individual differences in ambulatory heart rate variability and pre-ejection-period. Paul Groot demonstrates novel VU-AMS tools that allow stimulus presentation by E-prime to be cued by cardiac signals (e.g. R-waves) and the use of VU-AMS software to score the impedance and respiration signals obtained with the BioPac system. Finally, novel technological developments in the VU-AMS are be reviewed and a potential future agenda for ambulatory recording of human autonomic nervous system functioning is outlined by De Geus.

Temporal stability of ambulatory stroke volume and cardiac output measured by impedance cardiography

A.D. Goedhart, N. Kupper, G. Willemsen, D.I. Boomsma and E.J.C. de Geus

Ambulatory recording of heart rate and blood pressure has furthered our understanding of how repeated or chronic cardiovascular activation in response to naturalistic events can contribute to cardiovascular disease. The assessment of cardiac output would further complete this picture of hemodynamic regulation in naturalistic settings. Recently, devices have become available that allow non-invasive measurement of stroke volume and cardiac output through ambulatory thorax impedance recording. If such recordings have adequate temporal stability, they offer great potential to further our understanding of how repeated or chronic cardiovascular activation in response to naturalistic events may contribute to cardiovascular disease. An important question is whether ambulatory impedance-derived stroke volume and cardiac output show acceptable temporal stability. In this study, 24-hour ambulatory impedance-derived stroke volume, cardiac output and systolic time intervals were measured in 65 healthy subjects across an average time span of 3 years and 4 months. Because significant increases in stroke volume were seen during sleep, coupled to a decrease in heart rate and cardiac output, intraclass correlations were computed separately for sleep and daytime recordings. Intraclass correlations were moderate for stroke volume (.30–.39) and cardiac output (.32–.40) and good for systolic time intervals (.53–.82). Stroke volume derived from the dZ/dt¬¬(max) value based on the B-point in the impedance signal resulted in higher correlations to that derived from the dZ/dt¬¬=0 baseline, particularly during sleep. We conclude that moderate long-term temporal stability was found for individual differences in ambulatory stroke volume and cardiac output measured by impedance cardiography.

Validation of the thoracic impedance derived respiratory signal using multilevel analysis

J.H. Houtveen, P.F.C. Groot and E.J.C. de Geus

The purpose of the current study was to validate the change in thoracic impedance (dZ) derived respiratory signal obtained from four spot electrodes against incidental spirometry. A similar validation was performed for a dual respiratory belts signal to compare the relative merit of both methods. Participants were 38 healthy adult subjects (half male, half female). Cross-method comparisons were performed at three (paced) respiration frequencies in sitting, supine, and standing postures. Multilevel regression was used to examine the within and between-subjects structure of the relationship between spirometric volume and the respiratory amplitude signals obtained from either dZ or respiratory belts. Both dZ derived respiratory rate and dual-belts derived respiratory rate accurately reflected the pacing frequencies. For both methods, fixed factors indicated acceptable but posture-specific regression on spirometric volume. However, random factors indicated large individual differences, which was supported by variability of gain analyses. It was concluded that both the dZ and dual-belts methods can be used for measurement of respiratory rate and within-subjects, posture-specific, changes in respiratory volume. The need for frequent subject-specific and posture-specific calibration combined with relatively large measurement errors may strongly limit the usefulness of both methods to assess absolute tidal volume and minute ventilation in ambulatory designs. Additionally, the results of an exploratory multilevel path analysis on this data set will be presented. This analysis was used to examine the extent to which changes in posture and respiration can confound the Pre Ejection Period (PEP) and Respiratory Sinus Arrhythmia (RSA) as indices of cardiac sympathetic and vagal activity.

Heritability of cardiac autonomic balance from ambulatory recordings in a large-scale twin sample

G. Willemsen, N. Kupper, D.I. Boomsma and E.J.C. de Geus

Cardiac disease has been associated with a distinct shift in the autonomic control over the heart; parasympathetic activity is decreased, while sympathetic activity is increased. Laboratory studies have shown individual differences in parasympathetic activation, such as respiratory sinus arrhythmia (RSA), to be partly genetically determined. Little is known however about the heritability of indices of sympathetic control over cardiac contractility, such as pre-ejection period (PEP). Moreover, the contribution of genetic and environmental factors to the variation in autonomic activation has not yet been investigated during prolonged periods of ambulatory monitoring in a naturalistic setting. Using the VU-AMS, we recorded ambulatory ECG and thorax impedance signals during 24 hours in 790 healthy twins and their singleton siblings. Average PEP and RSA was determined for four periods of the day (morning, afternoon, evening, night). Multivariate genetic analyses showed significant heritability at all daily periods for PEP (48 to 62%) and RSA (40% to 55%). In conclusion, ambulatory measured indices of autonomic balance show substantial heritability at all periods of the day and during sleep.

Combining the VU-AMS, E-Prime and BioPac in laboratory designs

J. den Hartog and P.F.C. Groot

In this presentation a method is described which allows VU-AMS software to be used to analyze ECG, thorax impedance (ICG) and respiratory signals, which are measured by BioPac acquisition hardware in a laboratory setting. Since most laboratory designs make use of computerized tests during the physiological measurement, a mechanism is proposed for storing coded event markers as integral part of the recording to be able to classify specific parts of the data during the analysis phase. Part of this mechanism is a software procedure which translates BioPac signals and events markers to labeled signals in an AMS compatible file format, so existing VU-AMS analysis software can be used for in-depth analysis of thorax impedance signals to obtain PEP and RSA. Although EPrime is used during this presentation as a typical application to control the flow of psychological testing, any comparable stimulus presentation software can be used in a similar way, as long as it is able to output digital signals by using a standard printer port.

Vrije Universiteit Ambulatory Monitoring System: quo vadis?

E.J.C. de Geus, J. den Hartog and P.F.C. Groot

The Vrije Universiteit Ambulatory Monitoring System (VU-AMS) plays a key role in the cardiovascular research of the department of Biological Psychology and the Netherlands Twin Registry on the (interactive) influence of genetic and environmental factors (e.g. chronic stress) on cardiovascular disease risk. This ongoing scientific demand from within has maintained the VU-AMS in a continuous state of innovation since its first prototypes in the early 1990's. This demand for continuous innovation is reinforced by the over 40 external research groups that use the VU-AMS world-wide. To meet this demand, we will release the first VU-AMS of the new 5fs-series in the fall of 2005. The new full storage VU-AMS (version 5fs.1) will continuously record the ECG, thorax impedance and (optionally) skin conductance and store these to flash memory. The new version will not lead to a change in the main VU-AMS philosophy of rapid data-reduction based on ensemble averaging over labeled periods. Ensemble averaged values over the relevant labeled periods will still be extracted in as short a time frame as possible without compromising full visual interactive control over signal scoring. Put otherwise, the device will be downward compatible to existing software and data analysis strategies. The new device tops the old one, however, on three major accounts. A number of new variables can be extracted including the T-wave amplitude and QT interval variability (both indices of sympathetic drive that are less dependent on pre- and afterload), convergence of skin conductance and thorax impedance changes (can now jointly measured by a single device), and Fourier- or wavelet-based frequency decomposition of the variance in heart period and respiration signals. Secondly, because memory is flash-card based, repeated 24-hour measurements on the same subject can be scheduled by minimal contact of the researcher with the participant/patient. Finally, availability of continuous data sampled at a much higher frequency (1000 Hz) than before will allows us to reduce measurement error and increase automaticity of scoring of the impedance cardiogram (ICG) and the respiration signal.

Advancements in measurement of human motion

C. Baten[1] and J. Harlaar[2]
[1] *Roessingh Research and Development, Enschede, The Netherlands*
[2] *Vrije Universiteit Medical Center, Amsterdam, The Netherlands*

An increasing number of professional disciplines is interested in analysis of human motor behavior to assess its functioning or performance, e.g. rehabilitation medicine, orthopedics, physiotherapy, ergonomics, experimental psychology and sports. Besides observational analysis of human motion, state of the art methods provide accurate measurements of the kinematics of posture and movement, muscle activation patterns and external forces. These measurements enable the use of biomechanical models to provide a comprehensive assessment. Traditionally the most accurate methods apply some sort of video-based marker tracking methods. By tracking data with multiple fixed position cameras, 3D posture and motion data is gathered. Together with muscle activation data (gathered as electromyographic signals) and forces exchanged with the outside world (typically through the feet and measured with force plates in the floor) a complete assessment of performance or problems can be made. This comprises human motor control as well as the mechanical loading of the musculoskeletal system. However, fixed cameras and the fixed force plate limit possible motions and actions of the subject. Also the required stereophotogrammetric equipment requires high investments and a motion analysis session is typically a very time-consuming effort. So analysis of human movement (including gait analysis) is exclusively available to some (academic) laboratories. Currently technical and methodological alternatives are becoming available that form relatively cheap alternatives which are also very easy to apply in (clinical) practice. These methods promise to facilitate accurate motion analysis in any room or even in situations in which the subject is freely roaming about. This Special Interest Group introduces several new developments in miniature sensing and marker-less video motion analysis in a workshop in which all aspects of practical ambulatory or location-independent motion analysis are discussed.

Human motion analysis using inertial sensing
C. Baten

Recent developments have delivered new technology for ambulatory assessment of 3D kinematics of miniature motion sensors. These sensors combine signals from 3 accelerometers, 3 rate gyroscopes, 3 magnetometers and a temperature sensor, all inside one module, through optimal Kalman filtering technology into accurate estimates of 3D orientation, angular velocity and several components of 3D acceleration. Orientations are derived as the orientation of the sensor casing axes relative to a global inertial 'world' coordinate system and velocities and accelerations are derived relative to the sensor casing axes. To perform human motion analysis with this type of sensor technology human body segment kinematics have to be derived from the sensor kinematics. For this the relationship between the casing axes frame and the axes frame of the body segment on which the sensor is mounted has to be known.
Because mounting the sensor such that both body segment and sensor axes frame are in parallel is not a practically viable option, the sensors are mounted as stable as possible after which the relative 3D orientation of the sensor axes frame and the body axes frame are determined in a short calibration procedure. In this procedure sensor data is recorded during a few recordings in which a: the body segment is rotated purely around one of its defined segment frame axes or b: one of the body segment frame axes is kept parallel with one of its defined frame axes.
This 'helical axes'-calibration was tested and validated in normals for accuracy and robustness for 3 parts of the trunk, for shank and thigh and for upper and lower arm.
The quality of this calibration determines the quality of the clinical motion assessment. It depends in part on the ability of the subject to consistently perform the required motions. Therefore research is started to develop derived methods for use with patients in several areas of motion analysis and test accuracy, robustness of variations of this methods and protocols.

3D motion analysis applying inertial sensing
D. Roetenberg

Small inertial sensors like accelerometers and gyroscopes are more and more used in ambulatory motion analysis (Busseral. 1998; Baten et al. 2000; Veltink et al. 2003). Typically, angular orientation of a body segment is determined by integrating the output from the angular rate sensors strapped on the segment. Microelectromechanical (MEMS) gyroscopes are accurate for angular velocity measurements but can only be used for a short time to calculate angular orientation. A relatively small offset error due to temperature effects on the gyroscope signal and noise will introduce large integration errors. Linear accelerometers measure the vector sum of acceleration a of the sensor and the gravitational acceleration g. In most situations of human movement sensing, g is dominant, thus providing inclination information that can be used to correct the drifted orientation estimate from the gyroscopes. However, accelerometers cannot detect rotations around the vertical axis, therefore, additional magnetic sensing is required. The magnetometer is sensitive to the earth's magnetic field. It gives information about the heading direction in order to correct drift of the gyroscope about the vertical axis. Several filters like (Foxlin 1996; Bachmann 2000) have been proposed to fuse gyroscopes, accelerometers and magnetometers for body segment orientation measurements. They work in real-time and seem accurate but they have an important limitation. Ferromagnetic materials, like iron, in the vicinity of the sensor will disturb the local magnetic field and will therefore distort the orientation measurement. This magnetic interference impedes many applications; especially the ones outside the lab like back load estimation for ergonomic purposes at assembly lines. In this study, a Kalman filter based on (Luinge et al. 1999) is designed to fuse 3D gyroscope, accelerometers and magnetometer signals to estimate orientation and to compensate for magnetic interferences.

Ambulatory measurement of ground reaction forces

P.H. Veltink, C. Liedtke, E. Droog and H. van der Kooij

The measurement of ground reaction forces is important in the biomechanical analysis of gait and other motor activities. Many applications require full ambulatory measurement of these forces, but this is not supported by current measurement systems. We propose the use of two six degrees of freedom force and moment sensors under each shoe, which enables the ambulatory measurement of ground reaction forces and Centers of Pressure (CoP). The feasibility of this method is illustrated by experimental results in a healthy subject, using a force plate as a reference. The ground reaction forces and CoP recordings show good correspondence when they are evaluated for forces above 40 N and when it is simply assumed that the sensors are flat on the ground when they are loaded. The RMS difference of the magnitude of the ground reaction force over 12 gait trials was 15 ± 2 N, corresponding to 1.9 ± 0.3 % of the maximum ground reaction force magnitude. The RMS difference of the horizontal component of the ground reaction force was 3 ± 2 N, corresponding to 0.4 ± 0.2 % of the maximum ground reaction force magnitude and to 2 ± 1 % of the maximum of the horizontal component of the ground reaction force. The RMS distance between both CoP recordings is 2.9 ± 0.4 mm, corresponding to 1.1 ± 0.2 % of the length of the shoe, when the trajectories are optimally aligned.

Clinical applications of gait analysis

J. Harlaar

In clinical practice of rehabilitation medicine, the treatment of mobility can be enhanced by the introduction of gait analysis as an assessment tool. In order to provide meaningful information to the physician, clinical gait analysis should disclose the functions of structures of the human movement system that are a potential therapy target. This will include joint kinematics, muscle-functions and joint loads. Measurement of all this these variables, would require a highly sophisticated lab, yielding maximal accuracy. Fortunately, in order to serve clinical decision making, accuracy can be limited to clinically relevant values, based on biological variability.

A system for clinical movement analysis is presented, that aims at feasibility for its infrastructure in a standard clinical setup. This includes a biplanar video registration, surface EMG recording of muscle activity and ground reaction forces under the stance leg. Integration of all this signals, is realized by a multimedia software application . This application visualizes the measured signals in a way that is meaningful to the physician and can be interpreted very easily. Moreover, the application can be shared over the internet, in order to consult colleagues. It has proven a significant contribution to current clinical practice.

However, in order to assess movements in the transversal plane, biplanar video is not suited. For these kinematics and to perform more precise measures in other planes, a combination with inertial sensors is a relevant extion within the concept of clinical gait analysis.

Automatic assessment of dyskinesia in Parkinson's disease in daily life

N. Keijsers, M.W.I.M. Horstink and C.C.A.M. Gielen

During the first years of levodopa treatment, patients with Parkinson's disease (PD) have a stable response to levodopa. However, after several years of levodopa treatment, an increasing number of patients show fluctuations in motor response ("on"-"off" fluctuations) and levodopa induced dyskinesias (abnormal involuntary movements). These complications constitute a major problem in the long-term management of PD and add substantially to the patient's disability. New pharmacological and surgical treatments to reduce levodopa induced dyskinesias are becoming of more and more interest. Therefore, an automatic and portable device that can assess LID automatically and objectively in daily life is highly useful.

Thirteen patients were continuously monitored in a home-like situation for a period of approximately 2.5 hours. During this 2.5-hour period, the patients performed about 35 functional daily-life activities. Behavior of the patients was measured using triaxial accelerometers, which were placed on 6 different positions of the body. A neural network was trained to assess the severity of LID using various variables of the accelerometer signals. Neural network scores were compared with the assessment by physicians, who evaluated the continuously videotaped behavior of the patients off-line.

Neural network correctly classified dyskinesia or the absence of dyskinesia in 15-minute intervals in 93.7, 99.7 and 97.0% for the arm, trunk and leg, respectively. In few cases of misclassification, the rating by the neural network was in the class next to that indicated by the physician using the AIMS-score (0-4). The percentage of time that a segment was moving was the most important parameter used by the neural network. For the leg mainly parameters of both legs were important. For the arm and especially for the trunk, parameters, related to movements of other body segments, were relevant. Dyskinesia appeared to be more dominant in the lower frequencies than in higher frequencies.

The neural network could accurately assess the severity of LID and distinguish LID from voluntary movements in daily life situations. The results suggest that the method could be operating successfully in unsupervised ambulatory conditions.

Automated dynamic activity monitoring

R. Wassink and C. Baten

In some applications of ambulatory assessment of 3D human motion typically detailed kinematic and kinetic biomechanical data is gathered or derived over longer periods (hours), e.g. in ergonomic applications. For sensible interpretation of this data continuous context information is required, e.g. 'activity performed'. Currently this data typically is gathered by labor intensive manual observation, while all other data gathering is fully automated.

To get rid of the labor intensive manual observations a new method is proposed for automated activity monitor for human activities using the data already automatically gathered in the kinematic assessment.

Also in clinical motion analysis applications of ambulatory recording technologies (portable gait lab) a need has risen for automated context classification. Typically data is gathered of many motion cycles (e.g. steps in Gait in analysis) and to facilitate wide spread use by health care professionals automated cycle based interpretation assistance is required. For e.g. gait analysis this means estimation of kinematic and kinetic step cycle parameters and typical derived statistics. A requirement for this is automated step cycle detection and step cycle classification.

Both application areas of ambulatory motion analysis require automated activity recognition and classification including estimation of start and end times.

The proposed method applies self learning Hidden Markov Modeling (HMM) technique to the kinematic data. In a training phase for each activity a HMM is derived from training data. In the application phase for each the HMM the probability is estimated that the current activity is the one represented by this HMM. In a post processing phase these probabilities are used to decide on which activity is currently recognized. In a more sophisticated estimator also (estimated) a priori probabilities of occurrence are taken into account in the post-processing.

Pilot experiments have indicated that HMM methodology seems very capable of delivering the requires recognition and classification functionalities. Current research aims at examining accuracy, robustness and generalizability of the methodology. In this SIG contribution possibilities and experiences with applying HMM are discussed as well as its potential for a general Activity Monitor for use in functional evaluation in rehabilitation and ergonomy.

Identification of determinants of performance in sports: a case study in rowing

M. Hofmijster

Feedback on biomechanical and physiological parameters is of paramount importance to athletes. Apart from helping athletes to improve performance, feedback also enhances motivation, irrespective of the skill level. Traditionally, feedback is provided by a trainer or coach, relying on visual observation. In recent years, electronics and sensor technology have developed to the extent that it is now feasible to develop portable measurement and feedback devices that can provide athlete (and coach) with objective information on biophysical parameters (motion, forces, heart rate).

However, electronic feedback is useless when it does not contain meaningful information for the athlete in question. In this perspective experienced coaches provide a valuable source of information. To capture their implicit knowledge in explicit biophysical terms is not straightforward. A well defined theoretical framework is necessary. Rowing was chosen as a case study since performance is easy to quantify (namely average velocity over the race distance) and technique plays an important role. The approach described may also be applicable in other types of (endurance) sports however.

The analysis of rowing is based on the mechanical power equation (e.g. van Ingen Schenau and Cavanagh, 1990). Averaged over a rowing cycle, the rower delivers a certain amount of mechanical power. Part of this power is lost at the oar blades during the push-off, because water is set in motion. The remainder of the power is dissipated by the drag force on the hull. Conceptually, the power loss due to hull drag can be separated into an effective part that is related to average boat velocity, and an ineffective part that is related to substantial intracycle fluctuations in boat velocity. From the perspective of the power equation, the rower with the better technique has less power loss at the blades and/or less power loss due to intracycle boat velocity fluctuations, resulting in a higher average velocity.

Feedback on these power losses or feedback on forces or motions that lead to these losses may help the rower in achieving better performance. Ongoing cooperation between industry, science and sports is necessary to ensure new technologies will be used in the most advantageous way.

Ambulatory Measurement of Ground Reaction Forces

Peter H. Veltink, Christian B. Liedtke, Ed Droog, Herman van der Kooij
Institute for Biomedical Technology, University of Twente, Enschede, The Netherlands

Abstract

The measurement of ground reaction forces is important in the biomechanical analysis of gait and other motor activities. Many applications require full ambulatory measurement of these forces, but this is not supported by current measurement systems. We propose the use of two six degrees of freedom force and moment sensors under each shoe, which enables the ambulatory measurement of ground reaction forces and Centers of Pressure (CoP). The feasibility of this method is illustrated by experimental results in a healthy subject, using a force plate as a reference. The ground reaction forces and CoP recordings show good correspondence when they are evaluated for forces above 40 N and when it is simply assumed that the sensors are flat on the ground when they are loaded. The RMS difference of the magnitude of the ground reaction force over 12 gait trials was 15 ± 2 N, corresponding to 1.9 ± 0.3 % of the maximum ground reaction force magnitude. The RMS difference of the horizontal component of the ground reaction force was 3 ± 2 N, corresponding to 0.4 ± 0.2 % of the maximum ground reaction force magnitude and to 2 ± 1 % of the maximum of the horizontal component of the ground reaction force. The RMS distance between both CoP recordings is 2.9 ± 0.4 mm, corresponding to 1.1 ± 0.2 % of the length of the shoe, when the trajectories are optimally aligned.

Keywords

Ambulatory sensing, ground reaction force, center of Pressure, instrumentation, biomechanics

1 Introduction

Ground reaction forces (GRF) are currently measured using lab-bound force plates built in the floor. In most cases one or, at the most, two plates are available. This limits GRF measurement seriously: first of all, the subject needs to place his or her foot or feet completely on the force plate in order to perform a correct measurement of the total ground reaction force. In walking trials this means that subjects have to walk exactly over the force plates. Secondly, only one or two steps during a gait trial can be measured, while there is, in general, a large variation in the ground reaction forces between steps, related with differences in muscle activation and body movement. Thirdly, when standing on a single force plate, the GRF of a single foot can not be distinguished, the plate only measures the total GRF. Fourthly, the use of fixed force plates impede ambulatory GRF measurement in any place during daily-life activities at home and at work. The only ambulatory alternatives to force plates used to our knowledge are pressure sensor matrices, placed inside the shoe [1]. It should be noted, however, that matrices of pressure sensors only allow the estimation of the vertical component of the GRF, not the shear components. In addition, pressure measured by pressure sensor matrices applied inside the shoe, does not add up to the GRF under the shoe, because of the pressure induced by the fitting of the shoe. A method to estimate GRF using insole pressure sensors and additional knowledge of body movements has been proposed [2], but an independent measurement of GRF is preferred. Sheets that can measure all components of stress inside a shoe have been proposed [3], but are not yet used in regular human gait analysis. Recently, also carpets of pressure sensor matrices on the floor are used [1]. This allows more than two steps to be measured, but is, like force plates, not an ambulatory system.

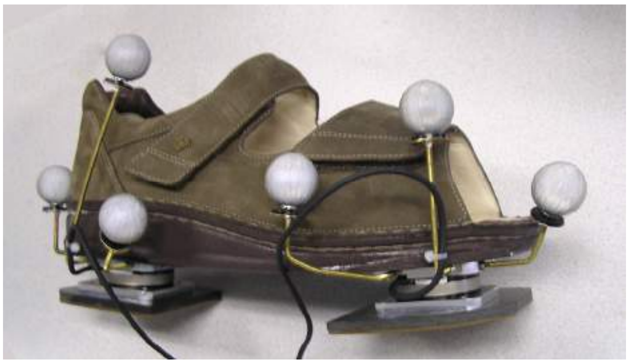


Figure 1. *Shoe instrumented with two 6-axis force sensors. Reflexive markers are added for position analysis using the VICON optokinetic system (not evaluated in this paper).*

The objective of the current paper is to present a new methods for ambulatory measurement of ground reaction forces, using an instrumented shoe, and to present preliminary data demonstrating the feasibility of this new method. A more extensive presentation of the results of this study is in press [4].

2 Methods

2.1 Instrumented shoes

We propose to measure GRF using two six degrees of freedom force and moment sensors under each foot, one under the heel and one under the forefoot (figure 1). This arrangement allows for the flexion of the sole of the shoe during push off in gait, resulting in normal to near-normal gait. In principle, one sensor per shoe could be sufficient, but the construction of a shoe that would allow normal gait would be difficult. The specifications of the applied sensor (ATI-Mini45-SI-580-20, supplier: Schunk, Arnhem, NL) are given in table 1.

Table 1. Main specifications of the 6 DOF force and moment sensor (ATI FTD-Mini45-SI-580-20) used for ambulatory GRF measurement.

Sensor parameter	**value**
Diameter	45 mm
Thickness	15.7 mm
mass	92 g
Vertical force range	± 1160 N
Horizontal force range	± 580 N
Moment range (in all 3 directions)	± 20 Nm

2.2 Preliminary experimental evaluation

A healthy subject participated in a preliminary experiment. He wore an instrumented shoe on the right foot containing two 6 DOF force and moment sensors (figure 2). The left shoe was equipped with dummies instead of actual sensors. The subject repeatedly walked over an AMTI force plate in the movement analysis laboratory of the University of Twente. The trial consisted of three or four strides, of which one step of the right foot was on the force plate. The ground reaction forces were calculated from the signals of both the force plate and the instrumented shoe. The estimates of the modulus of the ground reaction force were compared by determining their RMS difference. In addition, the center of pressure trajectories were estimated from the measured signals of both systems. The RMS distance of both trajectories after optimal alignments were evaluated during the right stance phase.

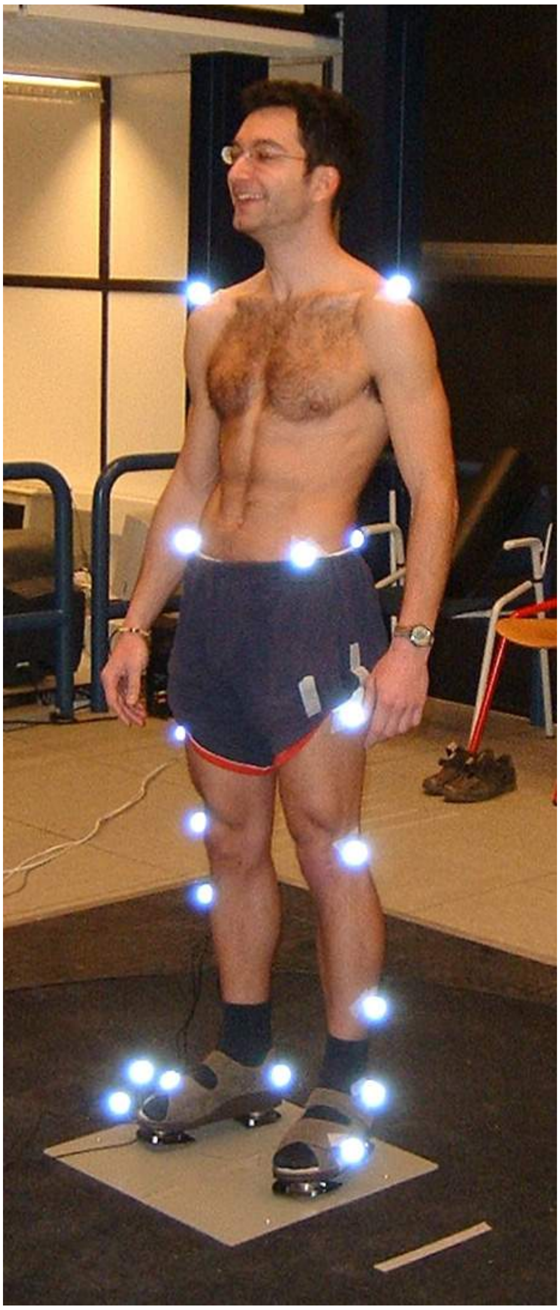


Figure 2. *Subject standing on instrumented shoes. Only the right shoe is instrumented with two 6 DOF forces, the left* shoe contains two dummies.

3 Results

The GRF measurement of the sensor system matches well with the GRF measured with the force plate. This is illustrated in figure 3. This figure shows the modulus of the heel and forefoot sensors and the modulus of the sum of both forces during a gait trail with a healthy volunteer. When comparing this sum to the modulus of the force plate GRF measurement, a good match was found (figure 3). . The RMS difference of the magnitude of the ground reaction force over 12 gait trials was 15 ± 2 N, corresponding to 1.9 ± 0.3 % of the maximum ground reaction force magnitude. The RMS difference of the horizontal component of the ground reaction force was 3 ± 2 N, corresponding to 0.4 ± 0.2 % of the maximum ground reaction force magnitude and to 2 ± 1 % of the maximum of the horizontal component of the ground reaction force. The RMS distance between both CoP recordings is 2.9 ± 0.4 mm, corresponding to 1.1 ± 0.2 % of the length of the shoe, when the trajectories are optimally aligned.

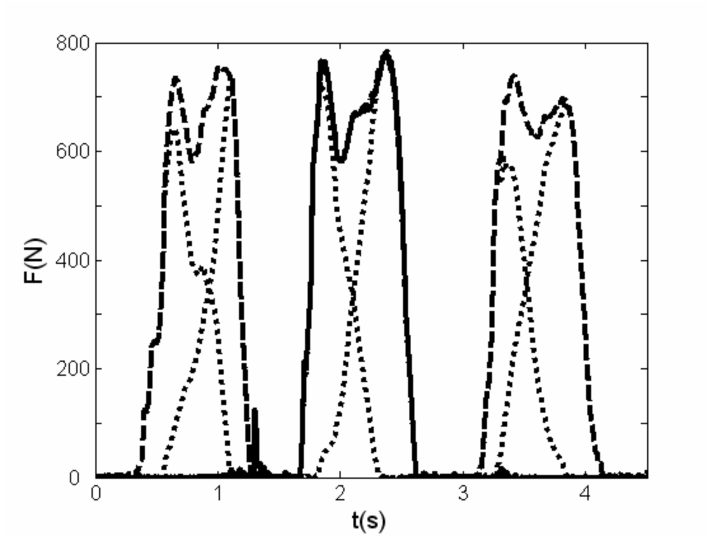


Figure 3. *modulus of forces measured during gait.*
The second step is on the forceplate: Modulus of force measured with the heel and forefoot sensors (dotted), modulus of sum of force (dashed line), modulus of force measured with force plate (solid line).

4 Discussion

It should be noted that the force, measured in sensor coordinates was directly used to constitute the modulus of the sum of heel and forefoot force. During the start and end of the step, the sensor may rotate with respect to the ground. Apparently, this does not cause measurable errors in the presented analysis, indicating that the implicit assumption that the sensors are flat on the ground when loaded is acceptable under the conditions of this measurement (see figure 3). It should be noted, however, that a correction for sensor orientation is certainly possible when this orientation is measured. In an ambulatory condition, this would require the use of ambulatory movement sensors [5-7]. In addition, it should be noted that separate GRF analysis is provided for each foot when instrumenting both feet.

5 Conclusions

Our study demonstrates the feasibility of ambulatory measurement of the ground reaction force under each foot by using two 6 DOF force sensors per shoe. This allows normal or near-normal gait, with the possibility to flex the sole of the shoe during push-off in gait. This development yields the possibility to perform full biomechanical analysis of body movements during daily-life activities at home and at work. In our future research we will further evaluate the accuracy of the estimate of modulus and orientation of the GRF and of the center of pressure, the influence of the instrumentation of the shoe on foot movement and the feasibility of using the instrumented shoes on patients with central neural disorders.

The financial support of this research by the Dutch ministry of Economic Affairs is gratefully acknowledged (ICT breakthrough project ExoZorg).

References:

1. Woodburn, J., 2003. The contribution of plantar pressure measurement to the understanding of foot structure and function in rheumatoid arthritis. in *Biomechanics of the lower limb in health, disease and rehabilitation*, Salford.
2. Forner Cordero, A., 2003. *Human gait, stumble and ... fall? - Mechanical limitations of the recovery from a stumble*, University of Twente: Enschede,.
3. Razian, M.A. and M.G. Pepper, 2003. Design, Development, and Characteristics of an In-Shoe Triaxial Pressure Measurement Transducer Utilizing a Single Element of Piezoelectric Copolymer Film. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **11(3),** 288-293.
4. Veltink, P.H., et al., 2005. Ambulatory measurement of ground reaction forces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, in press.
5. Luinge, H.J., 2002. *Inertial Sensing of Human Movement, in Electrical Engineering*, University of Twente: Enschede.
6. Luinge, H.J. and P.H. Veltink, 2004. Inclination measurement of human movement using a 3D accelerometer with autocalibration. *IEEE Trans. Neural Systems and Rehabilitation Engineering*, **12**, 112-121.
7. Veltink, P.H., et al., 2002. Three dimensional inertial sensing of foot movements for automatic tuning of a two-channel implantable drop-foot stimulator. Medical Engineering and Physics, 25, 21-28.

Seeing the complete picture: combination of telemetric monitoring and behavioral monitoring

K. Kramer
Vrije Universiteit, Amsterdam, The Netherlands

This SIG intends to bring together researchers interested in the combination of physiological data obtained via implantable radio-transmitters and obtained from behavioral observation tools. Three leading experienced researchers share the fruits of their years of laboratory experience working with fully implantable laboratory telemetry. Their presentations communicate the details of the telemetry applications and methods they have employed to dramatically reduce animal use, to improve data quality, to do better science, to maximize the amount of information gained from each experimental animal, and to reduce animal stress and discomfort. An emphasis is placed on the benefits of combining behavioral data obtained by different methods with physiological data.

The use of radio-telemetry in freely moving laboratory animals in combination with a behavior system

K. Kramer

Physiological parameters, such as body weight, hormonal levels in plasma and/or urine, heart rate, blood pressure and body temperature are often used for the assessment of the well-being of laboratory animals. When measuring physiological parameters as heart rate, blood pressure or body temperature in restraint small laboratory animals, the results must be interpreted with caution. E.g. blood pressure measurements in mice via tail-cuff technique cause an increase in both heart rate and blood pressure. In general, animals subjected to various laboratory procedures and environmental changes will react to novel situations with changes in their physiological responses.

Since radio-telemetry provides a way to obtain accurate and reliable physiological measurements from awake and freely moving animals in their own environment, it seems to be a valuable tool to assess animal welfare without the interference of, for instance, handling. Currently, radio-telemetry can be applied in all frequently used laboratory animal species. Implantable, miniaturized radio-telemetry devices enable the acquisition of non biased experimental data, lead to substantial reduction of the number of animals used and are often a refinement of the experiment. During this presentation the recent advances in the use of radio-telemetry in combination with a behavior system in the study of animal welfare in small laboratory animals, are discussed. Especially an emphasis is placed on the benefits of combining behavioral data obtained by the LABORAS behavior system with physiological data obtained by the DSI telemetry system.

Toward a novel approach for the simultaneous recording of behavioral and physiological data in rodents

B.A. Ellenbroek, D.J. Heeren and L.J. Lubbers

With the increase in techniques for altering the genetic make-up of mice (and more recently also rats) there is an increased need for developing sophisticated hardware and software that allow for the simultaneous analysis of a large number of different parameters in a relatively high throughput setup. Ideally such a system should use a standardized protocol, yet on the other hand allow for individual modifications necessary for specific research questions. Moreover, it should allow registration of both behavioral data using software analysis of specific behavioral elements, as well as the (telemetric) recording of physiological parameters such as heart rate, blood pressure and body temperature,; and these two data streams should be synchronized so that behavioral changes can be directly coupled to physiological changes. Finally, the system should be designed to register and monitor the behavior of several animals within the same cage, allowing the analysis of various kinds of social behavior, including play behavior, social interactions between conspecifics and sexual behavior.

We are presently evaluating and validating such a new approach for rats by using the Phenotyper, developed by Noldus B.V. and the PhysioLinQ developed by Telemetronics B.V. The Phenotyper allows for the continuous recording of behavioral data. These videostreams can then be analyzed using well known software programs such as EthoVision and/or the Observer. The PhysiolinQ represents a novel telemetric device that can record body temperature and heart rate. During the presentation we show the first results from our pharmacological experiments using singly housed rats, where alterations in body temperature and heart rate can directly be linked to changes in behavior. Moreover, we give a short demonstration of a social experiment, in which behavior, heart rate and body temperature during such a social encounter are registered and synchronized.

Although the validation of the entire system is still ongoing, the first results suggest that this represents a novel approach for the simultaneous recording of behavioral and physiological data which will become of crucial importance for the phenotypical analysis of small rodents such as rats and mice.

PhysioLinQ, a battery-less transmitter for remote measurement of physiological signals in socially housed rodents

G.J. van Essen and M.B. Jansen

An innovative telemetry system, PhysioLinQ has been developed by TeleMetronics biometry, a spin-off company of Wageningen University and Research Centre, The Netherlands. The system is based on proven technology; prior versions are in use since 12 years. On request of the European Space Agency (ESA) the system is modernized and miniaturized, by application of chip technology, for use in microgravity studies on mice onboard the International Space Station (ISS). The system is suited for monitoring heart rate, body temperature and animal activity, via acceleration, in group housed animals.

PhysioLinQ differs from other commercial telemetry systems: the signal transmission is not continuous but intermitted. This means that the collected physiological data is pre-processed on board of the implant. At predefined time intervals, for instance every minute, the accumulated data is transmitted to the receiver. This dramatically reduces the huge amount of redundant data that is normally produced by biomedical telemetry systems. Since every transmitter can be uniquely addressed by an identification code transmitted by the base station, a number of implants may be active at the same location. One base station is capable to monitor up to 20 animals simultaneously. This enables social interaction studies between animals.

A battery-less power system is developed to sustain long duration experiments as well to decrease the total volume of the implants. Successful in vivo experiments were carried out in mouse and rat. At the beginning of this year the system was accepted by ESA for use in the ISS.

PhysioLinQ benefits are: smallest implants, battery-less transmitters (no refurbishing), measurement on group-housed animals and integration with The Observer and EthoVision to analyze behavioral and physiological effects in social tests.

Possibilities and limits of EEG analysis in mouse models for epilepsy

H.E. Krestel

This talk discusses possibilities and limits of one and two channel EEG recordings and their analysis, exemplified by presenting data of the following mouse models for epilepsy. The first mouse model carries a point mutation in the channel pore of AMPA-type glutamate receptors (GluRs) that mediate the majority of fast excitatory synaptic transmission in the brain. These mice show myoclonic seizures, temporal lobe epilepsy and characteristic interictal activity (Krestel H.E. et al. 2004. J. Neurosci. 24: 10568). The second mouse model contains a point mutation in the ligand binding domain of GABA receptors that comprise the majority of fast inhibitory neurotransmission (GABA-A; submitted). This point mutation was first described in a family with absence epilepsy (Wallace, R.H. et al. 2001. Nat. Genet. 28: 49) and subsequently introduced into mice. However, genetically modified mice showed myoclonic jerks, comparable to mice with the GluR point mutation, but no tonic-clonic seizure attacks.

Third, mice deficient for a receptor subunit (GBR2) of GABA-B receptors that mediate slow inhibitory neurotransmission showed generalized tonic-clonic attacks without myoclonic jerks. EEG patterns and behavior of seizure attacks started and ended abruptly and interictal EEG showed predominantly polyspike activity (manuscript in preparation). No human epilepsy with a defect GBR2 subunit is known to date.

Technical and methodological possibilities and limits are discussed by data interpretation between these mouse models.

Telemetry monitoring: true physiological data from normal behaving animal models

E. Rieux

Implantable telemetry has become the gold standard to obtain physiological data from unstressed, free-moving animal models. A wide range of vital parameters such as Heart Rate, Blood Pressure, EEG, ECG, EMG, Temperature, Motor Activity, etc. can be monitored in animals ranging from small mice of 17 grams, to rats and similar size rodents and other larger animal models (dogs, primates, etc.).

In addition to physiological data, a number of environmental and behavioral data can be simultaneously collected: room conditions such as lighting and temperature, and events such as running wheel, drinking, and feeding activity. External stimuli (light, noise, air jet, etc.) can also be recorded together with the physiological data.

Physiological data obtained from implantable telemetry can also be imported into the Observer XT software. This provides a powerful tool to combine physiological and behavioral data.

This presentation gives an overview of current and future capabilities of telemetry monitoring, with a special focus on applications in behavior studies.

Autonomic function in mice during aversive emotional behavior tests

O. Stiedl and M. Meyer

The determination of autonomic function, i.e. sympatho-vagal interaction, has been introduced to supplement traditional behavioral measures during fear conditioning, an emotional learning paradigm in many species including mice. This paradigm is used to investigate the dynamics of heart rate and blood pressure in mice in order to determine three major aspects: (1) physiological adjustments, (2) pharmacological modulation, and (3) genetic approaches using mutant mice to investigate the contribution of specific neuropeptides to physiological cardiovascular conditions. The understanding of physiological characteristics is mandatory to draw proper conclusions as to the effects of pharmacological interventions on retention or expression of aversive memories. Furthermore, autonomic measures allow for stress-free behavioral tests in the home cage of mice when conventional behavioral parameters are unreliable. The normal heart rate displays complex fluctuations in time in response to environmental factors (breathing, activity, emotion), all of which are mediated by different central and peripheral feedback loops acting in parallel over a wide range of time scales. Measures of sympatho-vagal balance derived from non-linear analysis of heart rate variability reflect a highly dynamical function, the homeodynamics of which being accessible by advanced analytical techniques. Analysis of heart rate dynamics by techniques derived from statistical physics facilitates a translational approach from mouse to man. The methods are useful to uncover physiological principles and their pathological alterations, and hence, allow for assessment of cardiac risk in both man and mouse irrespective of species-specific differences in absolute heart rate. We hypothesize that (1) central autonomic dysregulation may contribute to elevated cardiac risk in the absence of genuine heart disease, and (2) autonomic dysregulation is potentially linked to the enhanced incidence of adverse cardiac events observed in epidemiological studies of psychopathological disorders in humans. With the implementation of autonomic measures, mouse models of chronic stress or depression will be investigated.

PhysioLinQ: Remote measurement in socially housed animals

G. van Essen, M.B. Jansen
Telemetronics Biometry BV, Wageningen, The Netherlands

Abstract
PhysioLinQ has been developed by Telemetronics Biometry as a device to measure heart rate, body temperature and activity in socially housed small animals.

Our product differs from other commercially available telemetry systems in some very important aspects

Signal transmission is not continuous, but intermittent: Collected physiological data are pre-processed within the implant and are sent at pre-defined time intervals. Since every transmitter can be uniquely addressed by an identification code, a number (up to 20) of implants may be active at the same location. This means that, contrary to other telemetry systems, PhysioLinQ can monitor social interaction between animals.

As PhysioLinQ uses inductive powering instead of batteries, it can be used for long duration experiments. Moreover, not using batteries makes a significant reduction of the volume of this telemetry system possible.

Promising results have been achieved with prototypes. PhysioLinQ will be introduced commercially at the end of 2006.

Keywords
Telemetry, socially housed animals, physiological body parameters, intermittent measurement, remote measurement.

1 Introduction
Major themes in designing biotelemetry implants are saving energy, reduction of volume and gaining biologically realistic data.

An implant that uses energy economically will be attractive for two very straightforward reasons: A given amount of energy can be used for gathering more data and there will be less stress for animal because replacing the energy source can be done less frequently. Most of the implants currently used have batteries as their source of energy. These batteries have to be replaced every 2-6 months.

Miniaturization of the implants is important because smaller implants open possibilities in a wider range of animals. Many commonly used implants can be used for rats and bigger animals, but turn out to be relatively big for use within mice.

From a biological point of view, the optimal environment for measuring behavior of animals would, of course, be their natural environment. Given the limitations of a laboratory setting, ways should be explored to find a realistic approximation. As, for instance, mice and rats are very social animals, the fact that many current biotelemetry systems can only monitor individual housed animals, seems to be a very severe limitation.

In developing PhysioLinQ we tried to find an optimal design that does justice to these themes.

2 PhysioLinQ Design
PhysioLinQ measures heart rate, body temperature and activity in socially housed animals. These are among the most frequently measured variables in animal research. Figure 1 shows a typical setup of the PhysioLinQ system.

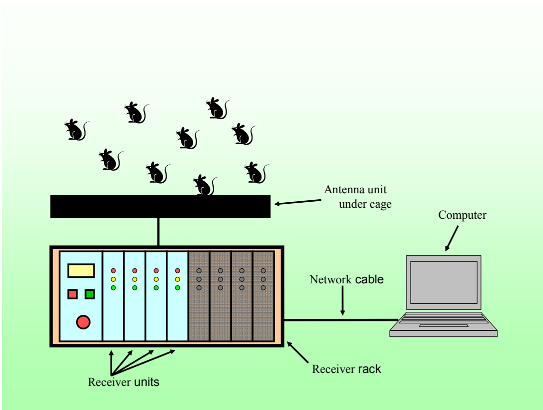


Figure 1. *Setup of a PhysioLinQ system.*

Several animals (mice, rats) in which PhysioLinQ is implanted, are housed together in a cage. This cage is placed on a 45x45 cm antenna unit that contains both an inductive powering system and a receiving antenna. The inductive field, generated by the powering system, serves as the energy source for the implants. The antenna unit is connected to a receiver rack. One or more receiver racks are connected to a (personal) computer via a TCP/IP network connection. This configuration makes it possible to place the receiver racks at the animal's location and the computer at a remote location.

Software, specifically designed for the PhysioLinQ system, stores the data. Data can be exported in CSV and native EXCEL format for further processing with commercially available software packages. As the software is designed to support TCP/IP, data can easily be made available via internet. This makes remote monitoring possible.

Each receiver rack can hold up to four receiver units. One or more receiver units can be connected to an antenna unit to facilitate synchronous measurements on a number of animals. Via a (TCP/IP) network, several receiver racks (up to five) can simultaneously be accessed. An example of a network setup is shown in Figure 2.

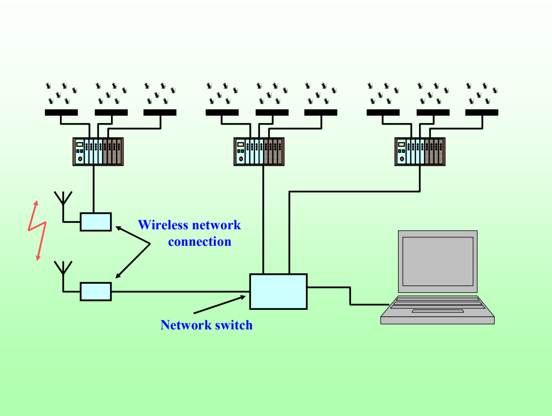


Figure 2. Possible PhysioLinQ network configuration.

Via the induction field, an identification code is intermittently sent to a specific implant. Upon decoding this call, the implant responds by sending its implant data. Each implant has a unique, 24 bit identification code. Implants can be called in any desired sequence. The period over which data will be measured and subsequently sent to the receiver, can be set in the software.

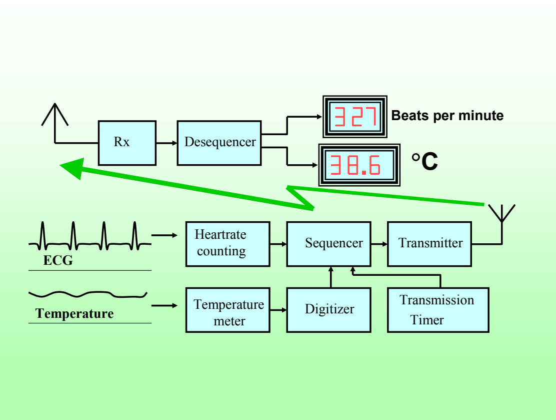


Figure 3. *Schematic representation of the intermittent transmission principle*

The intermittent character of the data transmission makes it possible to monitor more than one animal simultaneously with one receiver. Measurement data from one animal will be followed by data from a second animal and so on. Many implants can be monitored in the same available radio channel.

In addition to this simultaneous monitoring, it is also possible to make a set-up for synchronous measurement of instrumented animals. In this configuration each animal will be connected to one receiver unit. As the identification code for different animals can now be sent at exactly the same time, measurements over exactly the same period are possible.

2.1 Temperature

The temperature sensor circuit uses a resistive temperature sensing element (thermistor). Two reference resistors are added to the circuit with values that represent the lower and upper end of the temperature range to be measured. Because the values of these reference resistors are exactly known and all three values (of sensor and two resistors) are sent to the receiver system, a very accurate temperature measurement is possible. Absolute accuracy is +/- 0.1 °C and resolution is +/- 0.02 °C.

2.2 Heart Rate

The heart rate circuit gets its signal from two ECG electrodes that are placed along the electrical axis of the animal's heart (see Figure 4).

The ECG signal is picked up, amplified and filtered by a programmable band pass filter. The band pass filter suppresses all other signals such as muscle artifacts. Programmability is necessary because a QRS complex of ECG signals from different animal species differ considerably in width. So for each category of QRS complex the right frequency of the band pass filter is programmed.

The heart rate counter accumulates the received QRS complexes from the animal's heart. After a known time interval, the value is transmitted to the receiver and the counter is reset to zero. The counter value represents the average heart rate over the time interval in between two transmissions.

2.3 Activity

Animal activity is measured by using a piëzo electrical accelerometer. Output of the activity circuit is calculated as the number of times the acceleration signal amplitude exceeds a preset threshold level.

A small threshold level means that small movements of the animal will already generate sufficient activity counter values. A high threshold level means that only fast accelerations will be recorded. When it is expected that animals will be highly active, a prescaler can be switched on to reduce the activity movement counts.

The content of this counter is transmitted to the receiver system after a known time interval.

2.4 Dimensions

Figure 4 shows the PhysioLinQ implant in its current form.

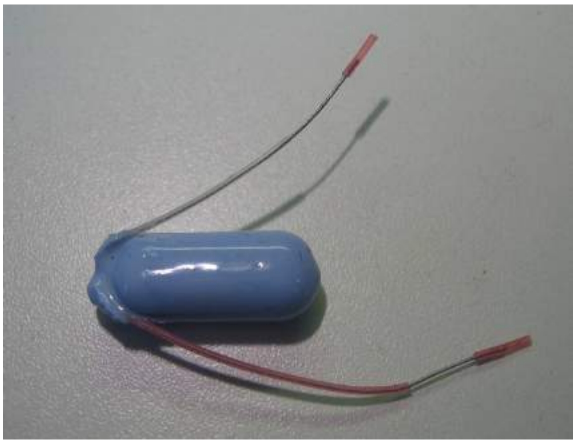


Figure 4. *The PhysioLinQ implant.*

The final version of a PhysioLinQ implant will have a volume of 1.0 ml and a weight of 1.5 gr.

3 Results and Discussion

At present, PhysioLinQ prototypes are being tested. Prior versions have been used for more than 12 years. In fact, they are still in use at the Rijksinstituut voor Volksgezondheid en Milieu (RIVM) in Bilthoven, The Netherlands. A prototype

of PhysioLinQ was successfully tested in February 2005 by the European Space Agency (ESA). PhysioLinQ will be incorporated in the Phenotyper, designed by Noldus Information Technology. Moreover, long-term duration studies are being set up. We expect to introduce PhysioLinQ in the fourth quarter of 2006.

Compared to other telemetry systems, PhysioLinQ has some clear advantages. The fact that no batteries are used means that the refurbishing problem does not exist. The inductive powering can in principle provide the implant with energy for many years.

Moreover, being battery-less, volume and weight of PhysioLinQ can be greatly reduced compared to other telemetric implants. A volume of 1.0 ml and a weight of 1.5 gram mean that PhysioLinQ not only can be used in rats, but also in animals, like mice, that are smaller and more sensitive.

Most important, however, is that PhysioLinQ makes it possible to monitor socially grouped animals. The intermittent character of data transmission means that many implants can be monitored, also if only one radio channel is available. Comparable telemetry systems that make use of continuous transmission are severely limited in this respect, since if only one radio channel is available, only one implant can be monitored at the same location. This makes social behaviour studies almost impossible.

In developing PhysioLinQ, we tried to find an optimal design. PhysioLinQ is unique in combining a very economic use of energy, miniaturization and the possibility to monitor socially housed animals.

The development of PhysioLinQ has greatly benefited from the kind cooperation of both Noldus Information Technology and the European Space Agency (ESA).

Using Virtual Reality for skills training and performance measurement

J.R. Kuipers[1,2] and P.A. Wieringa[2]
[1] *Green Dino BV, Wageningen, The Netherlands*
[2] *Delft University of Technology, Delft, The Netherlands*

Nowadays, the use of Virtual Reality for training and educating people has become widespread. Using simulators for training has many possible advantages such as increased safety, reduced costs and the fact that a conditioned training environment can be used. Currently, most researchers in the field are focusing on improving simulator realism and corresponding hardware requirements. However, not much attention has been paid yet to the automation of the training process. The current generation of training simulators is not sophisticated enough to fully replace a human instructor. It is hypothesized that a computer is better able to train a human being than a human instructor would, as a computer can objectively measure performance and automatically keep track of the student's skills on multiple tasks. Currently Delft University of Technology and Green Dino are developing a so-called Virtual Assistant (VA). The VA analyses human behavior and its operating performance (1. analysis), adapts to the student (2. adaptation), and efficiently communicates with the student (3. communication). Analyzing human behavior deals with measuring and judging performance from event-related data and continuous signals. Generic methods are being developed to quantify human performance from the complexity and frequency of movements. Adaptation deals with creating a student profile according to the student's short-term and long-term goals, needs and capabilities, such that optimal training is supplied. Communication deals with improving efficiency of communication by possibly using other modalities such as visual and haptic displays, rather than speech alone as used by human instructors. In the future, the VA should be able to fully replace a human instructor, and to supply more efficient and more effective training.

Modeling human decision making associated with discrete and continuous tasks

M.E. Vriezen, P.A. Wieringa and F. Vanderhaegen

The goal of this research is to develop a format for obtaining data to related to the human decision making. The operators perform a monitoring task of a complex system which includes discrete and continuous control tasks.
The basic idea is that the decision making is mainly based on personal weighting of the costs and benefits of the execution (model developed at the University of Valenciennes). The model consists of two main parts: the criteria definition and the decision making. The first part defines the increase of certain criteria (e.g. safety, productivity and workload) as a function of a situation S and a certain action A which can be performed in that situation. The second part of the model calculates a so-called preference for each possible action (Act*), as function of the situation and based on the differences between the criteria (i.e. ΔSafety(S,A)). These preferences are compared with the actual performed action (Act).
Data to fill out the model are collected during experiments with 9 participants. Data are recorded while a subject controls the a simulation program called NewTranspall. Data are collected in four 15-minutes sessions (two sessions on a complex system and two sessions on a less complex system). The first dataset on each system is used to fit the decision making of the model.
The subject based data are collected during a verbal protocol in which a human performs a rating scale questionnaire on the different criteria. A rules set is formed and validated which is the base of the criteria definition.
Preliminary conclusions are that a) the set of rules is probably sufficient, though not optimal, accurate to serve as input for the human decision making, b) The exponential trade-off has better performance modeling the decision making, than the fuzzy model using 1 antecedent, 1 consequent rules. This could be due to the fact that fuzzy model does not explicitly introduce trade-offs between the different criteria.

Behavior analysis of driving simulator students

J.R. Kuipers

The department Man-Machine Systems of the faculty Mechanical Engineering and Marine Technology started a cooperation with the company Green Dino Virtual Realities for research in the field of adaptive training systems. One of the topics is automatic instruction for driving simulators.
Green Dino Virtual Realities developed a sophisticated low cost driving simulator with an adaptive virtual instructor. Goal of the research is to improve the intelligence of the total system and the virtual instructor in specific. To measure the performance increase of the systems intelligence after adjustments we started a 0-measurement of the behavior of driving students. At the start the driving performance data of nearly 4000 students was available. The data analysis focused on the performance of driving students on driving tasks like not exceeding road speed, taking curves, approaching intersections etc. In the analysis the differences between genders got special attention. The results of gender differences in performance on driving tasks by students were compared with the differences in the performance of male and female experienced car drivers. The outcome showed a strong equality between the behaviors of drivers in the simulator as on the road. For instance men perform better on driving task like keeping lane and steering but women perform better on driving tasks like not exceeding the maximum speed and keeping a minimum following distance. These results show that the used simulator is valid alternative for measurement of driver behavior on specific driving tasks. Further more it gives more inside

information on the performance difference between genders. The results will be used to improve the curriculum and to personalize the instruction regarding to gender differences for example.

Radiotelephony understandability verification using a program distributed via Internet

P.J. Hoogeboom and G.D.R. Zon

In the continuous strive for increased flight safety levels, an experiment to verify the understandability of aeronautical radiotelephony messages has been conducted. The problem addressed concerns the pronunciation of the one hundred number either as "one-zero-zero" (International Civil Aviation Organization - ICAO - standard) or as "one hundred".

The main problem for this experiment concerns obtaining a sufficiently large number of pilots with different linguistic backgrounds. Pilots are scattered around Europe (or the globe). To overcome the associated financial burdens for travel, a dedicated PC program has been created which guides the pilots through the experiment. The program can be downloaded from the Internet and allows pilots to perform the experiment at their own pace, in their favorite environment and at a moment when they have time. From an experimental point of view, advantages are that participants can conduct the experiment in parallel, reducing the risk that they exchange information, and the reduction of the total experiment duration (multiple participants).

The program consists of an explanatory session, a training session including actual performance feedback, the main test with two sessions including rating forms (e.g. NASA TLX), and two questionnaires (biographic and topic related).

The participant needs to perform a domain relevant flying task using a simplified autopilot. Each time the participant receives an audio message from the simulated controller (10 different voices) directed to the own aircraft, the cleared values like heading, speed, altitude and communication frequency have to be selected using a graphically displayed keyboard. The flight status is displayed on a Primary Flight Display. A secondary task - analogous to the monitoring task from the NASA Multi Attribute Task Battery (MATB) - is added to ensure a workload level in accordance with what is experienced during normal flight.

The current paper includes the description of the program, some of the obtained experimental results, and the lessons learned with the set-up of this experiment like program creation, randomization of experimental conditions, recruiting participants, and the analysis of the reactions.

Allowance of larger inland waterway vessels on the river IJssel: Measuring workload and performance in a simulator

F.S.H. Verkerk

MARIN's nautical centre MSCN is specialized in nautical simulations. The activities of the centre can be divided in three items: the safety of ports and fairways, training of nautical personnel and the development of nautical simulators. MSCN has two full mission ship maneuvering simulators which are used for both training and research projects.

Design studies for ports and fairways are executed both for sea going vessels and for inland waterways. In most studies a first evaluation of the proposed fairway lay-out is executed with a fast time simulation model. In such a model the vessel is steered by a track following auto-pilot. In the finale phase of a design project simulations are executed on the real time simulator. The reason for simulations on the real time simulator is to take into account the effect of the human navigator. Normally the analysis of this type of simulations focuses on controllability of the vessel and the required maneuvering space. On behalf of the Dutch Ministry of Transport a study was executed to determine whether larger inland waterway vessel can be allowed on the River IJssel. Special concern was the passage of a series of sharp bends, not only due to the relatively narrow passage but also the workload of the skippers during this relatively long passage was of concern. For this reason a dedicated simulator study was set up together with TNO Human Factors. Experienced skippers executed simulator trials and at the same time their workload was measured. The study gave insight in the nautical feasibility of the maneuvers and on the safety of the passage. Recommendations were made how traffic regulations on the river can be adapted.

Tools for teaching observational methods

M. Roosjen and H. Theuws
Noldus Information Technology bv, Wageningen, The Netherlands

Researchers are aware of the many advantages technology has to offer in terms of increased efficiency and accuracy of data collection and analysis. Many behavioral researchers are also involved in teaching observational methods classes and try to improve the quality of teaching as well as to share this knowledge with their students.
This has lead to novel approaches, using multimedia techniques. Although the skills of systematic observation and assessment of behavior are central components of many undergraduate courses, they are becoming increasingly difficult to teach effectively using traditional approaches given the increases in student/trainee numbers.
This SIG meeting intends to bring together teachers of observational classes who share an interest in methods, techniques and tools for the study of behavior and to discuss the many difficulties that you will come across.

fOCUS: a tool for teaching observational skills to undergraduate psychology and postgraduate clinical psychology students

J. Oates

fOCUS is a generic multimedia application developed by the Open University and the BBC in collaboration with PSYCLE Interactive (www.psycle.com). It provides an innovative, multimedia approach to observation skills training and includes functions for observing, annotating, coding and analyzing digitized video and audio. Linked hypertext supports students' learning and the software functions. The media content can be easily versioned, making it a flexible, adaptable resource. fOCUS is already extensively used by the Open University to support undergraduate and postgraduate courses and won the European Academic Software Award 2000 for the social and behavioral sciences.
fOCUS has been further developed within a three year project funded by the Higher Education Funding Council for England (HEFCE) and the Department for Employment and Learning (DEL). The aim of the Observation Skills in Psychology project is to improve the quality and cost-effectiveness of training in systematic observation and assessment skills in undergraduate psychology and postgraduate clinical psychology courses. It is a consortium partner project between Leeds Metropolitan University, the Universities of Newcastle, Oxford and Sheffield, led by the Open University.
Learning units have been developed on coding schemes, reliability and validity, opening and closing clinical psychology sessions and taking clinical notes. These units, along with further developments of the fOCUS software, have been through two phases of field piloting and will be finalized during 2005. The learning units are designed to be stand-alone, and each can be completed within a 2-3 study session.
This tutorial explains the development of these learning units, the philosophy behind them and will demonstrate them in action. A discussion is encouraged to explore further uses of the software and the approach to teaching and learning that we have been developing.

Introducing the analysis of sequences in behavior

D.W. Dickins

After suitable experience with behavioral coding the student of behavior needs to move on to consider the structure of behavior. Examples of the ways in which individual behaviors, or interactions between individuals, may progress through a series of stages need to be observed. The statistical nature of unraveling the interdependencies between the emission of certain categories of behavior has to be understood in all but the most obvious examples of behavior. Students should do this by hand using simple examples before attempting to use supportive programs such as that provided by The Observer.
Probable examples used here to demonstrate how students might learn to conduct lag sequential analysis will be:

- for individual behavior: self-grooming (cat, monkey, human); exploratory sequences (mice); skilled sequences (humans).
- for interactions between individuals: nest relief in kittiwakes; courtship sequences in fruit flies.

Clues that observational analysis may provide as to the causal mechanisms underlying sequential structures in behavior are considered. Mention is made of temporal patterns in behavior that may only be discerned using more sophisticated algorithms, such as used in 'Theme'.

Distance teaching: behavior

J. van Rhijn

Almost 20 years ago we started to make a course on behavioral biology for our University: a textbook with theory and questions, and a practical based on interactive video programs on four 30 centimeter laser vision disks. Dispersed over the country we had 18 Study Centers where students could watch these programs and make a record of their own observations based on this material. Of course, image quality was marvelous, but the technical options to make a record were rather poor. When computers acquired a position in almost every household and most of these computers could be used to view videos, we decided that students should be able to do all the work at home. The videos were transferred to CD-rom, but unfortunately, the interactive component was dropped. In the mean time the course has been completely rewritten. It consists now of the theory in a printed textbook, an electronic workbook on CD-rom and three projects in The Observer 5.0 Student Edition on a second CD-rom. The two electronic components greatly expanded the possibilities to interact with the material. For the examination, the student has to write a short paper about a small piece of research or has to review some literature on a certain subject. In addition he or she has to comment on the papers by two other students. The ingredients for these concluding tasks can be found in the electronic workbook and in the three The Observer projects. I demonstrate some of their features and potentialities.

fOCUS: a distinctive pedagogic approach for maximizing the teaching and learning value of analyzing digital video

J. Oates
Centre for Childhood, Development and Learning, The Open University, Milton Keynes, UK

Abstract

This short paper describes the development of the fOCUS software and teaching materials which make use of the concept of 'learning object' to successfully meet the needs of teachers of observation skills.

Supported by the British Higher Education Academy, the project Observation Skills in Psychology has produced sets of learning units for undergraduate psychology students and trainee clinical psychologists, based on extensive surveys of lecturers wishes and needs for such materials.

The next stage of development is towards a new version of the software and the extension of the learning materials into new fields including health professions training, business studies and clinical psychology supervision.

Keywords

Teaching observation skills software pedagogy.

1 Introduction

Digital video is a rich and potentially powerful teaching medium. It invites engagement with its dynamic content, it is information dense and it can stimulate emotional as well as intellectual responses. By bringing events into the learner's immediate environment, direct contact with otherwise inaccessible experiences is made possible. Many fields of study involve extracting information from dynamic, real-world events. Fields as varied as vulcanology, musicology, animal behavior, theatre studies and developmental psychology can usefully employ video for teaching and research. The accurate and systematic analysis of video material is a key research method. In teaching observational skills, achieving a close engagement of learners in critical and systematic approaches to video materials is a crucial factor.

1.1 Observation skills and psychology

The skills of systematic observation and assessment of behavior are generally recognized as central components of undergraduate research methods training in psychology and form a major element of the defined core competencies for clinical psychology training in Great Britain. They are widely valued in related professional fields including nursing, primary care, social work and teaching, and are important 'employability' skills for clinical practice and behavioral research. They form an important part of linking theory and practice, through the application of psychological principles to observation, analysis and assessment of behavior.

2 The fOCUS project in the Open University

Stemming from experience with using digital video as a raw data source in professional software tools, such as The Observer, for research in child development and using video as a medium for teaching, the fOCUS project has been developing a pedagogic tool for presenting and analyzing video material. An initial version, produced in 2000 in collaboration with the BBC, won the European Academic Software Award for the social and behavioral sciences and was subsequently used as the delivery medium for a complete CD of audiovisual materials for the Open University Exploring Psychology course, and then also for the Masters level course Child Development in Families, Schools and Society, delivering video content and linked textual materials enabling students to learn how to analyze and interpret child behavior.

A new version of the software shell, fOCUS II, has been developed as part of the project. This is being used in the Open University course Research with Children and Young People, and is being used for the 2006 and subsequent presentations of the psychology course Child Development.

3 The Observation Skills in Psychology project

Initial funding for the development of the first versions of fOCUS came from the Open University. In 1999, the Open University Psychology programme was rated as 'excellent' by the British Quality Assurance Agency for Higher Education and this, along with the recognition of the role of the innovative fOCUS software allowed a successful bid to be made to the Higher Education Funding Council for England. This granted €360k funding under the Fund for the Development of Teaching and Learning to disseminate and develop fOCUS for use in undergraduate psychology and postgraduate clinical psychology training courses in English HEIs. This project, named Observation Skills in Psychology, is now in its third year and so far has achieved take-up of the materials by more than 30 other British HEIs. A major update and redesign of the user interface is currently underway towards a fOCUS III that will take full account of the extensive user testing that has now taken place in a large number of universities.

3.1 The consortium

A consortium of universities, Leeds Metropolitan University and the Universities of Newcastle, Oxford and Sheffield, led by the Open University, was established to carry the project forward. Its key aim was 'to improve the quality and cost-effectiveness of training in systematic observation and assessment skills in undergraduate psychology and postgraduate clinical psychology'.

The specific objectives of the consortium, in collaboration with additional participating departments were to:

1. Identify user requirements and intended learning outcomes for training psychology undergraduates and

clinical psychology postgraduates in observation and assessment skills;
2. Produce two customized versions of existing CD-ROM based teaching materials meeting the identified requirements, one relating to observational methods of studying behavior for psychology undergraduate courses, the second relating to observation and assessment in clinical psychology courses;
3. Pilot and evaluate the resulting teaching materials in ten psychology departments;
4. modify the materials in the light of the evaluation;
5. Disseminate widely information about the developed teaching materials, and support and evaluate their introduction into teaching programmes;
6. support up to ten departments in developing and using further versions of the CD-ROM based teaching materials;
7. Identify and evaluate opportunities and solutions for meeting the specific needs of students with disabilities;
8. Conduct a broad dissemination program to support and extend the take-up of the final sets of teaching materials; and
9. Plan a strategy to establish ongoing support for departments using or wishing to make use of the teaching resources, and to foster further development activities.

3.2 Progress towards objectives

At the time of writing (July 2005) we have achieved objectives 1-6 and have made substantial progress towards achieving objectives 7-9.

3.3 Identifying sector needs

During March 2003, questionnaire surveys to identify user requirements for observation and assessment skills training were circulated to 118 undergraduate and 32 clinical psychology departments; 39 undergraduate and 16 clinical departments replied.

The data from these returns gave us three key sets of information about:

• required learning outcomes
• preferred video contents
• curriculum and course constraints

The survey exercise also led to expressions of interest from several universities wishing to pilot test the teaching materials. Workshop sessions were held with teachers from these universities to agree the detailed requirements that then guided the authoring of the learning materials. One of the main functions of these meetings was to discuss and agree the required learning outcomes.

4 The pedagogic approach

4.1 What teachers want

It became clear from the surveys and workshop sessions that there is a high level of interest in improving the teaching of observation methods, since this has become so difficult to provide in recent years. It was also evident that a very specific pedagogic approach was needed, for which the fOCUS software tool was well suited.

Rather than requiring whole courses or modules produced by other people, it was established that what university teachers most want is to be able to use high quality, self-contained short ‘nodules’ of learning materials that could be easily incorporated into their existing teaching timetables. Such ‘nodules’ would ideally be deliverable within single sessions of 2-3 hours, the length of time usually allocated for practical classes. Materials like this are then felt to be ‘owned’ by teachers and hence more easily integrated into their broader teaching aims.

Given the time pressures that university teachers are increasingly experiencing, they need to be assured that materials provided by other people ‘do what they say on the label’, in other words, that they deliver successfully to clearly specified learning outcomes. They also expect all the components of the learning ‘nodule’ to be provided, so that neither they nor students are faced with assembling disparate elements from a range of sources that may or may not fit well together. Finally, they expect such materials to be relatively independent of any required prior learning in students, so that a range of prior backgrounds and experiences can be accommodated.

4.2 The ‘learning object’ concept

Reflecting on these findings, we found a lot of conceptual overlap with the relatively new notion of ‘learning object’ (Wiley, 2002; Weller, Pegler & Mason, 2003), which is being developed to describe similar ‘nodules’ of electronic teaching assets which have characteristics such as:

• reusability in different learning contexts
• self-contained; all assets included
• can be aggregated into larger units of study
• uniquely identified elements

These four criteria have been advanced in relation to virtual learning environments, but we have found them useful in developing what we have come to call ‘learning units’ built within the fOCUS shell. These units consist of raw data in the form of digital video files and linked sets of HTML files forming a coherent teaching progression. These take students through a structured sequence of activities involving various levels and types of analyses of the video sequences, embedded within an explicit set of defined learning outcomes and assessments designed to relate to those outcomes.

These latter components, the explicit specification of the desired learning and providing tools for the assessment of its achievement by students are important additions that we have made to the learning object concept, in the light of our experience in developing distance teaching materials in the Open University and elsewhere.

4.3 Implementing the learning unit concept

Thus informed by the synthesis of our survey findings and the learning object notion, we produced four learning units, two for undergraduate psychology students, Coding and Reliability and Validity and two for clinical psychology postgraduate trainees; Clinical Note-taking and Opening and Closing Sessions.

Each unit was designed to occupy about 150 minutes study time. To give an example of the pedagogic content, the following are the intended learning outcomes of the Coding unit:

Once students have completed the unit they should be able to:

• list features of systematic, scientific observation
• use an existing coding scheme to record behavior
• compare and evaluate the features of time-and event-sampling
• describe the basic characteristics of coding categories
• evaluate a coding scheme
• suggest modifications to a coding scheme.

4.4 Trialling and developing the materials

These materials were then assembled into editions of the fOCUS software and bulk copied onto CD-ROM for piloting in the ten partner universities, in which the first phase of evaluation trials was run (fig. 1). Following the systematic collection of evaluation data from these trials, the learning units were revised and minor modifications made to the fOCUS shell in preparation for the second phase of testing, which involved a further ten partner universities, during which further evaluative data were gathered.

Figure 1. *Student trials at Oxford Brookes University.*

In all, around 1,000 students, trainees, tutors and technical staff participated in the trials and we are currently finalizing the finished versions of the learning units in the light of the feedback for wide distribution in autumn 2005.

5 Conclusion

We found that extending the learning object concept to include the clear specification and assessment of the intended learning outcomes led to the development of sets of learning units delivered using the fOCUS shell which have met with wide acceptance and take-up.

The evaluative feedback gained from the extensive user testing of the learning units has contributed to a revised specification for the software interface and a new ‘look and feel’ is being developed for future applications of the fOCUS shell.

All of the institutions who piloted the learning units intend to continue using the materials that we have developed and we have had approaches from several disciplines other than psychology to extend our work into their areas.

We have recently gained further funding from the Higher Education Funding Council to continue this productive work.

The extension project that we start in 2006 will seek to adapt the learning units for use in business studies, health professions training, trainee supervision in clinical psychology and students’ own dissertation research using observation methods.

Supported by the Higher Education Funding Council for England (HEFCE) and the Department for Employment and Learning (DEL) under the Fund for the Development of Teaching and Learning (Grant # 72/02)

References:

1. Weller, M., Pegler, C. and Mason R. (2003) *Working with Learning Objects – Some pedagogical suggestions*, Milton Keynes, The Open University.
2. Wiley, D.A. (ed) (2002) *The instructional use of learning objects, the agency for instructional technology and the association for educational communications and technology*, Bloomington, Indiana (content also accessible via WWW http://www.reusability.org/read/).

Barnacle settlement behavior measured by EthoVision 3.0

J.-P. Maréchal[1], S. Esteban[1], C. Hellio[2], A. S. Clare[1]
[1] *School of Marine Science & Technology, Newcastle University, Newcastle uponTyne, United Kingdom*
[2] *present address: MFS, Université du Sud Toulon-Var, Toulon, France*

Abstract

Most sessile marine benthic invertebrates have a complex life cycle comprising several planktonic larval stages. As these larvae have to return to a surface to complete their life cycle, it can be argued that settlement strategies are of primary importance. The settlement stage of barnacles is the cypris larva. During settlement, the cyprid shows two main behaviors: swimming and searching. The latter involves the use of the cyprids' ambulatory organs – the paired antennules – which it uses to 'walk' over the substratum testing it for a suitable place to settle. Searching behavior can be further subdivided into: a) 'inspection' where the cyprid may make frequent turns but there is little or no movement over the substratum; b) 'close searching' which involves exploration of the substratum, characterized by small steps with frequent changes of direction, punctuated by the antennules testing the substratum; and c) 'wide search' during which the cyprid traverses the substratum in a comparatively straight line with few stops. To examine these behaviors in more detail, we developed a video-tracking protocol for individual cyprids in sea water using EthoVision 3.0. The raw data acquired during the experiments were used to isolate and characterize each behavioral category. Rates of transition between behaviors and the percentage of time spent in each behavior were also determined.

Keywords

Barnacles, settlement behavior, video-tracking, cypris larvae, marine invertebrate larvae.

1 Introduction

Throughout the complex life cycle of many marine benthic invertebrates a crucial stage occurs at the transition between planktonic life and benthic recruitment. At this precise moment the larvae need to find a suitable surface to settle on so that they can metamorphose into the juvenile. These pre-settlement processes are not very well understood in terms of behavior. However, it has been stated that behavioral changes at the small scale could have important effects on the organization of benthic communities [1].

Most marine larvae are small (<1 mm) and can swim very fast, making settlement behavior difficult to measure. Different methods have been developed to track marine invertebrate larvae over surfaces [2,3,4,5] but only recently the video-tracking software EthoVision [6] has been used [7].

We used cypris larvae of the tropical / semi-tropical barnacle, *Balanus amphitrite*, as a model for the development of a video-tracking protocol. This species was selected because it is relatively easy to culture in the laboratory and some details of its settlement behavior have previously been described [8,9]. Cyprid behavior has been described in ethological terms, but no quantifiable parameters have been determined.

B. amphitrite is a target species for antifouling studies [10] as it is an economically important fouling species. Fouling poses many problems for artificial structures such as ship hulls, oil platforms, aquaculture cages and netting [11,12]. Current barnacle antifouling research aims to interfere with cyprid settlement but the only widely used antifouling test is based on larval settlement rate over a 24-h period. The efficacy of an antifouling compound or surface is then measured according to its ability to inhibit settlement but the assay provides no information on behavior unless the cyprids are observed.

Cyprid behavior can be divided, simplistically, into searching or swimming; the former behavior being displayed when cyprids contact a surface. Cyprids may also 'rest' on a surface by lying on their side. When the larvae display searching behavior, three stages can be discerned [8]: "inspection", during which the cyprids stick by their paired antennules and do not move from their location; "close search" when the cyprids explore a small area often sticking by one antennule and probing the substratum with the other; and "wide search" which is characterized by 'walking' over the surface, generally in relatively straight lines.

These stages are quite easy to distinguish by eye. The challenge has been first to develop a suitable tracking protocol for individual cyprids to identify those behaviors, and then to use the EthoVision raw data to extract and segment the behavior of single cyprids observed during the assay, primarily into searching and swimming events.

2 Materials and methods

2.1 Animals

Adult *B. amphitrite* were kept at 22°C in the laboratory and fed with *Artemia* sp. (daily diet, 7 nauplii.ml-1). Cyprids were obtained by a modification [14] of previously described methodology [10,13]. The larvae were stored in the fridge at 6°C for use in video tracking. Three-day-old (D3) cyprids were used to measure behavioral parameters. Two additional batches of cyprids were used to examine the effect of age on behavior. These larvae (about 2000 individuals) were maintained on a 12h light: 12h dark cycle in a plastic tank filled with aerated artificial sea water (ASW) at 28°C. The cyprids were prevented from settling by containing them within a bag of plankton netting (250 µm).

2.2 Video-tracking protocol

We used Ethovision 3.0 to track individual cyprids in 30 x 10 mm Petri dishes (Bibby Sterilin Cat No 121V) filled with 4 ml artificial sea water (ASW). Cyprids were filmed using a Sanyo digital color CCD camera (model VCC 6592P) with a Cosmicar-Pentax lens (model TS2V314E), connected to a PC. The camera was placed either 4 cm above the arena for a single Petri dish or 12.5 cm for 3 Petri dishes.

Several tests were conducted to optimize the video-tracking protocol: type of dish, ratio of larvae/arena, light, temperature, acclimation time, tracking duration and number of replicates [7]. We finally used 1 to 3 30 x 10 mm Petri dishes placed within a 90 x 15 mm dish containing freshwater. This double-dish arrangement drastically reduced the light refraction problems in the

Parameters	Unit	Total/Mean	Filter	Description
Distance moved	Cm	Total	Min. Dist. Mov.	Total distance moved over a 5 min track
Velocity	mm/s	Mean	Min. Dist. Mov.	Mean movement speed of a moving cyprid (Inspection events not included)
Turn Angle path	Deg	Mean	Min. Dist. Mov.	Mean Turn angle of path
Turn Angle	Deg	Mean	No filter	Mean Turn angle taking into account body movements
Angular Velocity path	deg/s	Mean	Min. Dist. Mov.	Mean angular velocity of path
Angular velocity	deg/s	Mean	No filter	Mean Angular velocity taking into account body movements
Meander	deg/100µm	Mean	No filter	Indicator of Inspection behavior
Movement (moving and not moving)	%	Total	Filter: 0.001 + Start and stop velocity thresholds	% of total time swimming during moving

Table 1. List of parameters used to describe behavior of *Balanus amphitrite* cypris larvae measured by video-tracking. A downsampling (level 12) was applied to the raw data before analysis. Min. Dist. Mov. = Minimum Distance Moved filter.

walls of single dishes. The dishes were then placed over a light box. The light source did not affect the temperature of the test chamber which remained at 25˚C during the 5-min assay.

We previously found that the direct transfer of cyprids from 6˚C to 25˚C has a profound effect on cyprids behavior [7]. Accordingly, cyprids were acclimated at room temperature for 1 h prior to tracking. We also applied a 2-min acclimation time after the cyprid was transferred to the arena to reduce the immediate stress caused by transport and pipetting. For larvae maintained at 28°C, acclimation to room temperature was not necessary.

Each track was recorded for 5 min. In order to account for individual cyprid behavior variability, a total of 30 replicates were performed for each experiment, 20 being the minimum required. A total of 150 min of behavioral data were then acquired for each video tracking run.

2.3 Ethovision set up

The 500 µm long cyprid was tracked in a 3.5 cm diameter Petri dish. This means that the larvae appear as small black dots in the dish. We used the high acquisition resolution (768 x 576 pixels) and the grey scaling from the detection method in EthoVision. Cyprids size was a problem for visualization as we had to use a minimum object size of 1 pixel. Thus, any small object even if invisible to the eye, was detected by EthoVision. We therefore applied a dilation filter and added a 1 pixel layer to the cyprid's body. Using this filter we could set up a minimum object size detected of 8 pixels and a maximum of 50 pixels. The subtraction method with "object darker than the background" was used. A -15 to -20 low threshold was applied to remove noise. We finally used the scan window (50 pixels) to reduce problems with reflections.

The body of the cyprid is translucent in places which created contrast variation that could be interpreted as body position changes. Body size could also change very quickly if cyprids switched from swimming to searching behavior and *vice versa*. Consequently the size of the object detected varied greatly during tracking, rendering the surface area parameter unusable for further analysis.

2.4 Parameters used to describe cyprid behavior

We used the built-in statistic module of EthoVision to calculate parameters for cyprid behavior (Table 1). Recording at the sampling rate of 25 samples/sec gave a total of 7500 points over a 5-min track. We found that at this rate, the Total Distance moved of a non moving cyprid could reach more than 40cm. We then had to define filters to reduce the noise caused by cyprid body movements, particularly during inspection behavior, which created a virtual distance moved. For this reason, we downsampled the files to 625 points, i.e. every 12th point was sampled (applied to all parameters). Even with this down sampling, cyprids which did not move over the substratum were measured as doing so. Therefore a Minimum Distance Moved filter was applied. This filter was determined as previously described in [7]. It was not applied to all parameters so that we could take into account the effect of cyprid body movements.

2.5 Detailed description of cyprid behavior

We recorded 20 x 5 min movies of single cyprids in 3.5 cm Petri dishes with a magnification that provided a large arena for cyprid behavior and for the behavior to be characterized by an observer. These movies were used in EthoVision for tracking so that each behavioral event could be described by the parameters selected above. We could then characterize the stages described in reference [8] with quantifiable parameters.

3 Results

3.1 Detailed behavior description

Our objective was to characterize cyprid behavior using EthoVisions' parameters available in the data analysis module. Ninety four events were analyzed from 19 movies (95 min) with a duration of 0.56 s (inspection) to 300 s (swimming / inspection). No differences were observed between % behavior calculated from the raw data and from the downsampled data (Figure 1).

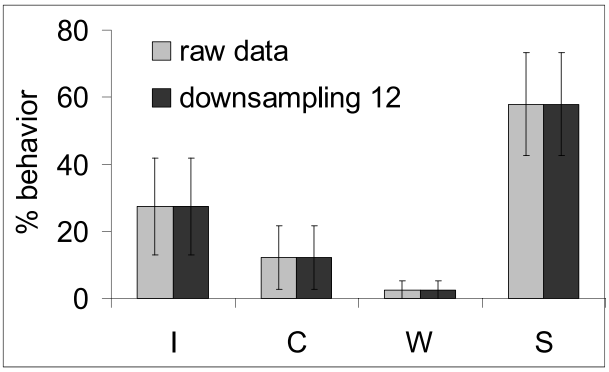


Figure 1. *Mean percent behavior of 19 single cyprids, calculated from 5min video-tracking recordings (total: 95min). I: inspection, C: close search, W: wide search, S: swimming.*

	Mean Distance moved between 2 points (µm)	Velocity (mm/s)	Turn angle path (deg)	Turn angle (deg)	Angular velocity path (deg/s)	Angular velocity (deg/s)	Meander (deg/100µm)
Inspection	4.2 +/- 5.9	0.02 +/- 0.07	34 +/- 62	117 +/- 12	8.2 +/- 19.6	201.8 +/- 22.8	139.1 +/- 30.3
Searching	19.7 +/- 6.9	0.07 +/- 0.05	71 +/- 39	109 +/- 11	13.9 +/- 15	197.9 +/- 24.9	119.4 +/- 26.8
Swimming	277.3 +/- 215.6	1.2 +/- 0.6	79 +/-31	87 +/- 21	93.6 +/- 50.8	152.8 +/- 35.2	77.3 +/- 37.4

Table 2. *Detailed behavior parameters calculated with the EthoVision analysis module (Mean +/- Stdev).*

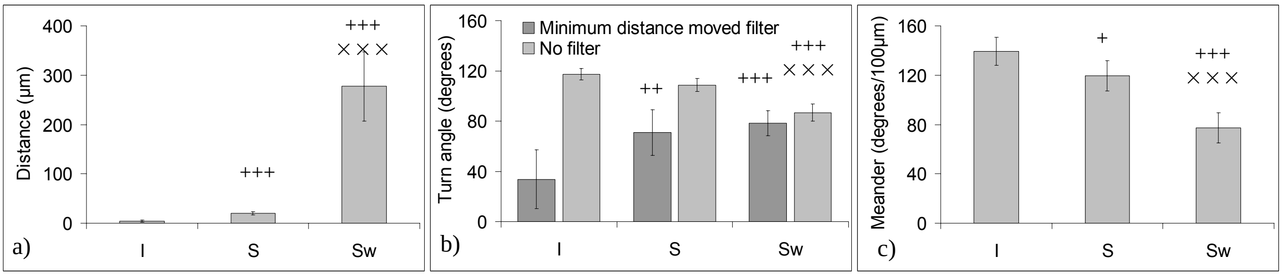


Figure 3. *Detailed behavior parameters calculated with the EthoVision data analysis module. +=compared to inspection, ×=compared to searching +++: P<0.001, ++: P<0.01, +: P<0.05, ×××: P<0.001. I: inspection, S: searching, Sw: swimming.*

Five events of 'wide search' were differentiated accounting for 2.5 % of behavior. Due to the low incidence of this behavior, 'close search' and 'wide search' were grouped together as 'searching'. We finally defined three categories: swimming (58 % of total behavior, 40 events), inspection (27.3 %, 32 events) and searching (14.7 %, 18 events). Inspection events were observed in almost all of the movies.
We calculated the percentage of each behavioral transition, e.g. from inspection to swimming, from all the transitions observed (Figure 2). The main switch was between 'swimming' and 'inspection' followed by 'close search' to 'swimming', 'inspection' to 'close search' and 'swimming' to 'close search'. All the other transitions were less than 3 % each.

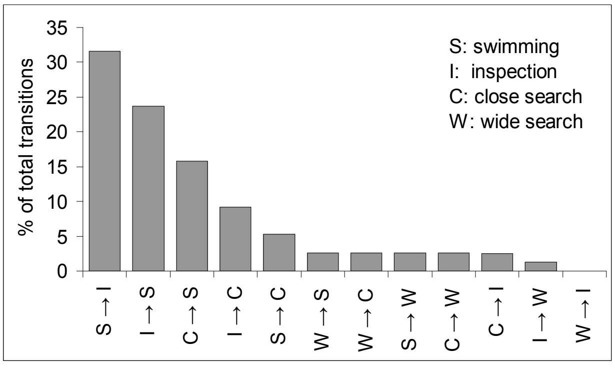


Figure 2. *The proportion of each behavioral transition calculated from all the transitions observed in the 19 movies analyzed.*

The values of the parameters calculated for each stage are presented in Table 2. 'Swimming' and 'searching' differed significantly for all parameters. The distinction between 'inspection' and 'searching' was not always significant. 'Close search' behavior is a mix between 'wide search' and 'inspection' and thus difficult to distinguish according to the parameter values calculated.
Distance moved (Figure 3a): the mean distance moved was calculated between 2 points for each of the 3 defined categories. According to the filters used, the distance was very close to 0 µm for 'inspection', which accurately reflected observations and accords with the removal of points associated with body movements. The mean distance moved during searching was close to 20µm whereas almost half a cyprid body length was moved while swimming (270 µm). Stop events during swimming and fast acceleration were responsible for the large standard error.
Velocity: velocities followed the same trend as the mean distance moved, with low speed during searching and about 2 cyprid body lengths/second during swimming.
Turn angle (Figure 3b): The application of the filters to the turn angle measure reversed the trend in the results. With the filters applied, the mean turn angle of the total tracking path was given, while with no filters, the turn angle associated with cyprid body movements was included and thus the value was much higher. This effect did not apply to swimming behavior where body movements are negligible.
Angular velocity: the speed of turning was higher during swimming as measured with filters applied. When no filters applied, the body movements masked the path mean value during searching compared to inspection. Nevertheless, a significant difference was found between these two behaviors with filters applied.
Meander (Figure 3c): we have previously shown that meander is a good indicator of inspection and searching as the number of deg/100 µm is very high through the cyprid pivoting on one or both antennules and/or rotating on the spot. Meander values decreased between 'inspection' and 'swimming'.

3.2 cyprid ages

The behavior of cyprids of different age (D0, D3, D6, and D9) was recorded. The parameters used to compare behavior were the same as used previously (section 3.1) except that the 'distance moved' is here used for the total distance moved during the 5 min assay. A swimming index [7], which was derived from the EthoVision movement parameter, was also calculated. This parameter gave the percentage of time that the cyprid spent swimming while it was moving.
Behavior changed with age for two different batches of larvae. No statistical differences were found between these

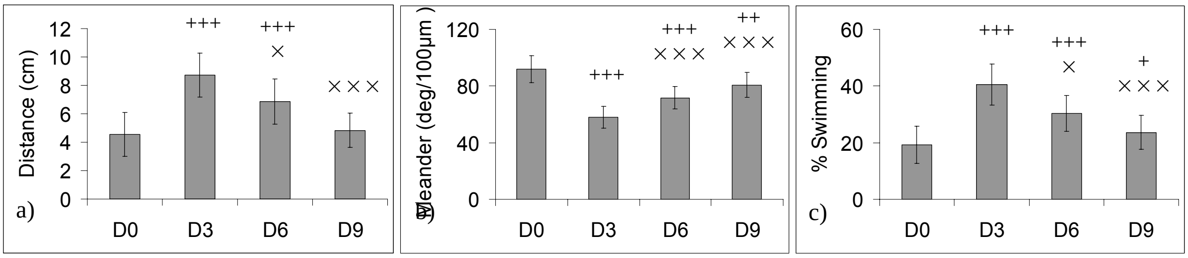


Figure 4. *Parameters calculated for cyprids aged D0 to D9 .Mean values for 30 replicates. +=compared to D0, ×=compared to D3, +++:P<0.001,+:P<0.05,×××: P<0.001.*

two batches. Similar parameter values were obtained for D0 and D9 cyprids, while an increase in activity was noted from D0 to D3 and a decrease from D3 through to D9. The most significant changes were between D0 and D3 cyprids, which gave the lowest and highest swimming values respectively (Figure 4c), with a corresponding increase in distance moved from 3 cm to 9 cm (Figure 4a). The meander value (Figure 4b) also dropped from around 110 deg for D0 cyprids to 60 deg for D3 cyprids. Differences between D3 and D9 cyprids were less marked, but still significant.

Comparisons between parameters calculated with or without applying filters were informative. The turn angle with filter applied (path) did not vary greatly which agrees well with visual examination of tracks obtained by EthoVision. However, when calculated from the raw downsampled data, the cyprid's body turn angle varied considerably with age, decreasing from D0 to D3 and increasing from D3 to D9, the latter being significantly different from D0. These observations were exactly the same for the angular velocity parameter.

4 Conclusion

EthoVision has been used to track cypris larvae in many different experiments and has proved capable of measuring swimming and searching behavior. Detailed behavioral analysis allowed us to gather essential information on quantitative parameters used to describe larval behavior.

These parameters are a valuable source of information as they can be used to measure differences in behavior in settlement experiments. The quantitative measures in this study were done for D3 cyprids. We have recently recorded detailed movies of D0, D3, D6 and D9 cyprids to further examine changes in behavioral parameters according to age, as significant differences in distance, meander and swimming were obtained from video tracking of differently aged cyprids. For example, young cyprids do not settle but are highly discriminatory with respect to the 'choice' of substratum. Our tracking results showed that D0 cyprids displayed a high level of searching behavior; swimming accounting for only 10% of the time spent moving. The high meander value also supports this assumption. D3 cyprids displayed a comparative increase in swimming behavior compared to D0 cyprids. This result was surprising as the settlement rate of cyprids rises as they age [15], at least until energy reserves are exhausted [16]. Swimming decreased in older cyprids, with D9 cyprids comparable to D0 cyprids in this regard although they were significantly different in other parameters such as meander. When correlated to settlement test results, we found that these cyprids, which were maintained at 28°C, settled at very low rates when D0 and D9, with the highest rate of settlement obtained for D6 cyprids. We found that switches from inspection to swimming and *vice versa* accounted for more than 50% of all the transitions observed for D3 cyprids. Swimming also represented 55% of all the behaviors observed. The decrease in swimming in D6 and D9 cyprids can be explained in terms of the desperate larvae hypothesis [17], whereby, as energy reserves deplete, larvae need to find a suitable place to settle before metamorphosis is compromised. However, we have to remember that all these results have been obtained under static conditions during a 5 min assay which does not reproduce the natural situation. Video-tracking of cypris larvae under flow is to be undertaken shortly.

Supported by awards from the Natural Environment Research Council (NER/B/S/2003/00273), the US Office of Naval Research (N00014-02-1-0311) and the European Union (CRAFT ENZYMEANTIFOULING G3ST-CT-2002-50300) to ASC.

References

1. Walters, L.J. (1992). Field settlement locations on subtidal marine hard substrata: is active larval exploration involved? *Limnology and Oceanography*, **37**(5), 1101-1107.
2. Hills, J.M.; Thomason, J.C.; Davis, H.; Köhler, J.; Millet, E. (2000). Exploratory behaviour of barnacle larvae in field conditions. *Biofouling*, **16**, 171-179.
3. Miron, G.; Walters, L.J.; Tremblay, R.; Bourget, E. (2000) Physiological condition and barnacle larval behaviour: a preliminary look at the relationship between TAG/DNA ratio and larval substratum exploration in *Balanus amphitrite. Marine Ecology Progress Series*, **198**, 303-310.
4. Walters, L.J.; Miron, G.; Bourget, E. (1999). Endoscopic observations of invertebrate larval substratum exploration and settlement. *Marine Ecology Progress Series*, **182**, 95-108.
5. Rittschof, D.; Sanford Branscomb, E.; Costlow, J.D. (1984). Settlement and behavior in relation to flow and surface in larval barnacles, *Balanus amphitrite* Darwin. *Journal of Experimental Marine Biology and Ecology*, **82**, 131-146.
6. Noldus, L.P.J.J.; Spink, A.J., Tegelenbosch, R.A.J. (2001). EthoVision: A versatile video tracking system for automation of behavioural experiments. *Behaviour Research Methods Instruments & Computers*, **33**(3), 398-414.
7. Maréchal, J.-P.; Hellio, C.; Sebire, M.; Clare, A.S. (2004). Settlement behaviour of marine invertebrate larvae measured by EthoVision 3.0. *Biofouling*, **20**(4-5), 211-217.

8. Crisp, D.J. (1976) Settlement responses in marine organisms. *In: Newell R C (ed) Adaptations to the Environment. Butterworth, London,* pp 83-124.
9. Lagersson, N.C.; Høeg, J.T. (2002). Settlement behaviour and antennulary biomechanics in cypris larvae of *Balanus amphitrite* (Crustacea: Thecostraca: Cirripedia). *Marine Biology,* **141**, 513-526.
10. Clare, A.S. (1996). Signal transduction in barnacle settlement: calcium re-visited. *Biofouling,* **10**, 141-159
11. Collinson, R.; Grant, G. (1994). An Australian shipping industry perspective. *In: Kjelleberg, S., Steinberg, P. 362 (Eds.), Biofouling: Problems and Solutions. The University of South Wales, Sydney,* pp. 14– 18.
12. Lewis, T. (1994). An Australian shipping industry perspective. *In: Kjelleberg, S., Steinberg, P. 362 (Eds.), Biofouling: Problems and Solutions. The University of South Wales, Sydney,* pp. 32– 38.
13. Rittschof, D.; Clare A.S.; Gerhart D.J.; Sister Avelin M.; Bonaventura J. (1992). Barnacle in vitro assays for biologically active substances: toxicity and settlement inhibition assays using mass cultured *Balanus amphitrite* Darwin. *Biofouling,* **6**, 115-122.
14. Hellio, C.; Simon-Colin, C.; Clare A.S.; Deslandes E. (2004). Isethionic acid and floridoside, isolated from the red alga, Grateloupia turuturu, inhibit settlement of Balanus amphitrite cypris larvae. *Biofouling,* **20**, in press.
15. Satuito, C.G.; Shimizu, K.; Natoyama, K.; Yamazaki, M.; Fusetani, N (1996). Age-related settlement success by cyprids of the barnacle *Balanus amphitrite*, with special reference to consumption of cyprid storage protein. *Marine Biology*, **127**, 125-130.
16. Harder, T.; Thiyagarajan, V.; Qian, P.Y. (2001). Combined effect of cyprid age and lipid content on larval attactment and metamorphosis of *Balanus amphitrite* Darwin. *Biofouling*, **17(4)**, 257-262.
17. Knight-Jones, E.W. (1953). Laboratory experiments on gregariousness during setting in *Balanus balanoides* and other barnacles. *Journal of Experimental Marine Biology* **30**: 584–599.

Investigating insect behavioral responses to insecticidal proteins

R. R. Binning[1], S. A. Lefko[2], F. Cheng[1] and K. Liao[2]
[1] *Pioneer Hi-Bred, Johnston, Iowa, United States*
[2] *Pioneer Hi-Bred, Newark, Delaware, United States*

Abstract
Worldwide in 2004, 15.6 million hectares of maize were protected from insect damage by plant expression of insecticidal proteins. This paper describes methods to measure the behavioral response of a target pest to a binary insecticidal protein in a quantitative, efficient, and statistically robust manner. Techniques were adapted from Strnad and Dunn (1990) for use with artificial insect diet and Ethovision®, a digital behavioral measurement tool (Noldus Information Technology).

Keywords
Diabrotica, corn rootworm, insect, Bt

1 Introduction
Building on the techniques of Strnad and Dunn (1990), Pioneer Hi-Bred developed an assay using Ethovision® (Noldus Information Technology) to observe and quantify search behavior of corn rootworm (CRW) larvae. The assay was designed to contrast the search behavior of CRW larvae after exposure to different food sources or hosts. The post-exposure search pattern is indicative of each larva's perception of host quality. The objective of the experiment was to describe behavioral changes resulting from larval exposure to the binary insecticidal proteins Cry34Ab1 and Cry35Ab1 (Cry34/Cry35), and to evaluate changes in behavior after exposure to Cry34/Cry35 at concentrations corresponding to those found in transgenic maize roots. The larval exposure scenarios included: no diet, healthy diet, diet containing Cry34/Cry35, unprotected maize roots and maize roots expressing Cry34/Cry35. A series of sensitivity analyses were used to: identify an optimal behavioral sampling strategy, identify behavioral parameters that best differentiated the negative and positive controls, describe optimal assay duration and to develop an experimental design that would help answer the research question.

1.1 Plant and Pest
Worldwide in 2004, 15.6 million hectares of maize were protected from insect damage by plant expression of insecticidal proteins from *Bacillus thuringiensis* (James, 2004). In some instances, plant protection may result from a combination of pest susceptibility and a delay in pest development resulting from insect exposure to Bt (Mohd-Salleh and Lewis, 1982; Gore et al, 2005). Measuring the contribution of both susceptibility and developmental delay is important to characterizing the mechanism of protection and considering the risk of insect resistance to the trait.

Western corn rootworm (*Diabrotica virgifera virgifera*, LeConte) is an insect pest of maize; the larvae damage maize roots and cost growers $1 billion per year in control costs and crop losses in the United States (Metcalf, 1986). Pioneer Hi-Bred International and Dow AgroSciences LLC have developed a plant-genetic solution for minimizing this loss; the protection mechanism is plant expression of the binary toxin resulting from Cry34/Cry35 insecticidal proteins.

Early observations of whole plant laboratory assays indicated differences in the feeding behaviors of western corn rootworm larvae exposed to maize seedlings with roots expressing Cry34/Cry35 and nontransgenic or control seedlings. These observations lead to an experiment to quantify the extent of a rootworm behavioral response to the Cry34/Cry35 insecticidal proteins.

1.2 Strnad & Dunn
Strnad and Dunn (1990) described search behaviors of larval rootworm associated with host quality. By exposing neonate rootworms to different plant roots, they were able to elicit and describe three distinct search patterns. Ranging is the normal behavior when no host is present; this search behavior is used to locate patches of quality hosts as quickly as possible. Strnad and Dunn presented results indicating ranging would continue after exposure to a plant that was of poor host quality to the rootworm. Local search is a change in behavior that results from contact with a good quality host. Once a food patch is located using the ranging behavior, local search is used to stay within the food patch. Local search is characterized by a decreased search area and velocity, and an increased number of turns and path crossings when compared with ranging. The third search pattern described was a mixture of ranging and local; this behavior was seen when rootworm larvae were exposed to plant roots that contained both a feeding stimulant and a feeding deterrent; search area and velocity decreased, but number of turns and path crossings stayed the same when compared to ranging.

2 Materials and Methods

2.1 Behavioral Assay Design

The assay was developed by identifying an optimal behavioral sampling strategy, identifying behavioral parameters that best differentiated the negative and positive controls, describing optimal assay duration and developing an experimental design that would help answer the research question.

Sampling Strategy - The Ethovision system for behavioral data collection is very versatile, and adjustments in the method of data collection were necessary to meet the objective of this study. The system collects a sequence of data points (X,Y coordinate information with an accompanying time stamp) and the usefulness of parameters that incorporate larval velocity are sensitive to sampling rate. More frequent sampling rates created artificial, usually multimodal, distributions that did not reflect true patterns of larval behavior. These distributions became more normal at less frequent sampling rates, and 6 samples per second was identified as optimal.

Parameter Identification – Nine potential behavioral parameters were identified using refereed literature and information from visualizations of preliminary tracks (Figure 1). These parameters included measures of larval velocity, distance traveled, turn angle, angular velocity and other parameters that are a function of these measures. Final behavioral parameters needed to be unique, have a low sensitivity to outliers, and a high probability of correctly detecting differences among treatments (power). Estimates for all potential parameters were calculated using preliminary data from positive and negative control exposure scenarios. The list of potential parameters was reduced to two by eliminating parameters that were highly correlated, eliminating parameters that were very sensitive to outliers and prioritizing parameters based on power analyses. The two parameters included in the analysis were: square area (max X – min X) * (max Y – min Y) and number of path crossings (the number of separate occurrences where one larva occupied the same X,Y coordinate during a track).

Optimizing Duration of Behavioral Data Collection - The efficiency of the assay and subsequent analysis increases with shorter tracking periods and consequently less data. Preliminary 60-minute tracks were used to identify optimal track duration. Pairwise comparisons of different time intervals revealed the assay and analysis of parameters was sensitive to duration. Tracks longer than five minutes were necessary to collect enough data to differentiate positive and negative controls. However, there was a diminishing return on precision with longer tracks, and longer tracks increased the probably of outliers. Ultimately, an analysis including the first 15 minutes of each track was identified as optimal for these assay conditions.

Sample Size – Preliminary tracks using positive and negative control exposure scenarios and the square area and number of path crossings parameters were used to estimate an appropriate number of replications. Using only the first 15 minutes of each track and a sample size of 35 replications for each treatment group, analyses of variance would have approximately 80% power to detect a true difference among treatments at the 0.05 level of significance. Thirty-five replications of each exposure scenario were used in the experimental design.

2.2 Experimental Design

To characterize a change in search behavior, rootworm larvae were exposed to various concentrations of the Cry34/Cry35 protein. The protein was incorporated into an artificial diet that western corn rootworm larvae perceive as a host. To further validate the test system and provide a more realistic exposure scenario, rootworms were exposed to maize roots.

The experiment used a randomized complete block design. The negative controls (filter paper), the positive controls (diet or root), and the diet or root treatments were randomly arranged within the light-proof box, and each set of three exposure scenarios was replicated over time. For each replication, a new piece of diet or root section was used and filter paper was replaced.

Non-diapausing WCR neonates (less than 4 hours after hatch) were used for all experiments. Neonates were exposed to diet, root treatments or moistened filter paper for 5 minutes. After the 5 minutes of exposure, larvae were moved to 15x100mm Petri dishes for observation. Each Petri dish was filled halfway with agar. The agar was covered with a piece of black filter paper, which became evenly moist after contacting the agar. A light-proof box held the Petri dishes, the Ethovision video camera and two infrared LED lights to illuminate the experiments. Larvae moved about the Petri dishes for 30-60 minutes creating a stream of x and y coordinates, each with a time stamp (Figure 1). The sample rate was 6 samples per second. Behavioral parameters were calculated using the coordinates and time.

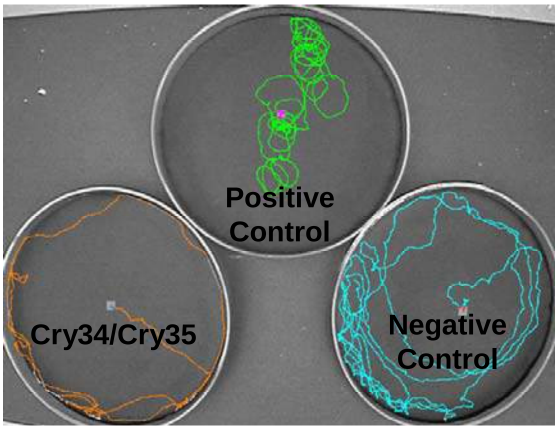


Figure 1. *A visualization three tracks representing each of the three treatments.*

2.2 Artificial Diet Exposure

The artificial diet was Bioserv southern corn rootworm larval diet (F9800B). Agar, ground wheatgerm, and sterile distilled water were added to the diet.

The diet exposures were a positive control using healthy diet, a negative control with no host and diet formulated with one of three different concentrations of Cry34/Cry35. The concentrations corresponded to the range of Cry34Ab1 and Cry35Ab1 expressed in roots. Five minutes provided adequate time for larvae to sample either diet.

2.3 Maize Root Exposure

Root exposures were a positive control using conventional maize roots, a negative control with no host and roots from a maize hybrid expressing Cry34/Cry35. Plants were grown to plant stage V7 in a Pioneer Hi-Bred greenhouse in Johnston, Iowa. One plant of each type was removed from its pot and the soil carefully removed from the roots. To expose the larvae, one root from the youngest node of each plant was trimmed about 7-8 cm from the growing point (root tip). Larval exposure to the positive control and the root treatment was achieved using the same Petri dish configuration used in the diet exposure. One larva was placed on each root tip, and five minutes of exposure provided adequate time for larvae to sample the roots. The negative control was a larva was placed on a moistened piece of filter paper for 5 minutes with no host stimulus.

3 Analysis

Several analysis methods were considered for this behavioral study. Simplicity and ease of communication were important factors in deciding on the appropriate analysis. To evaluate treatment effects on larval search behavior, analysis of variance (ANOVA) was performed on each parameter based on a normal theory mixed linear model. A number of different models for these data were considered. Examination of box plots and residual plots indicated that a normal theory model was appropriate for each parameter.

Behavioral parameters analyzed to date include number of path crossings and search area. Number of turns is another important parameter for differentiating local and ranging search behavior and work is ongoing to best incorporate it into this analysis.

4 Results

Overall, tracks from neonates exposed to plain diet had significantly more path crossings and a significantly smaller search area than tracks of neonates that were not exposed to a host (Figure 1). Tracks from neonates exposed to the highest concentration of Cry34/Cry35 give values that are in-between host and no-host tracks, without a significant difference from either.

Analysis is not complete for the experiment using exposure to maize roots; however, preliminary data indicate that the trends are similar to exposure to artificial diet.

5 Conclusions

Each parameter adds to the understanding of search behavior. As a result we are using multiple parameters and a weight of evidence approach to analyze the results. These results indicate that the test system and the parameters search area and path crossings are valid; negative control (ranging) and positive control (local search) are significantly different and the values for these parameters are intuitive.

This assay technique significantly decreases the input required to measure the behavioral response of an insect like western corn rootworm to an insecticidal protein and provides statistically robust quantitative results. This automated system takes minutes to achieve results comparable to a traditional assay that takes days of manual observations. Additionally, traditional assays typically do not provide the in-depth quantitative results that are produced using this assay technique. These methods may be applied to other insect species with results that contribute to target-pest resistance studies as well as non-target risk assessments. Using artificial diet for the exposure scenario allows a range of concentrations to be tested. Plant material may be used for a more realistic exposure scenario. Comparisons between different parts of a plant may provide answers to why certain pests are located on different plant structures in transgenic plants than in non-transgenic plants (Gore et al, 2002).

References:

1. Gore, J.; Adamczyk, J. J., et al (2005). Selective feeding of tobacco budworm and bollworm (Lepidoptera: Noctuidae) on meridic diet with different concentrations of *Bacillus thuringiensis* proteins. *Journal of Economic Entomology*. **98,** 88-94.
2. Gore, J.; Leonard, B. R., et al (2002). Behavior of bollworm (Lepidoptera: Noctuidae) larvae on genetically engineered cotton. *Journal of Economic Entomology*. **95**, 763-769.
3. James, C. (2004). Preview: Global status of commercialized biotech/GM crops: 2004. *ISAAA Briefs* No. 32. ISAAA: Ithaca, NY.
4. Metcalf, R. L. (1986). Foreword, pp. vii-xv. *In* J. L. Krysan and T. A. Miller [eds.], Methods for the study of pest *Diabrotica*. Springer, New York.
5. Mohd-Salleh, M. B.; Lewis, L. C. (1982). Feeding deterrent response of corn insects to beta-exotoxin of *Bacillus thuringiensis*. *Journal of Invertebrate Pathology,* **39,** 323-328.
6. Noldus Information Technology. Ethovision 3.0.
7. Strnad, S. P.; Dunn, P. E. (1990). Host search behaviour of neonate western corn rootworm (*Diabrotica virgifera virgifera*). *Journal of Insect Physiology,* **36,** 201-205.

Observing mother-infant sleep behavior

H.L.Ball[1], L.E.Volpe[2], E.Heslop[1], and S.J.Leech[1]
[1]*Parent-Infant Sleep Lab, Department of Anthropology, University of Durham, UK,*
[2]*Mother-Baby Behavioral Sleep Lab, Department of Anthropology, Notre Dame University, Indiana, USA*

Abstract
Behavioral analysis of mother-infant night-time caregiving was undertaken in two different studies: one in UK and one in US. Both studies involved similar behavioral observation techniques and utilized Noldus Observer Video Pro 5.0 for data capture and analysis. In each study observation periods lasted for the entire night and video coding was therefore complex and prolonged. This short presentation will describe our common experiences of coding and analyzing nocturnal mother-infant interactions using Observer Video-Pro.

Keywords
Nocturnal behavior, mothers, infants, The Observer Video Pro 5.0.

Introduction
Research teams from the Parent-Infant Sleep Lab at Durham University, UK, and the Mother-Baby Behavioral Sleep Lab at Notre Dame University, USA, both study aspects of mother-infant sleep behavior and night-time infant care-giving. Both teams utilize Noldus Observer software for data capture and analysis. The research questions we address are similar, deriving from anthropological perspectives on infant care practices within an evolutionary paradigm. The Durham team have recently completed a study of mother-newborn sleep behavior on the 1st two post-natal nights in a hospital setting, designed to examine how randomization to one of 3 alternative infant sleep locations affects breastfeeding initiation and other infant care practices. The Notre Dame researchers have recently examined night-time parenting behavior of adolescent and adult first-time mothers with their 4 and 8 months old infants in the setting of the sleep lab as part of an NICHD-funded project on early child neglect. In this paper we briefly summarize our respective research projects and discuss some of the methodological and analytical challenges faced by both teams in conducting these studies and some of the solutions we devised.

Methods

Hopital study
On the post-natal ward of the Royal Victoria Infirmary, Newcastle-upon-Tyne, England, mothers and newborn infants were randomized into three infant care conditions: bedding-in with cot side; bedding-in with side-car crib; rooming-in with stand-alone crib.

Mothers were recruited antenatally from breastfeeding workshops under the following inclusion criteria: intention to breastfeed, non-smoker, normal delivery, no opiate analgesics in labor. Mothers and infants were videoed in single rooms on the post-natal ward for up to 8 hours on each of the 1st and 2nd nights following delivery. Mothers were provided with a remote control with which they could start and stop filming.

Sleep Lab Study
In the Mother-Baby Behavioral Sleep Lab, Notre Dame, USA, adolescent and adult first-time mothers and their babies were filmed in a naturalistic setting for 1 night each when their the infants were 4 and 8 months old. Data on the sleeping and feeding patterns of 24 adolescent and 15 ethnically-matched adult primiparas were drawn from a larger longitudinal study on transition to parenting (N=684). Adolescent mothers were 15-18 years of age at childbirth and adult mothers were 22-37 years of age at childbirth. Participants were 56.4% European American, 43.6% African American, and 1% Latina. 62% of the infants in the sample were female. Mother-infant behavior was continuously recorded for up to 15 hours.
In both studies mother-infant sleep and care-giving behavior was captured to videotape using low-light intensity cameras illuminated via infra-red light. Cameras were either ceiling mounted (sleep lab) or mounted on a tall monopod positioned at the foot of the mother's bed (hospital) in order to obtain a clear view of both mother and baby during sleep.

Results
For both research teams the process of reconciling the complexity of the behaviors to be captured with the structure required by the software presented challenges. In some cases the limitations of the observation environment causes difficulty. In other cases aspects of software design thwarted some of our coding or analytical desires, or required creative circumvention. Below we highlight some of the challenges encountered and, where possible, the solutions we devised.

Nocturnal observation of mother-infant behavior
The research teams at Durham and Notre Dame both have many years experience in the ethology of mother-infant sleep and night-time care. We anticipated at the outset of the two projects discussed here that we would encounter certain observational difficulties given the limitations of coding from film, and the environments in which filming takes place. Both teams are committed, however, to maintaining as naturalistic a setting as possible and therefore were prepared for difficulties of observation such as making an assessment of infant sleep state when face is obscured by bedding; assessing feeding status if swallowing cannot be observed/heard; making a visual-only assessment of maternal sleep state; reliably coding highly variable behaviors such as flailing and crying. In order to manage these limitations coders discussed with each other any ambiguous situations that arose, and were trained (and tested) to apply strict coding rules and protocols devised for handling ambiguities.

Initialising behavioral taxonomies
For both teams of researchers the projects described above represent our first experience of using The Observer Pro

software. Our behavioral taxonomies, however, have been tried and tested over many previous studies, and have been cross-referenced within and between our research teams in order that our research outcomes are comparable. One of the major challenges both teams therefore faced was adapting our existing taxonomic structure to the format required by the Observer software. Individual behaviors and dyadic interactions (sometimes triadic in the hospital study) were coded as events or states according to similar multi-level behavioral taxonomies in both studies. The Durham (hospital) project employed 21 independent variables, 11 behavioral classes with up to 14 elements per behavioral class and 8 modifier classes with up to 20 elements per modifier class. Behavioral classes were predominantly states (10 of the 11 behavioral classes), many of which need to be continuously and simultaneously captured (e.g. infant's sleep state, sleep position, sleep location, proximity to mother are all continuously recorded, while simultaneously mother's sleep state, sleep position etc. are also being recorded). The Notre Dame (sleep lab) project employed a taxonomy consisting of 10 independent variables, 11 behavioral classes and 14 modifier classes. Selected behaviors focused on maternal and infant sleep state, infant feeding method, sleep location, presence of sleep-related risk factors, mother-infant proximity, amount of physical contact between mother-infant dyads, and potential neglect issues. Infra-red video recordings were coded by pairs of researchers using continuous event recording.

In addition to continuously recording state-based channels we needed to record transient states that mothers and infants may pass through such as crying, feeding, care-giving, and risk states. Modifier classes were used to record presence or absence of sub-groups of behaviors, such as practical vs. affectionate caregiving. The single events behavioral class was used to record brief behaviors such as visual and physical checks.

Due to the complex interaction of behaviors under observation with a limited number of modifier classes available we had to create a series of related yet mutually exclusive state categories. Because the Observer did not allow us to specify that a change in one behavior should automatically signal a change in another channel -- e.g. a mother cannot simultaneously 'observe infant' and 'care-give' under our taxonomic definitions but as these are on different channels one had to be switched off before the other was switched on -- our research staff developed complicated protocols for reminding ourselves which changes in certain states necessarily entailed new codes in other behavioral channels. In other cases we allowed certain behaviors to occur when 'technically' they couldn't (e.g. we decided to allow mothers' visual checks of the baby to occur while the mother was coded as 'asleep' in order to avoid having to stop and start sleep state channel at the point of every visual check). If some facility were available to automatically change the behavioral state in one channel when another behavior was coded this would be a huge benefit for this type of study.

Coding

A number of technical issues became apparent during coding and analysis. Some of these related to the length of the observation periods being coded and the fact that it is impossible to code an entire recording in a single coding session. It generally took at least two 8-hour days to code a single night's hospital recording with an estimated average of

15-1600 observations – often longer when there was lots of activity. In some cases in the sleep lab study recordings lasted 15 hours. It was necessary, therefore, to frequently pause coding to avoid coder-fatigue, and to be able to save coding in progress and restart again the next day. These two requirements revealed the following problems.

<u>Tape flutter</u>

We commonly experienced tape flutter when the tape was paused during coding which caused the machine-read time code to be read incorrectly into the software. Observations made immediately after pausing were then allocated an incorrect time which meant they could appear randomly through the existing observation list. We only became aware of these incorrect insertions when we were closing the package and an error message appeared (see below).

<u>Error messages</u>

Observations identified as containing an error were identified by the software on closing the observation package, however on doing so the observation session was then automatically locked out of the project. This meant that a partially completed observation containing an error could not be re-opened to fix the problem. As this may have involved an entire day's work it caused immense frustration! To overcome this problem we found that we could do the following: the .odf file had

to be saved separately with a new name, a new observation with an identical title to the original had to be opened in the observation package (replacing the original observation), and then closed, then the original (renamed) copy of the .odf file had to be copied back into the observation directory and the new observation reopened. We soon learned to make fewer coding mistakes!

A further problem encountered was that the error message indicating a false start time for a record entry provides the incorrect entry number making it almost impossible (and at least very time consuming) to manually locate in our long observation files.

<u>Twenty-four hour roll-over</u>

We encountered two issues with 24 hour roll-overs. Initially research staff wanted to leave their computers running overnight so they could resume coding where they ended the previous day. We discovered, however, that if you are coding at the turn of midnight, or leave the observation package running overnight, the software automatically inserts 24 hours into the observation – even if no code is being transmitted to the time code reader (i.e. video player is off).

Furthermore, an observation which includes a time code reading passing over the 24 hour mark will cause the observation package to attribute subsequent observations to before the observation began – i.e. it does not continue to count elapsed time, but begins again. For instance in the hospital study mothers sometimes were requested to initiate video recording on the 2nd night of observation if they were being filmed over a weekend. A problem arose in those cases where the power to the time-code generator was left on in between the first and second recording nights. This led to the time code passing through the 24-hour roll-over during the 2^{nd} night of filming. Once the tape was being coded it transpired that the software mis-interpreted the post-24 hour time signal and attempted to insert it prior to the 24-hour roll-over. Our solution in this instance was to begin coding after midnight on these tapes, otherwise a time-consuming process of exporting to Excel, manually altering time codes, and re-importing to the Observer was the suggested remedy.

Software limitations and bugs

One feature of the Observer software that caused difficulty involved the limitation in not being allowed to have the same actor as both a primary subject and as a modifier. In the hospital study, for instance, one of our projected outcomes was to ascertain whether the infants randomly allocated to each of the 3 sleep environments encountered any actual or potential risks in each environment such as falling, entrapment or head covering, and how any potential risks were resolved (i.e. prevented from becoming actual risks). In some situations an infant was able to resolve his own potential risk situation, for instance by moving the covers away from his head – however as the infant was the primary subject he couldn't also be coded as the modifier. This was not discovered until half the tapes were coded and led to the actor who resolved the risk in these situations being coded as 'missing'.

Frequently behaviors would appear displayed on the screen in other behavior channels for no apparent reason. For instance sleep state information (e.g. mother awake) would appear in the display box on the screen for a different channel. This made monitoring the status of each channel challenging!

Analysis issues

Simple analyses involve generating output of event frequencies and state durations per dyad (e.g. frequency of crying, duration of sleep bouts). More complex analyses of behavior sequences and time-lag analyses are also required (e.g. intensity of signaling required for baby to wake mother; time lag between baby's first signal and maternal response).

A small bug that took us a long time to circumvent involved selection of start-end events when the 'end' point of an event involved the 'start' of another event. This occurred when we wanted to calculate how long after attempting to feed babies actually began feeding. When selecting the interval period from start of behavior A to start of behavior B it was necessary to alter the entry in the 'event duration' window in order to make available for selection the options in the 'relative to' window that allows the option of ending an interval 'before start of behavior'. This circuitous way of selecting options does not appear in the manual.

Both research teams are still in the process of analyzing the data from these studies and the more complex analyses remain to be tackled. Further issues therefore may yet be encountered – and these will be discussed at the conference.

Durham University's hospital study was supported by a research grant from Babes-in-Arms. Notre Dame's sleep lab study was supported by NICHD HD 39456, Centers for the Prevention of Child Neglect 2001.

Behavior around oviposition by hens in cages with and without nest boxes

G.M. Cronin [a] and M.A. Desnoyers [b]
[a] *Animal Welfare Science Centre, Department of Primary Industries Victoria, Werribee, Vic 3030, Australia*
[b] *UMR INRA-INAPG, Physiologie de la Nutrition et Alimentation, Paris Cedex 05, France*

Abstract

The housing of laying hens in conventional cages is a contentious animal welfare issue. A potentially acceptable alternative is modified cages containing a nest box. However, not all hens in cages with nest boxes choose to lay in the nest box, suggesting we lack understanding of the factors influencing nest-site selection by hens in cages. Using The Observer program we measured the behavior of hens around egg laying from time-lapse video records. Hens were kept in pairs in observation cages either with or without a nest box (n=18 focal hens in each situation). Nine of the 18 focal hens with access to a nest box laid there; the other nine laid on the wire floor of the cage. This paper reports the methods used to measure the behavior of hens around egg laying and comments on the differences detected in behavior between hens that laid in compared to outside the nest box.

Keywords

Poultry, behavior, welfare, nest box, egg laying

1 Introduction

From studies of the pre-laying activities of domestic hens, it is generally accepted there are two phases of behavior involved in egg laying [11]. Beginning about 1 h prior to egg laying, the activity level of hens increases in a phase of behavior termed "searching" in which hens appear motivated to seek a nest site. In this phase hens increase locomotion and perform behaviors such as inspection of potential nests. Once hens have selected the preferred nest site the "sitting" phase commences, which includes the adoption of a sitting posture interspersed with nest-building activities such as scratching the floor/litter, rotating the body on the nest and collecting litter if available.

Activities performed in the searching phase are goal-directed or appetitive behaviors, occurring when hens are motivated to find a suitable nest for egg laying (the consummatory behavior). Thus, it has been shown that hens were motivated to lay their egg in a nest box [1,5], and if a nest box was not available hens performed more nest-searching behavior [3,6]. While an increased occurrence of appetitive behavior may indicate a stronger motivation to achieve the consummatory phase, it does not necessarily indicate that increased pre-laying locomotion reflects increased frustration and thus a potential welfare problem. For example, using an aversive task approach, it has been suggested that hens were only weakly motivated to reach the nest site during the searching phase, although the motivation to gain access to a nest site increased near the start of the sitting phase [7]. Further, a comparison of hens that were consistent and inconsistent in their use of nest boxes [4] reported no apparent difference between the two classes of hens in hens' motivation to use a nest box. The inconsistent hens were less responsive to the cues provided by nest boxes than consistent hens [4]. Based on evidence of motivation of hens to lay in a nest box, increased time spent in pre-laying behaviors in the absence of a nest and increased vocalisations indicative of frustration when access to a nest is blocked, it has been concluded that there is convincing evidence of the importance of a suitable nest site [9]. However, it is also known that even in the presence of a nest not all birds lay in the nest with reports of 0-80% of eggs being laid on the floor [12] and the question remains of what do hens perceive as a suitable nest site?

2 Materials and Methods

Commercial Hyline Brown hens aged 45-48 weeks were used in this experiment. The hens were part of a larger experiment investigating the effects of different cage "furniture" on hen welfare (ie nest box, dust bath and perch; Victorsson Trivselburen 8-bird Furnished Cages, Sweden). The cages were located in a controlled climate shed with control for temperature, light and ventilation. Temperature was maintained at about 17°C during the dark period and 23°C during the light period. Lighting was provided by incandescent globes and controlled by a computer and the hens were exposed to a light-dark regime of 16 h light : 8 h dark. Lights were programmed to come on at 0500 h and there was a daily 'sunrise' and 'sunset' of 30 min duration commencing at 0500 h and 2030 h, where the light level either slowly increased from dark or slowly decreased from light to dark, respectively. During the light period light levels were about 5 lux. Humidity was maintained at about 40%.

For the present experiment, two Victorsson Trivselburen Furnished Cages positioned back-to-back and with all furniture removed were used as observation cages. Each cage measured 1.2 m wide, 0.5 m deep and 0.45 m high at the rear of the cage. The 36 focal hens (18 pairs) selected for the experiment were from six specific "home" cages (8 hens per cage) that were without dust baths or perches. Three of these "home" cages however contained a nest box (NB treatment) and three were without a nest box (No NB treatment). There were nine replicates in time and the nest box was randomly allocated to one of the two observation cages for each replicate. The nest box measured 0.24 m wide, 0.5 m deep and 0.27 m high at the front of the cage and had a solid ceiling, rear and sides, apart from an entrance opening in one side wall. A blue vinyl flap covered the front of the nest box while the nest box floor was overlain with "astro turf" (0.37 m x 0.22 m x 15 mm thick).

At about 1600 h on day 1 of each replicate, a pair of hens was selected on an *ad hoc* basis from a home cage containing a nest box and transferred to the observation cage in which an identical nest box had been fitted. Similarly, a pair of hens from a home cage without any furnishings was moved into the other observation cage. Video recording, which commenced from the time the hens entered the cages, was assisted by infra-red (IR) light in the dark and low light conditions. Each hen was marked on the feathers of the back with carbon-based black ink for differentiation on the video record under normal light as well as IR light.

A total of 6 black and white video cameras with in-built IR lights were used to continuously record the activities of the hens. Two video cameras were fixed to the ceiling of the shed, one above each observation cage. These two cameras were connected to individual time-lapse video recorders (Panasonic AG-6124) situated in an adjacent shed. The four remaining video cameras were connected to a 4-channel, black and white quad video processor unit which in turn was connected to a third time-lapse video recorder. Two cameras were positioned to provide views of the front of both cages, showing the egg trays into which eggs normally rolled for later collection. The third camera was placed inside the nest box. The fourth camera that occupied the final channel of the quad-splitter unit, was placed inside one of the cages at floor level, at random, to provide an additional perspective to assist in determining hen behavior if required. A "dummy" camera was positioned in a similar position to this fourth camera in the other observation cage.

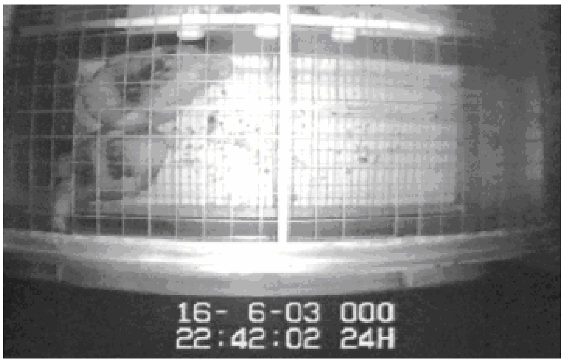


***Figure 1**. Image of the No NB treatment observation cage from above, captured from the time-lapse video record. The cage was illuminated using infra-red light. The feathers of the two hens were marked with carbon-based black ink for individual identification.*

To allow habituation to the observation cages there was a minimum of 24 h from placement in the cages to the commencement of data collection from 4 h of video records, commencing 2 h pre-laying on the day that both hens in the pair laid an egg on the same day. For 15 pairs, the data for analysis were captured from the second day in the observation cages and for three pairs it was the hens' third day in treatment. The mean (± std dev) time to the first egg laid on the observation day was 47.0 ± 10.7 h (min. 38.4 h, max. 67.5 h). Most hens laid one egg during this "habituation" period, although only the times and locations of egg laying were recorded from the video record. As the observation cages were of the same design as the home cages, and the pairs of hens originated from the same home cage, it was assumed that this would be a sufficient period of habituation.

An ethogram of 4 postures, 8 behavioral states and 16 behavioral events was developed to enable the transcription of the hen behavior data from the time-lapse video records using The Observer behavior recording program supplemented with the Support Package for Video Analysis (version 4.0 for Windows; Noldus IT, The Netherlands). Hen movement was tracked by dividing the video image from the overhead cage view into nine similar-sized areas, excluding the nest box or the equivalent area in the No NB treatment cage, and the frequency of hens occupying the different areas in the cage. This was achieved by marking lines on an acetate sheet attached to the TV monitor at data transcription. Frame-by-frame analysis of the video record was facilitated by the use of a Panasonic AG-7355 Video Cassette Recorder with jog-and-shuttle control linked to a computer.

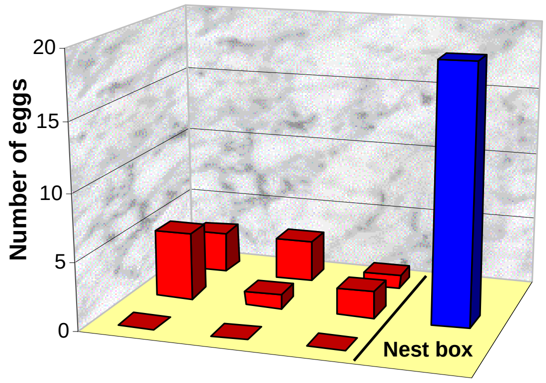
Using an iterative step in the data collation procedure in The Observer, the 4 h observation period was divided into eight, 30-min periods to facilitate detailed behavioral analysis in the pre- and post-laying periods. In general, hen behavior was similar in Time periods 1 and 2 (ie 120-60 min pre-laying) and thus were analysed together. Time periods 3 and 4 (ie 60-30 and 30-0 min pre-egg laying, respectively) were analysed individually. There were no differences in hen behavior in the four, 30-min periods post-egg laying, so data for these periods were combined for analysis. The statistical analysis involved a restricted maximum likelihood analysis (REML) [8] on individual hen data with random effects for home cage, replicate and pair of birds and fixed effects of laying status of the bird and its cage-mate and presence of a nest box. The random effect of pair of birds accounted for any correlation between the two birds in each pair. In the NB treatment, since some hens did not lay in the nest box data were thus analysed according to where egg laying occurred, ie hens were classed as either "nest-layers" or "floor-layers".

3 Results

Of the 18 hens in the NB treatment, nine laid in the nest box (nest-layers) and nine laid outside the nest box on the wire floor (floor-layers). For the analysis, the total number of hens in each analysis group was thus nine nest-layers and nine floor-layers for the NB treatment and 18 floor-layers in the No NB treatment, ie a total of 27 floor-layers. Based on observation of the hens on consecutive days, only one hen was observed to change her location of laying between days, ie laid in the nest box and on wire floor on different days. Every combination of "pairing" was observed, with both hens laying in the nest box or on the cage floor, one of each laying in the nest box or on the cage floor and overlap or separation of both egg laying and nesting behavior within a pair of hens (ie both hens laid eggs within the same 4-h period or with a clear separation of time periods).

A total of 69 eggs were laid by the 36 hens over the 2 or 3 days that hens occupied the observation cages. In the NB treatment, 19 of 34 eggs (55.9%) were laid in the nest box. Figure 2 shows the locations where eggs were observed to be laid in the two treatments, including in the nest box, over all days of the experiment.

Hen activity level around egg laying was assessed by the time hens spent walking and the frequency of entering areas 1-9 of the cages. Darkness strongly reduced hen activity. As the 2-h pre-laying period included darkness for six hens, the time in the dark was used as a co-variate in the analysis. These six hens alternated between standing and sitting in the dark and none were observed to locomote. After adjusting for darkness, there were no differences in the activity measures in the period 120-60 min pre-laying. However, in the 1 h before laying, nest-layers were less active than floor-layers; nest-layers performed less (P<0.01) walking behavior (excluding following behavior) and entered areas 1-9 less frequently (P<0.01) than floor-layers (Table 1).



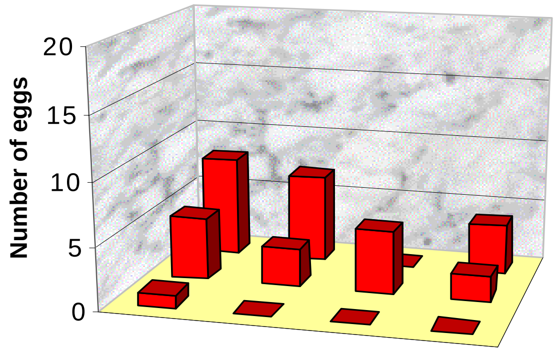


Figure 2. The number of eggs laid by hens in the NB treatment (upper) and No NB treatment (lower) during all days of the experiment, distributed according to location in the cage where each egg was laid.

Table 1. Pre-laying behaviors of nest-layers and floor-layers. Values shown are back-transformed means.

Behavior	Time pre-laying (min)	Nest-layers (n=9)	Floor-layers (n=27)
Stand stationary (min)	120-60	45.8	38.6
	60-30	16.4	18.1
	30-0	5.5 ^{p}	11.5 ^{q}
Sit stationary (min)	120-60	12.0	9.8
	60-30	8.7	4.5
	30-0	24.9 ^{p}	10.2 ^{q}
Walk (min)	120-60	4.8	5.3
	60-30	1.7 ^{x}	4.2 ^{y}
	30-0	0.3 ^{p}	4.5 ^{q}
Areas entered (freq)	120-60	93	98
	60-30	26 ^{x}	73 ^{y}
	30-0	3 ^{p}	90 ^{q}

Within rows, means with different superscripts differ significantly: ^{x,y} P<0.01; ^{p,q} P<0.001.

A behavior termed "following behavior" was observed. Following behavior, defined as the focal hen walking or running but following the other hen as she moved about the cage, was only observed in the floor-layers (16 of the 27 hens) and occurred during about 5% of the total time in the period 60 min pre-laying to 30 min post-laying and the peak frequency of occurrence was in the 30 min prior to egg laying.

On average (± std dev), the nine hens in the NB treatment that were nest-layers visited the nest box 3.9 ± 1.69 times and spent 49.0 ± 11.65 min in the nest box in the 2 h pre-laying. In comparison, six of the nine floor-layers in the NB treatment visited the nest box in the 2 h pre-laying, although they did not lay in the nest box. These six hens on average visited the nest box 3.7 ± 2.16 times and spent 5.3 ± 10.42 min in the nest box in the 2 h pre-laying. In the 2 h post-laying, one of the 9 nest-layers and two of the 9 floor-layers in the NB treatment entered the nest box.

While the combined frequency of occurrence of hens scratching the wire floor of the cage or the astro-turf floor inside the nest box was greater in the 30 min prior to egg laying for the nest-layers (14.7 bouts/hen) than floor-layers (9.8 bouts/hen), the difference was not statistically different (P>0.05). The mean values for the nest-layers and floor-layers in the period 60-30 min pre-laying were 3.4 and 3.0 bouts, respectively (P>0.05).

Nest-layers spent less (P<0.01) time feeding in the 30 min pre-laying compared to floor-layers (0.3% vs 2.9% of the time, respectively). The occurrence of feeding behavior in other time periods however was not affected by the hens' choice of laying in the nest box or on the wire floor. Similarly, there were no effects of egg location class on the time hens spent preening around egg laying.

4 Discussion

The hen's choice to lay her egg in the nest box compared to on the wire floor outside the nest box had a large effect on her behavior in the hour prior to egg laying. In the present experiment the time spent walking was significantly reduced if hens were nest-layers, supporting previous findings that if a nest box was not available hens performed more nest-searching behavior [3,6]. The second estimate of hen activity in the present experiment was the frequency of hens entering areas 1-9 in the cages. Using this parameter, the nest-layers were also significantly less active in the hour before laying than floor-layers. Nest-layers were less active during the hour before laying because they spent more time sitting. Similarly, because nest-layers spent more time sitting (in the nest box) before laying, analysis of the data also confirmed that nest-layers spent less time feeding in the 30 minutes pre-laying. Preening behavior however, which occurred during about the same proportion of observation time as feeding before egg laying, was not affected by choice of egg laying site. This finding was probably due to the fact that hens could preen at any location, whereas feed was only available outside the next box.

For eight of the nine pairs of hens in the NB treatment, egg laying by the two hens was separated by a minimum of 80 min (range 80 to 583 min). Thus it appears unlikely that either of the hens in the pairs was excluded from using the nest box. In one pair, egg laying occurred two min apart with one hen laying in the nest box and the other on the floor outside the nest box. In only one pair of hens, both hens laid in the nest box and for this pair the time between the respective egg laying events was 80 min. During the video observation period for this pair of hens (320 min), at no time were both hens in the nest box together. While these data could suggest that hens, on a daily basis, “choose” their laying site, it needs to be remembered that by the time these data were collected, the majority of the hens were already consistently laying at one site (for each hen). Thus further research is required on the development of choice of egg laying site.

This small experiment that involved observations on egg-laying behavior in 18 pairs of hens has provided more questions than answers. An interesting observation was “following behavior” which only occurred in hens that laid eggs on the wire floor, irrespective of the presence of a nest box. This behavior occurred during pre- and post-laying periods and involved the hen (follower) appearing to attempt to remain close to the other hen (followed), including when the followed hen was locomoting. When the followed hen was stationary and standing, the follower would often sit next to her and the follower would put her head under the body of the followed hen. A similar behavior has been previously described [10] in which hens appeared to follow and attempt to crawl underneath pen-mates. The reasons for this behavior are unknown as are the reasons why it occurred in 59% of floor-layers and occurred both pre- and post-egg laying. One explanation is that when a nest box was present which was utilised for egg laying, this environment provided appropriate cues for nest-site selection. Another explanation is the follower hen may have derived cues for nest-site selection from the followed hen.

This experiment has shown, as reported in other studies [4], that the use of the nest box for egg laying is highly variable between birds. In this experiment where experienced (with a nest box) hens were housed in pairs in a cage with a nest box, 50% of hens laid in the nest box and 50% laid on the wire cage floor outside the nest box. This was similar to the daily egg production records for the three NB treatment home cages, recorded over 5 months in the larger experiment (prior to the start of this smaller experiment), in which 55.8% of eggs were laid in the nest box. These data raise the question of hens’ preference for egg laying location. While the literature suggests hens are motivated to seek a “preferred” location for egg laying, the data from this experiment could be interpreted to suggest that either the nest box or the wire cage floor may be preferred locations. Alternatively, by one hen making a choice, this may or may not force the other hen into a less-preferred location. This experiment only used pairs of hens and the possible combinations for preference presumably become more complex in commercial settings of group sizes of 5 to 20 hens, particularly when the number of nest sites (boxes) is limited. Clearly further research is required to answer the following types of questions: What is the biological significance of following behavior, including any relationship with nest-site selection? Is consistency of nest-site selection associated with a preference for that site, or are some birds forced to choose a less-preferred site? Are these behaviors the same with larger group sizes? Are there any implications for welfare?

This research was co-funded by the Department of Primary Industries, Victoria, and the Australian Egg Corporation Ltd. We gratefully acknowledge Dr John Barnett for his advice on all aspects of this research, Mr Kym Butler for conducting the statistical analyses and Ms Samantha Borg, Mr Bruce Schirmer and Ms Tracie McCallum for assisting with the care of the hens. Funds to enable the first author to attend the conference were provided by the Department of Primary Industries, Victoria.

References:

1. Appleby, M.C. and McRae, H.E. (1986). Individual nest box as a super-stimulus for domestic hens. *Applied Animal Behaviour Science*, **15**, 169-176.
2. Appleby, M.C., Smith, S.F. and Hughes, B.O. (1993). Nesting, dust bathing and perching by laying hens in cages: effects of design on behaviour and welfare. *British Poultry Science*, **34**, 835-847.
3. Cooper, J.J. and Appleby, M.C. (1995). Nesting behaviour of hens: Effects of experience on motivation. *Applied Animal Behaviour Science*, **42**, 283-295.
4. Cooper, J.J. and Appleby, M.C. (1997). Motivational aspects of individual variation in response to nestboxes by laying hens. *Animal Behaviour*, **54**, 1245-1253.
5. Duncan, I.J.H and Kite, V.G. (1989). Nest site selection and nest-building behaviour in domestic fowl. *Animal Behaviour*, **37**, 115-231.
6. Freire, R., Appleby, M.C. and Hughes, B.O. (1996). Effects of nest quality and other cues for exploration on pre-laying behaviour. *Applied Animal Behaviour Science*, **48**, 37-46.
7. Freire, R., Appleby, M.C. and Hughes, B.O. (1997). Assessment of pre-laying motivation in the domestic hen using social interaction. *Animal Behaviour*, **54**, 313-319.
8. GenStat Committee (2000). *The Guide to GenStat®, Part 2: Statistics*. Lawes Agricultural Trust, Rothamsted, United Kingdom.
9. Keeling, L.J. (2004). Nesting, perching and dustbathing. *Welfare of the Laying Hen*, CABI, Wallingford, Ed. G.C. Perry, pp. 203-213.
10. Kite, V.G., Cumming, R.B. and Wodzicka-Tomaszewska, M. (1980). Nesting behaviour of hens in relation to the problem of floor eggs. *Reviews in Rural Science IV*, Uni. of New England, Armidale, Australia, Sept. 1979, Eds. Wodzicka-Tomaszewska, M., Edey, T.N. and Lynch, J.J., pp. 93-96.
11. Sherwin, C.M. and Nicol, C.J. (1992). Behaviour and production of laying hens in three prototypes of cages incorporating nests. *Applied Animal Behaviour Science*, **35**, 41-54.
12. Sherwin, C.M. and Nicol, C.J. (1993). Factors influencing floor-laying by hens in modified cages. *Applied Animal Behaviour Science*, **36**, 211-222.

Detecting temporal patterns in dog-human and robot-human interactions

A. Kerepesi[1], G.K. Jonsson[2], E. Kubinyi[1] & A. Miklósi[1]

[1]*Department of Ethology, Eötvös University Budapest, Hungary &Comparative Ethology Research Group, Hungarian Academy of Sciences, Budapest, Hungary*
[2]*Human Behaviour Laboratory, University of Iceland & Department of Psychology, University of Aberdeen*

Abstract

In this work we demonstrate that with the use of THEME we can detect idiosyncratic T-patterns in dog/robot-human interactions. In Study 1 we investigate a cooperative interaction between the dog and its owner and in Study 2 we compare spontaneous interactions with a dog puppy or AIBO in children and adults.

During the cooperative interaction we found complex and idiosyncratic T-patterns and a common sequence of behavior units were identified, which is important in the successful completion of the task.

In the analysis of the spontaneous interactions, we found a higher number of T-patterns in children's interaction with the puppy than in their interactions with the robotic dog, while no such difference were found in adults. No similar T-patterns were detected across dyads.

Highly structured T-patterns were detected both in a cooperative and in a spontaneous interaction, and the finding suggest that structured tasks resulted in similarities in T-patterns across dyads

Keywords

Dog-human interaction, robot-human interaction, temporal patterns

Introduction

Interaction between humans and dogs can occur in different contexts. In everyday life man interact with dogs through various forms of communicative and cooperative interactions. Most often humans walk their dog (without a leash), or humans of various ages can be seen to play with their pet [11, 12, 14]. To our knowledge only few studies investigated the temporal aspects of human-dog interactions [11, 12, 13] focusing on the synchronization of dog and human actions. However the applied time window was very short taking into account only actions that have followed directly one another. As an exploratory study we decided to investigate cooperation between the dog and its owner assuming that cooperative interactions provide a natural context for the emergence of temporal patterns.

We have also investigated humans' playing behavior to a dog in a spontaneous situation and compared to their play behavior to playing with a robotic dog, Sony's AIBO. AIBO has a sense of touch, likes being stroked, can move freely and interact with its human partner [6]. Questionnaires suggest that people consider AIBO as a companion and view it as a family member [3, 10], however there is a difference in the behavioral interaction between human-dog and human-AIBO [13].

In order to decide if humans interact with AIBO as a robotic toy or a dog, we propose analyzing the temporal structure of these interactions.

For the analysis of complex social interactions the preferred method should be highly flexible in detecting temporal patterns. The THEME program allows the analyst to detect complex repeated temporal patterns, even when a multitude of unrelated events occur in between components of the patterns, which typically makes them invisible to the naked eye and (to our knowledge) to currently available statistical methods and software [1, 8, 9].

Here we present two different behavioral situations that have been analyzed with THEME. In Study 1 we investigated cooperative interactions between the dog and its owner. In this exploratory study our aim was to describe temporal patterns in the behavior of interacting dog-human dyads observed in a cooperative task [for details see 7].

In Study 2 we investigated children's and adults' behavior during a spontaneous play session with either a dog puppy or AIBO. Contrary to Study 1, where we investigated a structured situation, we now focus on analyzing a spontaneous play between the human and the dog/robot. Our aim was to characterize the temporal structure of these interactions by analyzing and compare the temporal structure of the interaction with dog and AIBO in both children and adults.

Method

Study 1: cooperative interaction

Seven owners (3 men and 4 women) and 10 adult pet dogs (4 males and 6 females, mean age: 3.6 ± 2.7 years; 7 Belgian Tervuerens, 2 Vizslas and 1 Collie) participated in the test. Their task was to get the building blocks from a starting point to the target point with the goal of building a tower. We used 24 plastic cubic building blocks (a children's toy) in 8 different sizes, and 3 items for each size. In each case the experimenter determined the starting point (the location of the building blocks on the floor) and the target point (the location for the tower to-be-built) in the room 4 meters apart. The test consisted of three 5-min episodes. The behavior of the owner and their dog was recorded on video.

In the first episode only the owner carried the building blocks from one location to the other without any help from the dog. The owner could once call the dog's attention verbally at the beginning of the episode but after this she/he must not talk to the dog. In the second episode both the dog and the owner could carry building blocks to the target point. The owner was allowed talk to the dog but only the use of less familiar verbal utterances was permitted. In the third episode the owner sat on the floor at the target point with 5 building blocks in front of her/him and was not allowed to leave this position. In order to build the tower she/he had to rely on the help of the dog, so she/he could instruct the dog to carry the building blocks to her/him. The owner was allowed to talk to the dog but only less familiar verbal utterances were permitted (e.g. 'Please help me' or 'Come on' etc) and direct commands for retrieving were prohibited. No restrictions have been placed on gestural communication.

We have coded and analyzed only the third episode of this cooperative task with the Theme software package. In our samples we have coded 38 behavior units in dogs, and 32 gestural behavior units and 8 verbal behavior units in humans. The present analysis is limited to 21 behavior units that occurred most frequently in the behavior of the partners. Nine of the 21 behavior units were displayed by the dogs, 8 of them can be regarded as human gestural behaviors and 4 of them were verbal utterances.

From the point of view of the successful fulfilment of the task one of the most important behavior elements is the '*dog picks up the building block'*. Therefore we have focused our analysis on the (interactive) T-patterns involving this goal-oriented behavior. We have searched for T-patterns that occurred so often as the dog picked up the building block during the task minus one. We have analyzed the characteristics of T-patterns detected by investigating their number, length (the number of behavior units in the sequence) number of actors in a T-pattern and their composition of behavior units. T-patterns containing both dog and human behavior units are called interactive T-patterns. We have also looked for common T-patterns present in all dog-owner pairs.

Study 2: spontaneous interaction

Twenty eight adults and 28 children participated in this test. They were divided into four groups:

1. Adults playing with AIBO: 7 males and 7 females (Mean age: 21.1 years, SD= 2.0 years)
2. Children playing with AIBO: 7 males and 7 females (Mean age: 8.2 years, SD= 0.7 years)
3. Adults playing with dog: 7 males and 7 females (Mean age: 21.4 years, SD= 0.8 years)
4. Children playing with dog: 7 males and 7 females (Mean age: 8.8 years, SD= 0.8 years)

The test took place in a 3m x 3m separated area of a room. Children were recruited from elementary schools, while adults were university students. The robot was Sony's AIBO ERS-210, (dimension: 154mm × 266mm × 274 mm; mass: 1.4 kg; color: silver) that is able to recognize and approach pink objects. The living dog puppy was a 5-month-old female Cairn terrier, similar in size to the robot. It was friendly and playful, its behavior was not controlled in rigid manner during the playing session. The toy was a pink ball for the AIBO, and a ball and a tug for the dog-puppy. The tug was introduced in order to motivate the dog puppy to play.

The participants played for 5 minutes either with the AIBO or the dog puppy in a spontaneous situation. None of the participants met the test partners before the playing. At the beginning of each play we asked participants to play with the dog/AIBO for 5 minutes, and informed them that they could do whatever they wanted, in that sense the participants' behavior were not controlled in any way. Those who played with the AIBO knew that it liked being stroked, that there was a camera in its head enabling it to see and that it liked to play with the pink ball.

Two minutes (3000 digitized video frames) were coded for each of the five-minute-long interaction. The starting point of the coding procedure was when the participant first touched the dog or the robot.

The present analysis was limited to the behavior units connected with playing behavior ('dog/AIBO approaches the toy' and 'human moving the toy in front of the dog/AIBO') and stroking of the dog/AIBO. During the coding procedure we recorded the frequency and the duration of behavioral units. We have also transcribed the latency of the first human tactile contact with the dog/AIBO (this was also the starting point of the two-minute-long coding with ThemeCoder). Concerning the search for temporal patterns (T-patterns) we used, as a search criteria, minimum two occurrences in the 2 min. period for each pattern type, and only include interactive patterns (those T-patterns which contain both the human's and the dog's/AIBO's behavior units). A special focus is also on whether the human or the dog/AIBO starts and terminates the T-pattern more frequently. A comparison between the ratio of T-patterns initiated or terminated by humans, in the four groups, was carried out as well as the ratio of those T-patterns containing behavior units of playing and stroking. Tests were also conducted on the effect of the subjects' age (children vs adults) and the partner type (dog puppy vs AIBO) using two-way ANOVA.

Results

Study 1: cooperative interaction

In our sample we have found on average 218 T-patterns in total and 181 of them were interactive (83%). The difference between the "real data" and "randomized data" was significant (t_9= 2.914 p= 0.017), suggesting that T-patterns are not the results of chance effects.

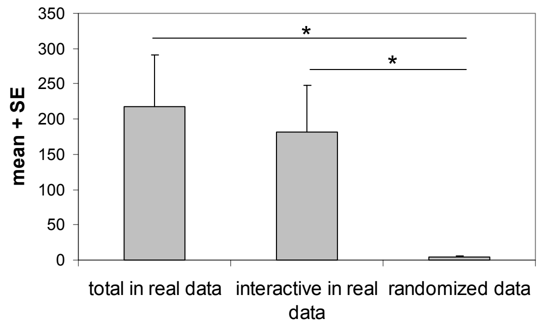


Figure 1*: Number of detected T-patterns during dog-owner cooperation. (N=10)*

In the present analysis we have found on average 11.3 T-patterns that contained the behavior unit *'dog picks up the building block*'. Interestingly, all but one occurrence of this behavior unit was found to be part of a T-pattern (74 out of 75 in our sample of 10 dogs). In the one exceptional case the dog accidentally dropped the brick, and the second 'pickup' is not included in any T-pattern. This means that 98% of occurrences are parts of a pattern, which is more than for any other behavior unit. Similar high levels of preference for being in pattern can be found for *'dog goes to human'* and *'dog goes to bricks'* (96% and 91% respectively). In the case of the human *'directing words'* can be found most frequently in T-patterns (66%) that is followed by '*pointing*' with 59%. Results show that not only behavioral units naturally associated with the particular action are present at a higher frequency but also communicative behaviors like '*dog looks at the owner's face*' or '*owner looks at the dog*'.

Considering all T-patterns containing the '*dog picks up the building block*' unit together, we can construct a matrix that shows all possible transitions from one behavior unit to the next. Note, however, that this is not a traditional transition matrix but one where all transitions have been previously verified statistically by THEME. If now we look for the most frequent transitions we get a cycle shown on Fig. 7.

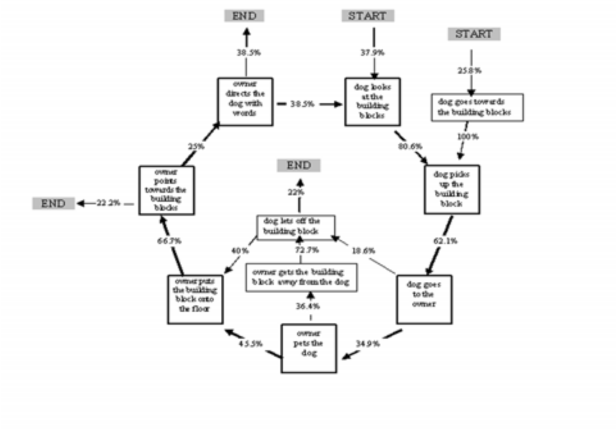


Figure 2. The typical behavioral sequence emerging during the cooperative task

The typical sequence emerging during the task was cyclic (the starting and the terminating behavior unit was the same) and consisted of 7 behavior units. This sequence was the outline of the successfully completed task. The part of this cycle or even the whole cycle occurs in the majority of T-patterns containing the '*dog picks up the building blocks*' unit.

Study 2: spontaneous interaction

The number of interactive T-patterns detected did not differ significantly among the groups (Fig. 3).

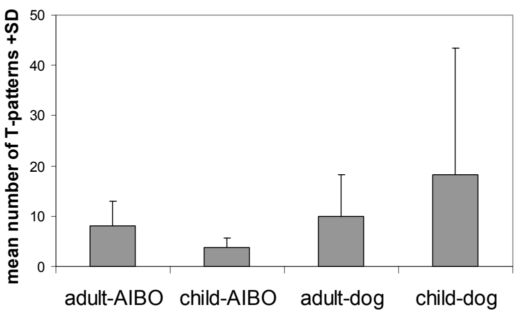


Figure 3. *Number of detected T-patterns in the spontaneous interactions in dog or AIBO and humans. (N=14 in each group)*

Comparing the ratio of T-patterns started by humans among the groups ($F_{3,56}$= 5.270, p= 0.003), we have found that adults started T-patterns more frequently when playing with AIBO, than participants of the other groups. Both the age of the human ($F_{1,56}$= 10.493, p= 0.002) and the partner type ($F_{1,56}$= 4.514, p= 0.038) had a significant effect, but their interaction was not significant.

Adults initialized T-patterns more frequently when playing with dog than children, but no such difference were found in their play with the AIBO.

Comparing the ratio of T-patterns terminated by humans we have found that the partner's type ($F_{1,56}$=10.725, p=0.002) had a significant effect. Both children and adults terminated the T-patterns more frequently when they played with AIBO than when they played with the dog puppy. (Fig 4)

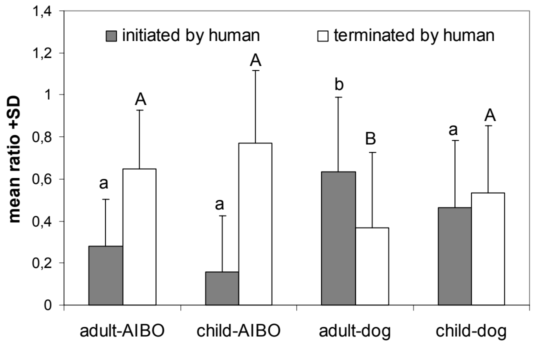


Figure 4. *The ratio of T-patterns initialized and terminated by humans in the spontaneous play.*

The age of the human had a significant effect on the ratio of T-patterns containing "*the dog/AIBO approaches the toy*" ($F_{1,56}$= 4.229, p= 0.045), and the interaction with the partner's type was significant ($F_{1,56}$= 6.956, p= 0.011). This behavior unit was found more frequently in the T-patterns of adults playing with dog than in the children's T-patterns when playing with dog (see Fig.5).

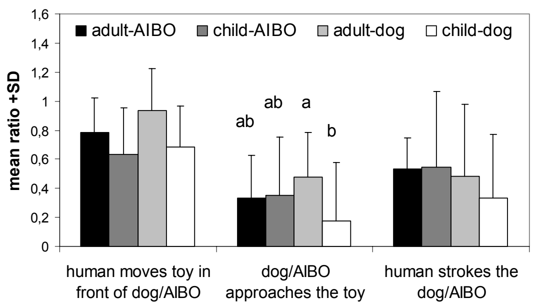


Figure 5. *The ratio of T-patterns containing "the dog/AIBO approaches the toy" in the spontaneous situation.*

Discussion

The present analysis provides strong support for long-term temporal sequences both in dog-human and robot-human interaction. This is the first time that such temporal patterning has been shown by statistical methods.

In Study 1 we have shown that during cooperative interactions dogs' and humans' behavior becomes organized into highly complex interactive temporal patterns. Although one can assume that cooperation or communication requires some regularity in behavior of the partners this notion has been often taken for granted without experimental or statistical verification

This analysis suggests that in the course of the present cooperative task many task-related actions enacted by the partners became spontaneously organized into T-pattern. The repetition of the same sequence (similar to the one shown on Fig 2.) allows the behavior units to organize into a pattern that occurs each time when the dog picks up a building block. By its very nature the detected T-pattern does not only represent a sequential organization but a temporal relationship among these units that is also relatively constant and gives a typical behavioral rhythm to the pattern. It is very likely that such connection over a long period of time could have not been detected by traditional methods of sequential analysis where one is only looking for first order relationship between behavior units.

Similarly to human interactions [4, 9] and human-animal interactions [7], the results of Study 2 shows that human-robot interactions also consist of complex temporal patterns. In addition these temporal patterns are comparable to those T-patterns detected in dog-human interactions in similar contexts.

One important finding of the present study was that the type of the partner affected the structure of the T-patterns detected. Adults initialized T-patterns more frequently when playing with dog than children, however, we did not found such difference when they played with AIBO. In contrast humans terminated T-patterns more frequently when playing with AIBO than playing with dog puppy. This can be explained on the basis of the programming of the AIBO, because the robot never stops interaction until his visual systems detects a pink ball in the environment. Other ethological observations also support this difference. Most of the owner's action remains without reaction [11, 12, 14], so it is more likely that the dog terminates their interaction. Humans' age also affects their interaction with the dog [11, 12], which can explain our results that not just the partner's type but also the age of humans affects on the structure of the T-patterns. Our results show that humans and dogs tended to terminate the T-patterns with the same frequency, since no significant difference were found in the ratio of T-patterns terminated by dogs and humans during their interaction, while humans tended to terminate T-patterns more frequently when playing with AIBO. However the number and the complexity of the detected T-patterns are similar, the structures are different suggesting that people distinguish AIBO from a living dog puppy in both age groups. Contrary to the suggestions of previous studies [5], analysis of T-patterns in human-AIBO interaction has showed that interaction with a robot is still different from the interactions detected while playing with a real puppy. Therefore our results show a clear advantage of using THEME in the analysis of the hidden behavioral structure of both cooperative and spontaneous interactions.

This work was supported by a grant from the Hungarian Science Foundation (OTKA T043763) and the Hungarian Academy of Sciences (F01/031). We would like to thank K. Soproni for allowing us to use her video recordings to run this analysis in the 1st study, F. Kaplan for providing AIBO for the 2nd study and M. Gácsi the owner of the dog puppy (Nudli) in the 2nd study for their help in data collection.

References:

1. Anolli, L., Duncan, S. Jr., Magnusson, M.S., Riva, G (eds) (2005) The Hidden Structure of Interaction: From Neurons to Culture Patterns. *Emerging Communication: Studies in New Technologies and Practices in Communication. 7.*
2. Bartlett, B., Estivill-Castro, V., Seymon, S. (2004) Dogs or robots: why do children see them as robotic pets rather than canine machines? *Proceedings of the fifth conference on Australasian user interface. Dunedin, New Zealand*
3. Beck, A.M., Edwards, N.E., Kahn, P., Friedman, B. (2004) Robotic pets as perceived companions for older adults. *IAHAIO People and animals: A timeless relationship, Glasgow, UK*
4. Grammer, K., Kruck, K.B., Magnusson, M.S. (1998) The courtship dance: patterns of nonverbal synchronization in opposite-sex encounters. *Journal of Nonverbal Behavior, 22, 3-29*
5. Kahn, P.H., Friedmann, B., Perez-Granados, D.R., Freier, N.G. (2004) Robotic pets in the lives of preschool children. *Conference on Human Factors in Computing Systems. Vienna, Austria*
6. Kaplan, F. (2001) Artificial Attachment: Will a robot ever pass Ainsworth's Strange Situation Test ? *Proceedings of Humanoids 2001: IEEE-RAS International Conference on Humanoid Robots*
7. Kerepesi, A., Jonsson, G.K, Miklósi Á., Topál, J., Csányi, V., Magnusson, M.S. (2005) Detection of temporal patterns in dog-human interaction. *Behavioural Processes, in press.*
8. Magnusson, M.S. (1996) Hidden Real-Time patterns in Intra- and Inter-Individual Behavior Description and Detection. *European Journal of Psychological Assessment, 12, 112-123*
9. Magnusson, M.S. (2000) Discovering hidden time Patterns in Behavior: T-patterns and their detection. *Behavior Research Methods, Instruments & Computers, 32, 93-110*
10. Melson, G.F., Kahn, P., Beck, A., Friedman, B., Roberts, T. (2004) Children's understanding of robotic and living dog. *IAHAIO People and animals: A timeless relationship, Glasgow, UK*
11. Millot, J.L., Filiâtre, J.C. (1986) The Behavioural Sequences in the Communication System between the Child and his Pet Dog. *Applied Animal Behaviour Science, 16, 383-390*
12. Millot, J.L., Filiatre, J.C., Gagnon, A.C., Eckerlin, A., Montagner, H. (1988) Children and their pet dogs: how they communicate. *Behavioural Processes, 17, 1-15*
13. Mitchell, R.W. and Thompson, N.S. (1993) Familiarity and the rarity of deception: two theories and their relevance to play between dogs (*Canis familiaris*) and humans (*Homo sapiens*), *Journal of Comparative Psychology, 107,, 291-300*
14. Rooney, N.J. Bradshaw, J.W.S. Robinson, I.H. (2001) Do dogs respond to play signals given by humans? *Animal Behaviour, 61, 715–722*
15. Turner, D.C., Ribi, F.N., Yokoyama, A. (2004) A comparison of children's behaviour toward a robotic pet and a similar sized, live dog over time. *IAHAIO People and animals: A timeless relationship, Glasgow, UK*

Using Theme to Analyze Convention-based Face-to-Face Interaction

Starkey Duncan
Department of Psychology, University of Chicago, Chicago, USA

Abstract

Mutually contingent action by the participants is a defining feature of face-to-face interaction. That is, each participant's action is influenced, at least in part, by one or more preceding actions by the partner. In analyzing sequential action patterns, Theme, originally designed for analyzing face-to-face interaction, is the analytic tool of choice. In addition, it is possible to view much of face-to-face interaction as convention based or rule governed. Developing hypotheses concerning rules in interaction is greatly facilitated by Theme patterns but requires the investigator to consider and integrate groups of patterns, combining them in various ways, as opposed to focusing exclusively on single patterns. The techniques of Theme-based analyses of rule-governed analyses are briefly described.

Keywords

Theme, face-to-face interaction, convention, rule

1 Introduction

A defining characteristic of face-to-face interaction is that it is composed of mutually contingent action by the participants. That is, each participant's actions are influenced, at least in part, by preceding actions. It is an empirical issue to identify the one or more preceding actions by the partner, the participant, or both that influence a participant's action at some defined point in the stream of interaction.

In any event, the ongoing process of mutually contingent action requires that in interaction research the data carry sequential information, and that analyses can accommodate sequential patterns. Interaction simply cannot be productively studied using simple counts of actions [4]. This sequential pattern analysis is, of course, precisely what Theme was designed to accomplish. Although it is now applied to a wide variety of fields, this discussion will focus on its original purpose: analyzing face-to-face interaction.

(For simplicity, I shall refer only to two-person interactions. However, it will be obvious that the principles and techniques described can be applied to interactions involving any number of participants.)

In addition, much of interaction can be viewed as convention based or rule governed. That is, strictly speaking, interaction may be said to have a grammar, just as does language. Of course, the type and complexity of an interaction grammar may be quite different from a linguistic grammar.

2 Interaction Structure

We shall use the term *structure* to refer to a hypothesized convention operating in an observed interaction. The structure would apply to each instance of the interaction or part of the interaction. An interaction can include a number of conventions that operate simultaneously or sequentially, each requiring its own structure. For example, a conversation between adults might include conventions for greeting, forms of address, relative space between participants, appropriate topics, general length, taking turns, and closing. In addition, there may be a higher level structure specifying how these different conventions are arranged within the conversation as a whole.

Very broadly, a structure includes three main components (a) a set of appropriate actions, termed elements in this discussion, (b) a set of rules that link the elements sequentially, and (c) the participants who may legitimately carry out these actions. Discussion of the use of Theme in developing hypotheses of structure is necessarily preceded by a brief description of some major components of rule-governed interaction.

2.1 Elements

A structure contains for each participant at least one *element* that describes the action(s) that the participant may appropriately take at specified points in the stream of interaction. An element may contain one or more actions. In the case of multiple actions, all the included actions must have the same effect on the course of the interaction. Because of this, these actions are hypothesized to be *interchangeable* components of the element [1]. The element occurs if any one or more interchangeable actions occur.

2.2 Rules

In addition to elements, a structure also contains a set of rules specifying appropriate sequences of elements. We shall focus on two types of rules: obligatory and optional.

Obligatory rules designate points where an element must or must not occur. For example, in the structure for exchange of speaking turns in conversations between adults [2,3], the speaker's gesticulating while talking marked points at which the auditor must not attempt to take the turn.

Optional rules designate points where a participant may legitimately choose between two or more alternative elements. That is, the participant may choose to perform and action or not to perform it. For example, in the speaking turn structure [2,3], an auditor may or may not take the turn in response to the display of a turn signal by the speaker. Alternatively, another convention may specify that, at a given point in the interaction, the participant may choose between two legitimate actions, such as saying either yes or no.

Although by definition a specified choice is entirely legitimate within the convention, two constraints apply to optional rules. First, the participant must exercise the choice at that particular point in the interaction. That is, the convention requires the participant to make a choice, taking one of the alternatives. In the speaking-turn example, upon display of the speaker turn signal, the auditor cannot avoid choosing between acting to take the turn or not.

The second constraint stipulates that the available alternatives in an optional rule define a point at which a change in the course of interaction may occur. For example, if the auditor acts to take the turn upon display of the speaker turn signal, the interaction will take a different course: the previous auditor will now be the speaker. If the auditor does not so act, then the current speaker will retain the turn. This is, in essence, the distinction between interchangeable actions and optional actions. Interchangeable actions do not change the course of the interaction, while optional actions do.

3 Sequential Probabilities

The problem for exploratory analyses of interaction structure is to develop and evaluate tentative hypotheses of both elements and rules. Both the composition of elements and the rules connecting these elements are evaluated by examining the relevant sequential probabilities.

In all data containing sequential information, two probabilities can be calculated. More familiar is the *consequent probability*: the probability that, given an action *A*, action *B* occurs after it. That is, if *A* occurs, then the probability that *B* will follow is calculated. This probability is widely used because it indicates the degree of predictability of *B*, given an occurrence of *A*.

Less familiar but crucial to these analyses is the *antecedent probability:* the probability that if *B* occurs, *A* precedes it. For example in the speaking-turn structure it was found that virtually all smooth exchanges of the turn were preceded by the turn signal.

These two probabilities are jointly used to evaluate a possible relation between two elements in an interaction structure. In obligatory sequences, both consequent and antecedent probabilities are expected to be high, optimally 1.00. If action *A* occurs, action *B* always follows.

In optional sequences, the probability of an antecedent must be similarly high, optimally 1.00, while the probability of a consequent is not specified. The consequent probability of a particular alternative action depends on the rate at which the participant chooses to exercise it. That is, given a display of the turn signal, one auditor may act to take the turn at a very high rate, while another auditor may be more reticent.

It will be apparent that analyses of interaction structure begin with the probability of an antecedent. If it is sufficiently high, there is evidence supporting some structure or part of a structure. Given the high probability of an antecedent, the probability of a consequent can be calculated to determine whether the potential rule is obligatory or optional.

4 Theme Analyses

Analysis of complex or large data sets is greatly facilitated by Theme. Theme identifies sequential patterns (t-patterns) that are necessary for developing hypothesized structures. Developing these hypotheses is further facilitated by the extensive descriptive statistics Theme provides with each t-pattern. These include the critical consequent and antecedent probabilities.

The hypothesis of an obligatory rule is relatively straightforward. Two actions *A* and *B* are consistently associated each other in t-patterns. Scrutiny of the accompanying descriptive statistics indicates that *B* consistently follows *A*, and *B* is consistently preceded by *A*.

However, in the case of interchangeable actions and optional rules, the t-patterns returned by Theme cannot be used directly in developing structures. T-patterns do not directly yield information on these phenomena. Developing structures concerning interchangeable actions and optional rules requires the investigator to compare, contrast, and combine t-patterns in order to obtain strong results and thus firm hypotheses of structures.

A t-pattern does not contain a set of interchangeable actions. Rather, interchangeable actions will be suggested by a set of t-patterns in which each pattern shows a relationship between different actions (e.g., *A* and *B*) and a subsequent action (*D*). In this case, one pattern might show a sequential relation ($A \rightarrow D$). Another pattern shows a relation ($B \rightarrow D$). This suggests the possibility that *A* and *B* are interchangeable actions. Once a set of interchangeable actions has been tentatively identified—in this case *A* and *B*—the investigator may check the antecedent probability: D is always, or almost always, preceded by either *A* or *B*. Thus, the requirement of a strong probability of an antecedent is satisfied. The number of interchangeable actions may, of course, be greater than two. In any event, careful comparison of t-patterns is required.

In the case of optional rules the process involves the same sort of comparison process but reversed. In this case, there may be two or more t-patterns in which different alternative actions are associated with a single preceding action. For example, actions *D* and *E* are preceded in t-patterns by action *A*. This suggests that, given the occurrence of *A*, either *D* or *E* may legitimately follow. This possibility can be tested by calculating the probability of an antecedent *A*, given either *D* or *E*.

The study of speaking turns in conversations provides an example of both interchangeable action and optional rules. The speaker turn signal involved five interchangeable actions (very briefly, a certain intonation contour, completion of a grammatical clause, paralinguistic drawl, end of a gesticulation, and a sociocentric sequence, such as “you know” or “or something”). It was found that smooth exchanges of the turn were preceded by at least one of these alternative actions. Auditor attempts to take the turn in the absence of a speaker turn signal tended to result in simultaneous turns (a speech overlap).

However, the speaker turn signal was not always followed by a smooth exchange of the turn. In this way, the hypothesis of an optional rule was suggested: the auditor may choose to take the turn or not to take the turn in response to the speaker’s display of the turn signal. If the auditor chose to take the turn, this action was always preceded by the turn signal.

It will be apparent that t-patterns identify sequential relationships between actions, but the investigator's further integration of Theme results is necessary to deal with the complexities that arise in the hypothesizing of rule-governed interaction.

In the end, a structure is evaluated by fitting it to the transcribed data. In the case of turns, each display of the turn signal and each auditor attempt to take the turn were tested evaluated in terms of the structure. If the structure fits a high percentage of the data (optimally 100%), the inferential work leading to the structure is successful. If there is not a high percentage of fit, then the investigators work is not done. T-patterns must be further scrutinized for possible alternative actions and optional rules.

Thus, t-patterns are invaluable in studying rule-governed face-to-face interaction, but integrating the information in a set of t-patterns is essential in pursuing this type of research.

This research was supported in part by National Institute of Mental Health grant (MH-38344) and The McCormick-Tribune Foundation.

References

1. Duncan, S. D., Jr., & Farley, A. M. (1990). Achieving parent-child coordination through convention: Fixed- and variable-sequence conventions. *Child Development, 61*, 742-753.
2. Duncan, S. D., Jr., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
3. Duncan, S. D., Jr., Fiske, D. W., Denny, R. Kanki, B. G., & Mokros, H. B. (1985). *Interaction structure and strategy*. New York: Cambridge University Press.
4. Duncan, S. D., Jr., Kanki, B. G., Mokros, H. B., & Fiske, D. W. (1984). Pseudounilaterality, simple-rate variables, and other ills to which interaction research is heir. *Journal of Personality and Social Psychology, 46*, 1335-1348. (Republished: H. M. Mokros (Ed.), *Interaction & identity: Information and behavior. Vol. 5* (pp. 71-96). New Brunswik, NJ: Transaction Press, 1996.)

Assessment of play interaction in football through SOF-CODER and hidden patterns

G.K. Jonsson[1], M.T. Anguera[2], A. Blanco-Villaseñor[2], J.L. Losada[2], A. Hernández-Mendo[3], T. Ardá[4], O. Camerino[5], J. Castellano[6]

1 *Human Laboratory Institute, Reykjavík, Iceland & Department of Psychology, University of Aberdeen, UK*

[2] *Methodology of Behavioral Sciences, University of Barcelona, Barcelona, Spain*

[3] *Department of Social Psychology, University of Málaga, Málaga, Spain*

[4] *INEF-Galicia, University of A Coruña, A Coruña, Spain*

[5] *INEFC-Lleida, University of Lleida, Lleida, Spain*

[6] *IVEF, University of Basc Country, Vitoria, Spain*

Abstract

In this paper we study the actions between players in collective sports, mainly football. In this sport there are a diversity of interactive structures and a relevant influence of the context. Our main approach is the detection of hidden patterns in the multiple actions that emerge during the game, and considering different aspects of the match, like team, laterality, zone, balloon contacts, and so on.

Our interest is focused on habitual activity of professional players that is produced in a non-restrictive context, and consequently it adopts the approach of observational methodology. The observational design is nomothetic (two teams), punctual (one game) and multidimensional (criteria of the field format). The observational instrument is SOF-3, built as a mixed instrument that combines a field format with a category systems for game events. It is an *ad hoc* instrument that considers the game actions of both teams, each recorded from the same criteria.

Keywords

T-patterns, positive interaction, negative interaction, observational record, football.

1 Introduction

The aim of this paper is the study of the hidden structures underlying an interactive situation, notably a game of football. In other words, the way in which temporal patterns are able to reveal those aspects of playing interaction which are not immediately observable. Every interactive flow is governed by behavioral structures of varying stability that can be visualized by obtaining temporal patterns (*T-patterns*).

Social interaction is a very complex phenomenon. The assumptions underlying social interaction may be considered as a basic perspective for the study of social relationships, although there is no unified theory for the study of such processes. The most important of these are the inevitable multiplicity of individuals involved, the diversity of existing interactive structures, the effect of the context and interactive content, and both inter- and intra-session permanence and changeability, among many others. The study of social interaction is undoubtedly difficult from a methodological point of view, and in this regard the contribution of observational methodology proved decisive in all situations [1].

There are many specific questions that may be researched in the context of social interaction, as the latter constitutes an extensive framework from which varied elements may be selected. One of the specific questions that is proving to be highly fruitful, and which constitutes the *leitmotif* of the present paper, is the study of the hidden structures underlying an interactive situation such as a game of football. In other words, the way in which temporal patterns are able to reveal those aspects of social interaction which are not immediately observable. Every interactive flow is governed by behavioral structures of varying stability that can be visualized by obtaining temporal patterns (*T-patterns*). These temporal patterns, detected by means of the powerful algorithm of the THEME software, developed by Magnusson [11, 12], constitute the object of study here.

THEME proves to be extraordinarily productive and fruitful in the study of any of the multiple facets or fields of application of social interaction, as has been the case in the field of team sport [2, 5, 6, 9, 10].

2 Method

2.1 Participants

This current study is a part of an ongoing and broader research concerning the analysis of all games played by Barcelona Football Club (Spain) during the National League Championship and the Champions' League for several seasons. From these games we have selected a one particular match (Barcelona-Valladolid) from the period 2000-2001, in order to distinguish the little differences between both teams. The score at the end of first period was 0-0, and at the end of second period 3-1.

2.2 Instruments

The observational instrument is SOF-3, one of a series of instruments developed for the study of football interaction. The SOF-3 is a combination of field format and a category system developed into an agreement between Department of Methodology of Behavioural Sciences (University of Barcelona) and Barcelona Football Club [6,7,8].

SOF-3 includes the following criteria: Competition, times of the game, player, lateral spaces, reception area, pass area, balloon contact, interruption, interception, shoot, and scoreboard. Duration of each action are also recorded.

2.3 Procedure

The observational design is nomothetic (two teams), punctual (one game) and multidimensional (these dimensions show correspondence with criteria of the observational instrument).

Observational records were obtained using Excel, with binary codes (1 and 0) enabling detailed coding of digitized video files (Figure 1). The record of this play consists of 1074 rows.

The Excel files have been transformed to .txt files, in order to analyze the collected data. In this file, we have a Dataname column, a T column (with the temporal frames), and the different codes in adequate columns.

Figure 1. Excel file with the observational instrument SOF-3.

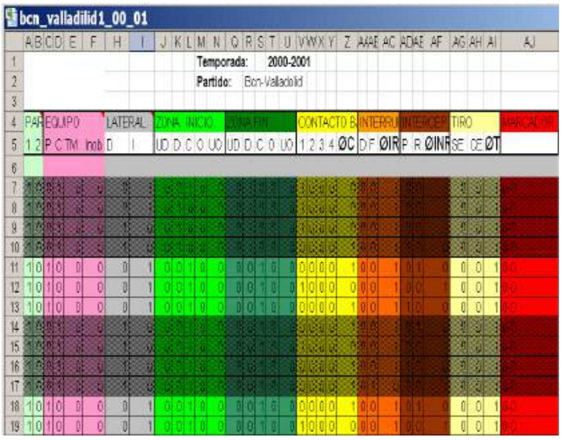
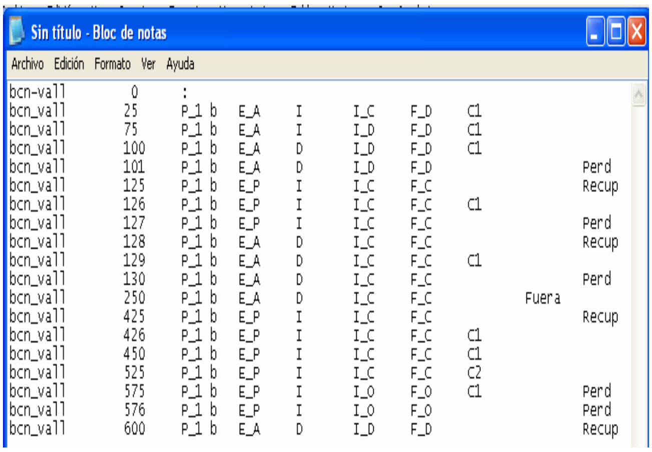


Figure 2. Fragment of *.txt* file (transformation of Excel file).

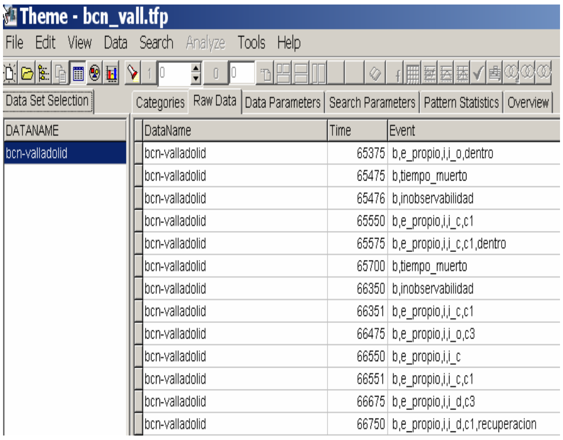


```
Sin título - Bloc de notas
Archivo  Edición  Formato  Ver  Ayuda
bcn-vall    0    :
bcn_vall    25   P_1 b  E_A  I  I_C  F_D  C1
bcn_vall    75   P_1 b  E_A  I  I_D  F_D  C1
bcn_vall    100  P_1 b  E_A  D  I_D  F_D  C1
bcn_vall    101  P_1 b  E_A  D  I_D  F_D              Perd
bcn_vall    125  P_1 b  E_P  I  I_C  F_C              Recup
bcn_vall    126  P_1 b  E_P  I  I_C  F_C  C1
bcn_vall    127  P_1 b  E_P  I  I_C  F_C              Perd
bcn_vall    128  P_1 b  E_A  D  I_C  F_C              Recup
bcn_vall    129  P_1 b  E_A  D  I_C  F_C  C1
bcn_vall    130  P_1 b  E_A  D  I_C  F_C              Perd
bcn_vall    250  P_1 b  E_A  D  I_C  F_C       Fuera
bcn_vall    425  P_1 b  E_P  I  I_C  F_C              Recup
bcn_vall    426  P_1 b  E_P  I  I_C  F_C  C1
bcn_vall    450  P_1 b  E_P  I  I_C  F_C  C1
bcn_vall    525  P_1 b  E_P  I  I_C  F_C  C2
bcn_vall    575  P_1 b  E_P  I  I_O  F_O  C1       Perd
bcn_vall    576  P_1 b  E_P  I  I_O  F_O              Perd
bcn_vall    600  P_1 b  E_A  D  I_D  F_D              Recup
```

The .txt file is then imported into the THEME software in order to detect hidden patterns. The columns are: Dataname, T (time in frame numbers), and third column includes all codes separated by commas, in order to detect hidden patterns. Figure 3 displays a fragment of a data file after importing into THEME.

The codes of observational instrument SOF-3 are included in the .vvt file. The criteria of the instrument are sorted as columns, and respective codes of each one in rows. Figure 4 displays these codes as viewed in the THEME software.

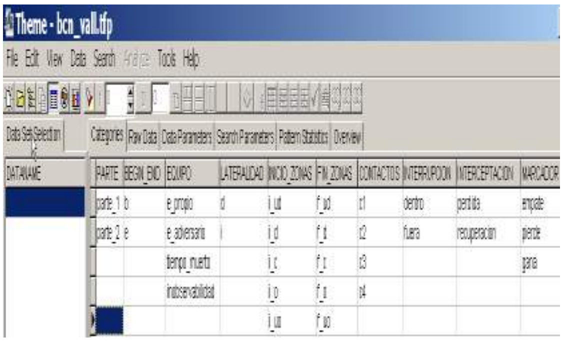
Figure 3. Data displayed in THEME.



Theme - bcn_vall.tfp

DataName	Time	Event
bcn-valladolid	65375	b,e_propio,i,i_o,dentro
bcn-valladolid	65475	b,tiempo_muerto
bcn-valladolid	65476	b,inobservabilidad
bcn-valladolid	65550	b,e_propio,i,i_c,c1
bcn-valladolid	65575	b,e_propio,i,i_c,c1,dentro
bcn-valladolid	65700	b,tiempo_muerto
bcn-valladolid	66350	b,inobservabilidad
bcn-valladolid	66351	b,e_propio,i,i_c,c1
bcn-valladolid	66475	b,e_propio,i,i_o,c3
bcn-valladolid	66550	b,e_propio,i,i_c
bcn-valladolid	66551	b,e_propio,i,i_c,c1
bcn-valladolid	66675	b,e_propio,i,i_d,c3
bcn-valladolid	66750	b,e_propio,i,i_d,c1,recuperacion

Figure 4. Category system.



The THEME program allows a detection of complex repeated temporal patterns, even when a multitude of unrelated events occur in between components of the patterns, which typically makes them invisible to the naked eye. The basic assumption is that the temporal structure of a complex behavioural system is largely unknown, but may involve a set of particular type of repeated temporal patterns (T-patterns) composed of distinguishable event-types. Essentially, within a given observation period, if two actions, A and B, occur repeatedly in that order or concurrently, they are said to form a T-pattern (AB) if found more often than expected by chance, assuming as h0 an independent distributions for A and B, there is approximately the same time distance (called critical interval, CI) between them. More complex T-patterns are gradually detected as patterns of simpler already detected patterns through a hierarchical bottom-up detection procedure [11, 12].

3 Results

Analysis of transcribed records has revealed a high number of repeated complex temporal structures within the game events of the particular match analyzed. Figure 5 displays an event frequency chart and Figure 6 the event time plot from the analyzed game.

Figure 5. Event frequency chart.

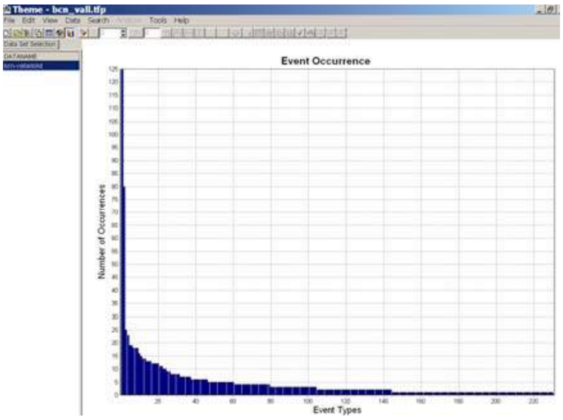
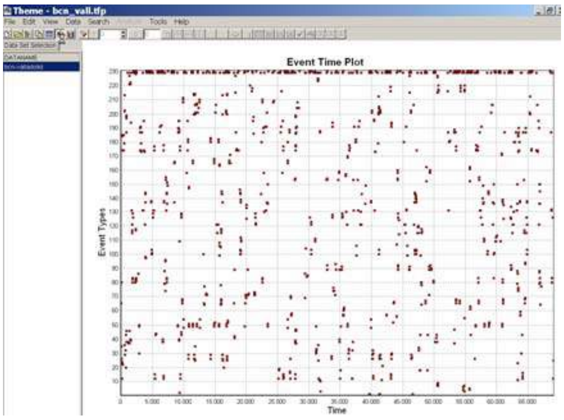


Figure 6. Event time plot.



The data show that a high number of temporal patterns exist in football. The number, frequency and complexity of the detected patterns, indicates that sport behavior is more synchronized than the human eye can detect. This synchrony was found to exist on different levels, with highly complex time structures that extended over considerable time spans within performances with patterns occurring in both cyclical and acyclical fashion [4]. An example of a detected pattern, involving 3 event types, is displayed in Figure 7 and a more complex, involving 8 events, in Figure 8.

Figure 7. An example of a detected pattern, involving 3 event types.

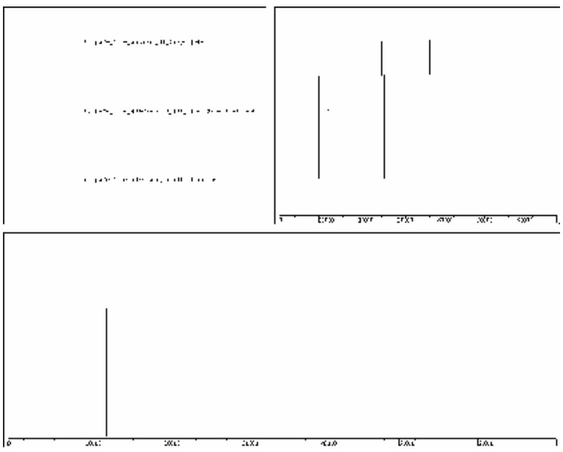
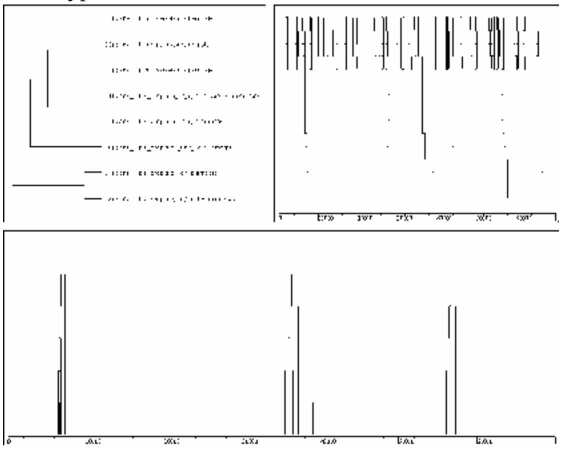


Figure 8. An example of a detected pattern, involving 8 event types.



The pattern displayed in Figure 7 occurred 4 times during first half of the match, the duration was 5338 frames (8% of the observation period) and describes the following sequence of action: When the own team plays in defensive zone, the opposite team retrieves the balloon and plays in his ultra-defensive zone (with the same goals in scoreboard). . The pattern displayed in Figure 8 occurred 3 times during first half of the match, the duration was 6454 frames (9% of the observation period) and describes the following sequence of action: There is an alternance between situations of inobservability and some periods of time without playing (later fouls); also, the own team plays from the middle zone until ultra-defensive zone having very short balloon possessions, and later plays in offensive zone, where the opposite team commits a foul in favor own team..

Supported by DGES, Spain (Grant N. BSO2001-3368)

References:

1. Anguera, M.T. (2005). Microanalysis of T-patterns. Analysis of simmetry/assimetry in social interaction. In L. Anolli, S. Duncan, M. Magnusson & G. Riva (Eds.), *The hidden structure of social interaction. From Genomics to Culture Patterns* (pp. 51-70). Amsterdam: IOS Press.
2. Anguera, M.T.; Ardá, T. (2003). Hidden patterns in seven-a-side football: how do children interact during play? *XIth European Conference on Developmental Psychology*, Milan, Italy, June.
3. Anguera, M.T.; Blanco-Villaseñor, A.; Losada, J.L.; Ardá, T.; Camerino, O.; Castellano, J.; Hernández-Mendo, A.; Jonsson, G.K. (2003). Match & player analysis in soccer: Computer coding and analytic possibilities. *International Journal of Computer Science in Sport (e-Journal), 2* (1), 118-121.
4. Anguera, M.T. & Jonsson, G.K. (2002). Detection of real-time patterns in sports: Interactions in football. Third Meeting of the European Research Group on *"Methodology for the analysis of social interaction"*. Milan, Italy, June.
5. Anguera, M.T. ; Jonsson, G.K. (2003). Detection of real time patterns in sport: Interactions in football. *International Journal of Computer Science in Sport (e-Journal), 2* (2), 118-121.

6. Anguera, M.T.; Jonsson, G.K.; Magnusson, M.; Losada, J.L.; Blanco-Villaseñor, A. (2004). Variability of temporal patterns intra-team and between teams in football. *Fourth Meeting of European Research Group Methodology for the analysis of social interaction.* Barcelona, Spain, June.
7. Blanco-Villaseñor, A.; Castellano, J.; Hernández-Mendo, A.; Anguera, M.T.; Losada, J.L.; Ardá, A.; Camerino, O. (2004). SOF-5: Instrumento de codificación y registro de la interacción en fútbol. *III Congreso Vasco del Deporte. Socialización y Deporte / Kirolaren III Euskal Biltzarra. Sozializazioa era Virola*. Vitoria, Spain, November
8. Anguera, M.T.; Blanco-Villaseñor, A.; Losada, J.L.; Ardá, T.; Camerino, O.; Castellano, J.; Hernández-Mendo, A.; Jonsson, G.K. (2004). SOF-4: Instrumento de registro y codificación en el fútbol. Presentación Multimedia. *II Congreso Internacional de Actualización en Psicología del Deporte*. Buenos Aires, Argentina, March.
9. Bloomfield, J.; Jonsson, G.K.; Polman, R.; Houlahan, K.; O'Donoghue, P. (2005). Temporal patterns analysis and its applicability in soccer. In L. Anolli, S. Duncan, M. Magnusson & G. Riva (Eds.), *The hidden structure of social interaction. From Genomics to Culture Patterns* (pp. 237-251). Amsterdam: IOS Press.
10. Jonsson, G.K.; Bjarkadottir, S.H.; Gislason, B.; Borrie, A.; Magnusson, M.S. (2003). Detection of real-time patterns in sports: interactions in football. In C. Baudoin (Ed.), *L'éthologie appliquée aujourd'hui (Applied Ethology Today)*, Volume 3 - Ethologie humaine. Levallois-Perret, France: Editions ED.
11. Magnusson, M.S. (1996). Hidden real-time patterns in intra- and inter-individual behavior. *European Journal of Psychological Assessment, 12* (2), 112-123.
12. Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers, 32* (1), 93-110.

Scientific Tours

The following behavioral research facilities in and around Wageningen opened their doors during the conference.

Experimental Zoology Group (Department of Animal Sciences, Wageningen University)

The Experimental Zoology group applies mechanical principles to understand the architecture and functions (in particular locomotion) of animals. The tour will give an overview of a selection of ongoing projects on fish swimming and bone remodeling. Among the techniques that will be demonstrated are high-speed video analysis, digital particle image velocimetry (DPIV) and computational fluid dynamics (CFD) to analyze the body motions and the related flow around swimming fish. For more information: www.zod.wau.nl/ezo/index.html.

Laboratory of Entomology (Department of Plant Sciences, Wageningen University)

One of the research themes of this Laboratory is chemical ecology of insects, directed at understanding how chemical cues are exploited by insects to find their food source. This includes herbivorous insects, carnivorous insects that prey on herbivores, and haematophagous insects. Examples are malaria mosquitoes that exploit human odors, parasitic wasps that exploit their host's pheromones or SOS-odors from plants and butterflies or beetles that exploit plant odors. The research extends from production of the chemicals, through an analysis of their perception by the insects towards the behavioral responses. The equipment used for this includes wind tunnels, olfactometers and electrophysiological and neurobiological setups. The equipment is used on the one hand for assessing the active components from complex mixtures of compounds and on the other for gaining knowledge in understanding the biology of chemical information exploitation by insects. For more information: www.dpw.wau.nl/ento/english.

Green Dino Virtual Realities

Green Dino Virtual Realities is the Dutch specialist in Intelligent Instruction Systems for training and education. We started in 1992 as a spin-off from Wageningen University. Through research and development Green Dino offers state of the art technology, tailored to the client's needs and tested for validity. This research is done in cooperation with Delft University and Twente University. We develop the Dutch Driving Simulator, a new tool in category B driver education. With the aid of this compact simulator students learn to perform different driving tasks. This is done under supervision of a Virtual Driving Instructor that adapts its instruction to the performance level of the student. The data form automatic registration of events is used to improve the intelligence of the system. The architecture of the driving simulator is generic and can be used for all kinds of procedure training for operators. The automated instruction with a Virtual Instructor is the key for cost effective training! The performance of simulator students is better than students who followed lessons in a car with real instructor. The Dutch Driving Simulator was awarded with the National ICT Award MKB 2004. During the scientific tour we will demonstrate the Dutch Driving Simulator and give insight in the architecture. For more information: www.dutchsimulators.com.

Netherlands Institute of Ecology (NIOO-KNAW)

Bird character, insect learning, global change, and favorite-food foraging to name just a few of the ingredients of the tour at the Netherlands Institute of Ecology (NIOO-KNAW) in Heteren, near Wageningen. The bird *Parus major* is one of the model species, especially for research on the 'personality' of animals and the ecological consequences thereof. Next to behavioral studies, also the effect of global climate change on ecosystems' food chains are an important research issue. Other species studied are water birds - for which the Netherlands are invaluable as breeding or wintering grounds. How are their interactions with water plants structured, and how does that influence their feeding behavior? Thirdly, the insects are considered. Their interactions with plant life and each other can not easily be overestimated, as they are influencing the food chain several steps down the line. And then we have not even mentioned the DNA lab, the soil research, and our own FBI (fungal-bacterial interactions). We'll visit one out of NIOO's three ecological centers in the Netherlands - NIOO forms the largest group of fundamental ecologists in the country. For more information: www.nioo.knaw.nl.

Maritime Research Institute Netherlands (MARIN)

The Maritime Research Institute Netherlands (MARIN) is specialized in consultancy and research for the maritime industry. The institute was founded in 1932, today it is a company with 250 employees and state-of-the art model test facilities and simulation tools. MARIN executes research and design studies for the offshore industry, ship builders, shipping companies, navies, governments and other companies in the maritime world. MARIN's nautical centre MSCN is executing studies for the design of ports and fairways, set up of maritime operations and the safety of shipping. Furthermore they provide training for pilots and captains of sea going vessels. For this purpose they have two full mission maneuvering simulators and four part task simulators.

Occasionally these simulators are also used to measure the workload of navigators when executing a specific task. During the visit to MARIN the simulators will be demonstrated. For more information: www.marin.nl.

Numico Research

Numico Research is the innovative life science organization of Royal Numico serving among others Nutricia and Milupa. Numico Research focuses on the core activities of the company: Baby Food and Clinical Nutrition. The role of Numico Research is to develop, test and validate hypotheses resulting into new scientific insights, principles and concepts. These will then be transformed into safe and responsible specialized nutritional products. The development of innovative product concepts requires a thorough study and understanding of the mechanisms of action within the involved metabolic pathways. Thus, basic *in vitro* and *in vivo* model studies are being complemented with clinical efficacy and safety studies. Over 60 bio test systems are used at molecular, cell and tissue level in combination with clinical trials in a world-wide network. Numico Research has an extensive range of laboratories and a pilot plant. Activities include: discovery of new ingredients and product opportunities, identification of new biological activities and health benefits, and preclinical research. For more information: www.numico.com.

Agrotechnology & Food Innovations BV

A&F is an applied research institute within Wageningen University and Research Centre. The department of Consumer & Market Insight studies eating and drinking behavior of humans and the factors that influence this behavior. The department participates in the Wageningen Centre for Food Sciences. During the lab tour, we show our sensory/physiological facilities where sensory measurements are performed simultaneously with physiological measurements. In addition, we will show our instrumentation to gain insight into processes as they occur in the mouth during oral processing of foods. Finally, we will present our plans for the new research facilities (opening no later than December 2005), in which we aim at establishing an interactive future-oriented research centre where factors underlying food choice and eating and drinking behavior are investigated. The factors are studied in a behavioral context, which ensures validity larger than is usually encountered in many current research paradigms. The facilities will be housed in a new research centre, which will include a restaurant equipped with video surveillance and possibilities to simulate different environments and excellent sensory and physiology laboratories. The centre is set up with different partners and will operate on a 'facility-sharing' basis with its research partners. For more information: www.agrotechnologyandfood.wur.nl.

Experience Lab, Noldus Information Technology BV

Noldus Information Technology develops innovative software tools and integrated solutions for behavioral research. In our new office building we have constructed a state-of-the-art Experience Lab, equipped with the latest computer and digital video technology. The lab has been designed for human-computer interaction studies, multimodal user experience research, focus groups, and usability testing of hardware and software. The Experience Lab includes a user room and observation room, separated by a one-way screen, and a conference room. By unobtrusively watching and recording user behavior, we can measure the usability of product prototypes from a user's point of view. The user room can easily be converted from an office environment into a living room, thus allowing us to evaluate a large variety of applications and devices in various environments. Endpoints of usability tests assess the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments (ISO 9241). During the tour you will see demonstrations of remotely controlled dome cameras and IP-based video cameras, multi-channel video and audio recording, digital screen capture, event logging and multimodal data integration. The facility is used by our internal Research & Development department, but is also available for rent. Come over and have a look inside the place where the next generation of The Observer, EthoVision and other applications are tested and improved. For more information: www.noldus.com/usability.

Sponsors

Measuring Behavior 2005 gratefully acknowledges the financial support from the following organizations:

Wageningen University and Research Centre

Dutch-Belgian Society for Neuroscience (EKN)

SenterNovem / IOP Human-Machine Interaction

City of Wageningen

Johnson & Johnson Pharmaceutical Research and Development, L.L.C.

Exhibitors

Biopac Systems, Inc. (www.biopac.com)
Biopac develops, manufactures, and supports data acquisition and analysis systems for life science research and education, including wireless Ethernet and USB analog to digital converters, high-speed acquisition solutions, amplifiers, transducers, electrodes and accessories.

Brain Products (www.brainproducts.com)
Solutions for neurophysiological research, BrainVision product line, fMRI-compatible EEG amplifiers.

Frank Buschmann International (www.fbiscience.net)
FBI-Science provides research equipment for behavioral experiments with animals and humans. Specialized in the development of modern software and hardware, FBI-Science offers the whole range of technology that is necessary to run an experiment. Computer hardware, wireless network solutions, databases, and online analysis tools, experimental animals/subjects administration packages, programs for easy design of psychometric testing and stimuli delivery, multimedia operant learning chambers and much more. Products are developed and tested by scientists providing you with much expertise and unique ease of operation.

Data Sciences International (www.datasci.com)
DSI provides complete systems for monitoring and collecting data from conscious, freely moving laboratory animals. No wires or tethers are needed, reducing stress to the animal. Parameters such as blood pressure, ECG, EEG, EMG, temperature and activity can be collected without the use of restraint and tethering.

Harlan (www.harlaneurope.com)
There is more to Harlan than the company you know as a reliable supplier of laboratory animals, isolators and diets. We're a specialist organization with all expertise needed to support your research.

Hogrefe & Huber Publishers (www.hhpub.com)
Hogrefe & Huber has published psychology books, journals, and tests for more than 50 years, including the journals *European Psychologist*, *Journal of Individual Differences*, and *Methodology*.

Institute for Zoo and Wildlife Research (www.izw-berlin.de)
The network TTM (Telemetry - Telematics - Monitoring) offers the storage telemetry systems ETHOSYS and ETHOLOC (activity and GIPS recording), and the wireless acceleration measurement system WAS for behavior recording in humans and animals.

Metris B.V. (www.metris.nl)
Metris is rapidly improving its developments in animal behavior measuring equipment. The product line has been extended to the integration of behavior and physiological data, ultrasound measurement and sleep analysis. Metris will increasingly provide you with the solution for automating your safety pharmacology and drug development processes.

Mini Mitter (www.minimitter.com)
Mini Mitter offers medical and scientific devices for ambulatory monitoring of physiological and behavioral parameters. Our small animal telemetry devices provide temperature, heart rate and activity data and don't require batteries or refurbishment.

NewBehavior (www.newbehavior.com)
NewBehavior provides systems for the fully automated behavioral and cognitive screening of animals. We offer solutions for free-ranging animals and laboratory animals housed individually as well as in social groups.

Noldus Information Technology (www.noldus.com)
Noldus Information Technology develops software and instrumentation for recording and analyzing human or animal behavior. Besides standard software packages, Noldus also provides integrated data acquisition and analysis systems, including computer hardware, audio and video components, wireless network technology, as well as complete behavioral lab setups. Well-known products include The Observer, EthoVision, PhenoTyper, UltraVox, MatMan and Theme. Noldus serves its customers in 75 countries through its offices across Europe and USA, and a network of international distributors.

SensoMotoric Instruments (www.smi.de)
SMI, founded in 1991, is the world's leading provider of eye tracking solutions for the scientific communities in fields including vestibular, ophthalmologic, psychological, reading, neurological, human factors, sports, ergonomic and marketing research as well as in medical diagnostic and therapy. SMI´s iVIEW and 3DVOG systems are applied in a broad range of environments from laboratories over driving and airplane simulators, shopping malls, virtual environments to cars and planes. SMI services customer needs and support through its Berlin-based headquarters, Boston (USA) subsidiary and various distributors throughout Europe and Asia.

Tracksys Ltd (www.tracksys.co.uk)
Tracksys are the distributors for Noldus Information Technology, SensoMotoric Instruments and SIMI Reality Motion Systems in the UK and Ireland. They integrate a variety of products, many of which are designed in-house, to provide complete research solutions. These solutions include the use of eye trackers, behavioral and motion analysis systems, portable labs and remote site 12-volt video equipment. Tracksys also build and supply tethered and wireless mobile device cameras, infra-red lights and infra-red mazes for studying transgenic animals.

Xsens Technologies (www.xsens.com)
Xsens is a supplier of 3D inertial motion sensors for human motion analysis as well as industrial applications. We supply miniature 3DOF orientation trackers (MTx) which measure body segment orientation and kinematic data. With a set of MTx's we demonstrate an innovative ambulatory full body motion capturing solution.

Author Index

Keyword Index

Page numbers in this index indicate the first page of the paper in which the keyword is mentioned.