Proceedings of

Measuring Behavior 2010

7th International Conference on Methods and Techniques in Behavioral Research

Eindhoven, The Netherlands August 24-27, 2010

Editors

Andrew Spink Fabrizio Grieco Olga Krips Leanne Loijens Lucas Noldus Patrick Zimmerman



2010

Bibliographical reference

Proceedings of Measuring Behavior 2010, 7th International Conference on Methods and Techniques in Behavioral Research (Eindhoven, The Netherlands, August 24-27, 2010)

Edited by A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

ISBN 978-90-74821-86-5

Copyright

© 2010 Noldus Information Technology bv P.O. Box 268

6700 AG Wageningen

The Netherlands

www.noldus.com

Printed in the Netherlands

No part of this publication may be reproduced, transcribed, stored in a retrieval system or translated into any other language, in whole or in part, in any form or by any means, without the prior written permission of the publisher.

Some of the product names referred to in this book are registered trademarks or proprietary names, even though specific reference to this fact is not always made in the text. Therefore, the appearance of a name without designation as proprietary is not to be considered as a representation by the publisher that it is in the public domain.

Table of Contents

| Welcome to Measuring Behavior 2010 | 1 |
|--|----|
| The Measuring Behavior Conferences | 2 |
| Measuring Behavior 2010: Highlights of the scientific program | 5 |
| Keynote Lectures | |
| Multi-Dimensional Recording in Social Primates: Method and Application | 7 |
| Measuring Behaviour in Human-Robot Interaction Studies Professor Kerstin Dautenhahn | 7 |
| Measuring What the Brain Does, What It Experiences, and What It Controls: Mobile Brain/Body Imaging | 8 |
| Symposia | |
| Symposium 1: Improving Sport Performance Symposium: Improving Sports Performance | 11 |
| Performance Monitoring in Equine Sports Patricia de Cocq, Carolien Munsters | 13 |
| Methodological Approach to Evaluate Interactive Behaviors in Team Games: An Example in Handball | 16 |
| The Measurement of the Visual Search Behavior in Sport. Can It Be a New Avenue into Talent Identification and Development? Geert J.P. Savelsbergh | 20 |
| Faster Marathon Times by Measuring Human Performance | 21 |
| Symposium 2: Large and Small Scale Physiological Recordings in Behavioural Context | |
| Symposium: Large and Small Scale Physiological Recordings in Behavioural Context | 24 |
| Epidural EEG Recordings Using Microchips in Behavioural Context Bettina Platt, Andrea Plano, Amar Jyoti, Gernot Riedel | 26 |
| Timed Behaviors in Mice Valter Tucci, Glenda Lassi, Patrick M. Nolan | 30 |
| Route Finding in a Complex Maze in Wild-Type and CA1 NR-1 KO Mice: Hippocampal Local Field Potentials, Single Units and Relationship with Behavior | 33 |
| Use of Behavioral Outcome to Assess Cognitive State | 36 |
| Simultaneous Measurement of Brain Activity, Physiology & Behavior in Large Animals Nadine Reefmann, Thomas Muehlemann, Martin Wolf, Beat Wechsler, Lorenz Gygax | |

| Platform for Ambulatory Assessment of Psycho-Physiological Signals and Online Data Capture | 1 |
|--|---|
| Experimental Design for Sternocleidomastoid Muscle Stress Measurement | 4 |
| Peripheral Arterial Tone As an Index of ANS Trade-Off | 8 |
| Using EEG Recordings to Examine the Relationships Between Sustained Attention and Types of Background Music in Individuals with ADHD | 2 |
| Chelsea Chew | |
| Psychophysiological Data Collection in an Organizational Setting: Studying Interaction Between the Manager and Subordinate During Performance Review Discussion | 5 |
| Mikko Salminen, Pentti Henttonen, Niklas Ravaja, Mikael Saarinen | |
| Extracellular Multi Unit Recording in Fear Conditioning in Mice Using a Telemetry Approach in an Automated Home Cage (DualCage) Environment | 8 |
| René F. Jansen, Anton W. Pieneman, Andries Ter Maat, Manfred Gahr, Oliver Stiedl | |
| Symposium 3: Measuring Surgical Behavior | |
| Symposium: Measuring Surgical Behavior | 9 |
| Teaching Arthroscopy: Analysis of Verbal Communication in the Operating Room | 1 |
| Methods for Automatic Statistical Modeling of Surgical Workflow | 4 |
| In-Vivo Measuring Surgical Workflow Activities in the OR | 5 |
| Learning Curve Assessment and Identification of Surgical Pitfalls of a New Hip Prosthesis Using Time-Action Analysis | 8 |
| Symposium 4: Monitoring of Social Interactions and Initial Signs of Conflicts in Isolated Groups during Long-Term Missions | |
| Symposium: Monitoring of Social Interactions and Initial Signs of Conflicts in Isolated Groups during Long-Term Missions | 0 |
| PSPA - Test for the Study of Individual Values and Mutual Perception in Small Group72 Vadim Gushin, Alla Vinokhodova | 2 |
| Electronic Partners That Diagnose, Guide and Mediate Space Crew's Social, Cognitive and Affective Processes | 3 |
| Automatic Mental Heath Assistant: Monitoring and Measuring Nonverbal Behavior of the Crew During Long-Term Missions | 7 |
| Symposium 5: The Significance of Voluntary Exploration to Monitor Emotional Behavior of Rodents | |
| Symposium: The Significance of Voluntary Exploration to Monitor Emotional Behavior of Rodents | 2 |
| Fear Conditioning in an Automated Home Cage (DualCage) Environment | 3 |
| Spontaneous Behavior in the Home Cage: a New Test Environment for Measuring Neuropathic Pain | 5 |

| An Automated Maze for Studying Working Memory and Decision-Making in Rodents |
|---|
| Genetic Dissection of Motor Activity Levels and Avoidance Behavior in The Home Cage; Translational Phenotypes for Mood Disorders89 Martien 1.H. Kas, Berend Olivier, Ametrude (I.G.) de Maoii-van Malsen |
| Understanding Exploratory Behavior Step by Step |
| Symposium 6: Unveiling Affective Signals |
| Symposium: Unveiling Affective Signals |
| Unveiling Affective Signals |
| Measuring Affective and Social Signals in Vocal Interaction |
| Facial EMG as a Tool for Inferring Affective States |
| Motor, Emotional and Cognitive Empathic Abilities in Children with Autism and Conduct Disorder |
| Mimicry as a Tool for Understanding the Emotions of Others |
| Social Signal Processing: Understanding Nonverbal Communication in Social Interactions |
| Automatic Measurement of Affect in Dimensional and Continuous Spaces: Why, What, and How? |
| Relative Affective Blindsight for Fearful Bodily Expressions |
| Full papers |
| Full paper session 1: Measuring the Behavior of Laboratory Animals |
| Microstructural Assessment of Rodent Behavior in the Hole-Board Experimental Assay |
| A Novel Conditioning Paradigm Enables the Dissociation Between the Formations of Context- and Cue-Dependent Memories of Drug Reward |
| Automation of Continuous Spontaneous Alternation to Increase the Throughput for In Vivo Screening of Cognitive Enhancers. Optimization of the Ethovision System for the Y-maze Test in Mice |
| Home Cage Testing of Decision-Making |
| Trainable, Vision-Based Automated Home Cage Behavioral Phenotyping |
| Refinement Through Better Animal Monitoring: How Behavioural Researchers Can Contribute |
| Natural Colour Preference in the Zebrafish (Danio rerio) |

| Full paper session 2: Statistical Techniques and Pattern Analysis | |
|--|----|
| Estimating Slope and Level Change in N=1 Designs | 58 |
| DyaDA: An R Package for Dyadic Data Analysis | 62 |
| The Algorithm for Detection of Fuzzy Behavioral Patterns | 66 |
| Machine Learning Techniques for Parameter Selection and Automated Behavioral Classification | 71 |
| Using Ideas of Kolmogorov Complexity for Studying Animal Behavioural Patterns | 74 |
| Ball Recovery in the Handball Tournament of the 2008 Beijing Olympic Games: Sequential Analysis of Positional Play as Used by the Spanish Team's Defence | 77 |
| DataPrism: a Tool for Visualizing Multimodal Data | 81 |
| Video Deep Tagging and Data Archiving in the Comparative Mind Database | 85 |
| Using Postural Control System Measures to Detect Hypovigilance | 89 |
| Full paper session 3: Measuring Gaze, Eve Tracking and Facial Expression | |
| Measuring Instant Emotions During a Self-Assessment Test: The Use of FaceReader | 92 |
| Comparison of Eye Tracking Systems with One and Three Cameras | 96 |
| Gaze Metrics for Efficient and Safe Operations of Hemodialysis | 00 |
| Automatic Recognition of Lower Facial Action Units | 04 |
| Facial Expression Classification Based on Local Spatiotemporal Edge and Texture Descriptors | 08 |
| Comfort and Acceptability Study of 3D devices | 12 |
| Eye Tracking Analysis: Application in a Case Study of a Fast Moving Consumer Goods Product | 15 |
| Physiological Measures, Eye Tracking and Task Analysis to Track User Reactions in User Generated Content | 18 |
| Measuring Reading Characteristics Using Eye Tracking and a Context Adapting Classification System | 23 |
| DogCam: a Way to Measure Visual Attention in Dogs | 26 |

| Full paper session 4: Psychological Measurements of Behavior |
|--|
| A Dyadic Approach for Measuring and Testing Agreement in Interpersonal Perception |
| Acquisition, Translativity and Transfer of a Matching to Sample Task Under Different Language Modes |
| Early Social Skills as a Function of Gender: An Observational Study of Children Between 3-8 Years in Andhra Pradesh, India |
| Measuring Potential Cues for Depression in Adolescents |
| Observing the Paraverbal Communication of Coaches in Competitive Match Situations |
| Temporal Dynamics of Multimodal Multiparty Interactions: A Microgenesis of Early Social Interaction |
| Measuring Behavioral Indices of Cognitive Processing in Children |
| Full paper session 5: Studving Human Machine Interaction |
| A Method for Remote and Semi-Automatic Usability Evaluation of Web-based Applications Through Users Behavior Analysis |
| Beyond Usability: A Methodology to Evaluate the Affective Experience of Interaction with E-Commerce Websites |
| A Framework to Measure User Experience of Interactive Online Products |
| iHeartrate: A Heart Rate Controlled In-Flight Music Recommendation System |
| Measuring Motor Actions and Psychophysiology for Task Difficulty Estimation in Human-Robot Interaction |
| Evaluating Multimodal Human-Robot Interaction: A Case Study of an Early Humanoid Prototype |
| Measurement of Feet-Ground Interaction for Real-Time Person Localization, Step Characterization, and Gaming |
| Web Based Measuring System for Health Monitoring at Home |
| Towards the Use of Psychological Variables in User Profiling |
| MAESTRO: Orchestrating User Behavior Driven and Context Triggered Experience Sampling |
| Full namer session 6. Tracking and the Analysis of Motion |
| Multi-Method Analysis of the Relationship Between Individual's Space-Time Behavior, Built Environment and the Evolution of Cognitive Representations: Application of Tracking Technologies and Internet Surveys |

| Using Motion Sensing to Study Human Computer Interaction in Hospital Settings Katherine M. Sellen, Mark Chignell, Jeannie L. Callum, Jacob Pendergrast, Sharon E. Straus, Allison Halliday | |
|--|-----------------|
| A High-Resolution System for Recording the Daily and Lifetime Behavioral and Movement Patterns of Individual Tephritid Frui James R. Carey, Sige Zou, Pablo Liedo, Leopoldo Robles, Amy Morice, Hans G. Müller, Jane-Ling Wang, Donald Ingram, Nikos Papadopoulos | t Flies.301 |
| Identification of Vertical and Horizontal Movement Patterns in Cod Behavior Gudberg K. Jonsson, Vilhjalmur Thorsteinsson, Gunnar Gudni Tomasson | |
| Measuring 3D Arm Movements for Activities of Daily Living Thomas Haslwanter, Jürgen Waldhör | |
| Full paper session 7: Human Behavior in the Urban Environment; Measurements in the Field and Realistic Sin | tuations |
| Review of Real-time Visual Driver Distraction Detection Algorithms | |
| Using Semi-Automated Shadowing for Analysing Stress-Induced Spatio-Temporal Behaviour Patterns of Passengers in Public Temporal Behaviour Patterns of Passengers in Public Behaviour Patterns of Passengers in Public Temporal Behaviour Patterns of Passengers in Public Temporal Behaviour Patterns of Passengers in Public Behaviour Patterns of Passengers in Public Behaviour Patterns of Passengers in Public Behaviour Passengers in Public Behaviour Passengers in Passengers in Public Behaviour Passengers in Passengers in Public Behaviour Passengers in Passenger | ransport 314 |
| Alexandra Millonig, Gudrun Maierbrugger | |
| Assessment of Aging Effects on Drivers' Perceptual and Behavioral Responses Using Subjective Ratings and Pressure Measures Gyouhyung Kyung, Maury A. Nussbaum | |
| Applying Nonlinear Dynamics Features for Speech-based Fatigue Detection | |
| Detecting Fatigue from Steering Behaviour Applying Continuous Wavelet Transform Jarek Krajewski, Martin Golz, Sebastian Schnieder, Thomas Schnupp, Christian Heinze, David Sommer | |
| Creating a Richer Data Source for 3D Pedestrian Flow Simulations in Public Transport Veronika Egger, Helmut Schrom-Feiertag, Lisa Ehrenstrasser, Gregory Telepak | |
| Identifying Operators Monitoring Approriately Dietrich Grasshoff, Catrin Hasse, Hinnerk Eißfeldt, Carmen Bruder | |
| A Detailed Analysis of Eating Behaviour Ioannis Ioakeimidis, Per Södersten, Modjtaba Zandian | |
| Field Measurement Protocol for Team Communication: A Study of Medical Rehabilitation Team Interaction Balázs Péter Hámornik, Márta Juhász, Ildikó Vén | |
| Measuring Professional Teams' Information Sharing Behaviour Juliánna Katalin Soós, Márta Juhász, Balázs Péter Hámornik | |
| Child-Activity Recognition from Multi-Sensor Data Sabri Boughorbel, Jeroen Breebaart, Fons Bruekers, Ingrid Flinsenberg, Warner ten Kate | |
| Studying Media Usage in the Living Room by Measuring Infrared Signals Thomas Mirlacher, Thomas Fischer, Regina Bernhaupt | |
| Optimization of User Satisfaction on an Augmented Reality Based Guided Tour Marília Felismino, Ricardo Luiz, Marta Castela, Pedro Marques | |
| Posters | |
| Poster group A | |
| A New Method of Computer-Supported Measurements of Distances Moved by Animals Julia Brendle, Hoy Steffen | |
| | |

| EEG Data Logging with the NeuroLogger [®] System: Spontaneously Occurring Electrophysiological Correlates of Migraine in Mutant Mice368 Alessandra Di Lio, Dimitrula Arabadzisz, Hans-Peter Lipp, Hanns Ulrich Zeilhofe |
|---|
| Ethosearch: A Comprehensive Repository of Ethograms for Use in Animal Behavior Research |
| The Complexity of Reliably Investigating Effects of Novelty Stress in Rats: Dynamics and Correlations of Behavioral, Cardiovascular, Endocrinological and Electroencephalographic Responses Under Simultaneous Measurement |
| Development and Evaluation of an Operant-Based Reversal Learning Task in the Rat, Relevance of the Visual Cue |
| Measurement of Akinesia in Rats: Design and Validation of a Side Effect Paradigm in Freely Moving Animals |
| Visualizing Group Dynamics Signals |
| Automated Video Registration in Home Cage for Measuring Maternal Behavior of Wild, Aggressive and Tame Rats |
| Measuring Behavior in Complex Maritime Operations |
| Diagnostics of Motivational Conflicts Through Thinking Activity |
| Use of Ethological Rodent Behavior to Assess Efficacy of Potential Drugs for Alzheimer's Disease |
| Who Said Pedestrian Means Dull: Teaching Methods in the Measurement of Behaviour Through Observational Studies of Pedestrian Behaviour |
| Development of Behavioral Measures of Osteoarthritis-induced Pain in Rabbits |
| Progress With MINDS, a Testmanager for Psychological Assessment, Research and Education. Applications in the Forensic Psychiatric Domain |
| A. N. Brand, A. K. L. von Borries, B. H. Bulten |
| Effects of CGP7930 on Spontaneous Behavior, Anxiety and Learning in Immature Rats |
| Stereoscopic Motion Analysis of Giant Honeybees |
| Use of Automated Tracking System Across Anxiety and Depression Models in Rodents |
| Induction and Measurement of Basic Emotions in Subjects and the Influence of Multisensory Stimuli |
| Microstructural Assessment of Rodent Behavior in the Hole-Board Experimental Assay (also presented as paper) |
| Using Multivariate Analysis for Comparing Locomotor Patterns of Sprague-Dawley Rats from Different European Breeders |
| Development and Evaluation of a Rat Water Maze System: Impact of Maze Size, Water Temperature and Other Parameters on Performance |

| Video Tracking Analysis of Contextual Fear Conditioning for the Screening of Anti-Alzheimer Drugs after Acute Administration in Tg2576 Mice |
|--|
| Measurement of Locomotor Activity in Rodents: Design of a Compact, Multifunctional, User-friendly and High Throughput System414 Hansfried V. M. M. Van Craenendonck, Darrel J. Pemberton, Nancy Aerts, Koen Hens, Anton A. H. P. Megens |
| The Links Between Behavioral Markers of Performance, Personality and Communication |
| Food Attractiveness and Gazing Behaviour |
| Delay Intolerance Task: Different Impact of Timeout Duration on Impulsive and Non-Impulsive Rats |
| Poster group B |
| Quality Monitoring and Analysis of Open Field Behavior in Rats by Means of Multivariate Analysis |
| Different Interpretation of the Hot Plate Test in Rats |
| Preliminary Study of Behaviour and Sound Emitted by European Spiny Lobster Palinurus Elephas During the Interaction with Predators |
| Combining an Operant Chamber Paradigm with [18F]Fluorodeoxyglucose MicroPET Imaging: A Study on Conflict Processing in the Rat |
| |
| The FischFITMonitor – A New System for Monitoring Multiple Physiological and Behavioural Parameters in Fish |
| Spatial ECG of the Rat During Spatial Navigation |
| Characterization Of Foraging Behavior Of Yellowjacket Workers And Worker-Larva Trophallaxis Towards Feeding Baits Laced With Fipronil |
| Software Employment for Automatic Emotion Recognition in Online Psychotherapy |
| Use of Environmental Light Cycles to Distinguish Circadian Regulation from Direct Light Modulation of Learning and Mood |
| Postnatal Exposure to Clonazepam Leads to Cognitive Deficits in Adult Rats |
| Studying the Modulation of Brain Rhythms by Dynamic Cues |
| Combinatorial Measurement of Sleep and Wheel Running Activity to Examine the Interaction Between Light and Drug Administration in Mice. |
| Tara LeGates, Samer Hattar |
| Measuring Stress and Cognitive Load Effects on the Perceived Quality of a Multimodal Dialogue System |
| A Systematic Review on Randomization and Permutation Tests in the Educational and Behavioral Sciences |

| Pattern Recognition Methods – A Novel Analysis for the Pupillographic Sleepiness Test Jarek Krajewski, Thomas Schnupp, Sebastian Schnieder, David Sommer, Christian Heinze, Martin Golz | 459 |
|---|-----|
| Group Preparedness For Risk in the Environment of Social Instability | 463 |
| Trackmate (V. 1.0): A Versatile Program for Automated Tracking of Behaviour in Laboratory Rodents Esther Remmelink, Michael T. Bowen, Iain S. McGregor, Paul Halasz, Robert A. Dielenberg | 464 |
| Trackmate Social (V 0.9): Beta Software for the Automated Measurement of Rodent Social Behaviour Michael T. Bowen, Esther Remmelink, Robert A. Dielenberg, Paul Halasz, Iain S. McGregor | 465 |
| The Use of Wildlife Cameras to Monitor Pig Behavior During Commercial Long Distance Transport Katja van Driel, Richard Heath, Janet Talling, Heleen van de Weerd, Peter Kettlewell, Malcolm Mitchell | 466 |
| An Automated Tracking Algorithm for Quantitative Group Behavior Studies | 468 |
| The Compensatory Tracking Task: A Pattern Recognition Based Approach for Classifying Vigilance Corinne Schenka, Thomas Schnupp, Christian Heinze, Jarek Krajewski, Martin Golz | 470 |
| Automated Home Cage Assessment of Sca17 Mouse Model | |
| The BacHD Rat: A New Rat Model of Huntington's Disease Expressing the Full-Length Mutant Huntingtin Laura Emily Clemens, Libo Yu, Jasmin Simone Ehrismann, Esteban Portal, Olaf Riess, Huu Phuc Nguyen | 475 |
| Crowd Safety Architecture - Measuring Safety Levels Maarten Douwe Bredero | 477 |
| Free Exploration of an Elevated Plus-Maze in Mice A. Arabo, C. Potier, V. Roy | |
| Work Units in Tutorials Software for Learning Mathematics | |
| Workshops | |
| Measuring (Ambient) Persuasive Technologies Wolfgang Reitberger, Alexander Meschtscherjakov, Manfred Tscheligi, Boris de Ruyter, Jaap Ham | 489 |
| Prerequisites for Successful Measuring of Ambient Persuasive Technology | 491 |
| Wildlife Surveillance Using GPS: From Movement Tracking to Behavior Recognition Herbert Prins, Andrea Kölzsch, Arnold Bregt, Tamme van der Wal, Lucas Noldus | 495 |
| Measuring Consumer Behavior René de Wijk, Jos Mojet | 497 |
| Teaching a Course in Measuring Behavior Richard E. Brown, Timothy P. O'Leary | 499 |
| Acknowledgements | |
| Author index | |

Welcome to Measuring Behavior 2010

It is my great pleasure to welcome you to Measuring Behavior 2010, the 7th International Conference on Methods and Techniques in Behavioral Research. This conference edition is hosted at the High Tech Campus in Eindhoven. With over 90 companies and institutes, the HTC brings together a dynamic mix of multinational companies, small and medium-sized businesses and technology start-up companies. Campus residents share knowledge, experience, open laboratories and technical infrastructure, enabling better, faster and more cost efficient innovation. An open environment that fuels opportunities for valuable R&D, for successful business partnerships.

In this year's conference we bring closer together a diversity of communities ranging from neuroscience and zoology to psychology and consumer behavior. The conference location could not have been more symbolic for meeting with a multidisciplinary community interested in sharing methods and techniques for conducting behavioral research.

This year's Measuring Behavior conference features a very strong technical program, assembled under the expert leadership of Program Co-Chairs Emilia Barakova and Andrew Spink. Together with the Scientific Program Committee and expert reviewers from the community, they undertook the difficult job of carefully evaluating the large number of submitted papers, considering the merits of each through detailed reviews and selecting a technical program of the highest caliber. With a technical program bringing symposia, paper sessions, demonstrations, tutorials, user meetings, workshops, scientific tours and exhibitions there offer the best setting for productive cross-fertilization between research fields in the area of measuring behavior.

Again, welcome to Measuring Behavior 2010 at the High Tech Campus. I wish you a very productive and informative conference and hope that you will take the opportunity to strengthen your network with the Measuring Behavior community.

Boris de Ruyter Measuring Behavior 2010 Conference Chair

The Measuring Behavior Conferences

Measuring Behavior is a unique conference about methods and techniques in behavioral research. 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 find yourself among ethologists, behavioral ecologists, neuroscientists, experimental psychologists, human factors researchers, movement scientists, robotics engineers, software designers, human-computer interaction specialists... to mention just a few. 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. Crossing the boundaries between disciplines and species (from insects to astronauts) can be extremely inspiring. For many delegates, attending a *Measuring Behavior* meeting is an eye-opening experience, to find out which interesting (and often highly relevant) developments are taking place in domains they usually don't venture into.

Measuring Behavior started in 1996 as a workshop in the framework of a European research project "Automatic Recording and Analysis of Behavior", aimed at sharing the results of our project with colleagues from abroad. Organized by Noldus Information Technology and hosted by Utrecht University, *Measuring Behavior '96* attracted over 150 participants from 25 countries. Encouraged by the international interest, it was decided to make *Measuring Behavior* a recurring conference. In the years that followed, the conference travelled to six other Dutch university towns: Groningen (1998), Nijmegen (2000), Amsterdam (2002), Wageningen (2005), Maastricht (2008) and now Eindhoven (2010).

Over the years, *Measuring Behavior* has developed a formula with a mix of ingredients that has proven quite successful. The meeting is always held in a university town where research on human or animal behavior is prominent, with local scientists playing a prominent role in the conference organization (see table below). Noldus Information Technology serves as conference organizer and main sponsor. For a small company like ours, the conference is a major investment. The registration fees just cover the direct expenses associated with the meeting; the hours spent on the organization (several person-years) are on our account. We gladly do this, because we believe that the focused attention on behavior research methods and techniques will eventually lead to a higher demand for our tools. To prevent commercial bias, however, the scientific program is put together under auspices of an independent Scientific Program Committee, consisting of international experts from a broad variety of disciplines (see the Scientific Program Committee on page 501). We are very grateful for their effort to review papers and the helpful input during email exchanges as well as all the other reviewers of the papers in the scientific program.

| City | Conference chair |
|------------|--|
| Utrecht | Berry Spruijt |
| Groningen | Jaap Koolhaas |
| Nijmegen | Alexander Cools |
| Amsterdam | Gerrit van der Veer |
| Wageningen | Louise Vet |
| Maastricht | Harry Steinbusch |
| Eindhoven | Boris de Ruyter |
| | City Utrecht Groningen Nijmegen Amsterdam Wageningen Maastricht Eindhoven |

Over the years, the conference has grown significantly in size, from 153 delegates in 1996 to more than 400 in 2008. At this size, the event is large enough to cover a wide range of topics, yet still small enough for a social program with all delegates. *Measuring Behavior* has also become a truly global meeting: delegates come from dozens of countries on all continents. 2010 is the most international version so far with participants from more than 35 countries and for the papers submitted individually the host country is not longer the one with the most papers (that honor goes to the USA).

In the scientific program, well balanced between human and animal research, one finds a variety of formats for presentation, interaction and exchange of information. The traditional oral papers (full papers) and poster presentations have always been central to the conference. Increasingly, special symposia– focusing on a current methodological or technical theme – are proposed by experts from various disciplines. These symposia illustrate the widening scope of the conference as well as trends in science. From the start, there have been symposia on topics in behavioral neuroscience (such as animal models for human disease or automatic behavior recognition in rats and mice) and data analysis and statistics (such as sequential analysis and pattern detection). Subsequently, the scope extended towards psychology, human factors, ergonomics and movement science. And at this year's conference we see the arrival of novel topics such as human-robot interaction, measuring behavior in the operating theatre and behavior of forensic scientists.

Besides oral presentations in symposia or free paper sessions, the conference program always includes ample time for posters and demonstrations of software or equipment by participants. The latter are actively encouraged, because it is a format not supported by most other conferences. Full demonstrations in a seminar room were new at the previous Measuring Behavior and we are pleased to see that has really taken off, with 12 separate demonstrations on the Thursday afternoon. For several academic inventors, the presentation of their prototype software or hardware tool at *Measuring Behavior* paved the way towards commercialization. This is how CatWalk and FaceReader found their way to the Noldus product portfolio. We hope that scientists will continue to present inventions at *Measuring Behavior* and discuss commercialization opportunities with the vendors present at the meeting.

Another attractive element of the conference is the scientific tours, guided visits to behavioral research facilities and laboratories in and around the hosting university. Tutorials, short courses – mostly about software tools and instruments – taught by expert instructors, have also become a popular program element. Other program elements are user meetings (organized by manufacturers of research tools), and workshops. At this year's conference, the latter have become more prominent: seven workshops are being held, about topics ranging from autism research to GPS tracking and behavior recognition in wildlife. Finally, there is the commercial exhibition of scientific instruments and software related to behavioral research.

Measuring Behavior is a scientific conference, so special attention is paid to publication of the work presented at the meeting. We started off with a program book and an abstracts book. In 2005 we added printed conference proceedings with short papers and a conference CD. Because of the overlap between the abstracts and proceedings books, we have gone back to two books: a program book and printed proceedings of extended abstracts (short papers). Then there is the conference website (www.measuringbehavior.org). 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. Presenting authors will have noticed that they were asked to submit using a rather complex template. This was to enable us to archive all the reviewed papers to the Digital Library of the ACM (the Association for Computing Machinery), which is an important publication channel for those studying human-computer

interaction. Finally, selected presentations will be published as full papers in the Journal of Integrative Neuroscience.

Now you find yourself at the 7th *Measuring Behavior* conference. The organizers have done their best to prepare an optimal mix of scientific, technical, social and culinary ingredients. We hope that you will find *Measuring Behavior 2010* a rewarding experience and wish you a pleasant stay in Eindhoven.

Lucas P.J.J. Noldus Managing Director, Noldus Information Technology BV l.noldus@noldus.nl

Measuring Behavior 2010: Highlights of the scientific program

The purpose of the Measuring Behavior conferences has been to provide a broad and interdisciplinary forum for novel methods to define, measure, and analyze human and animal behavior. The conference looks at individual organisms and their behavior along with the behavior of groups, and societies, and processes which can explain the specific behaviors. Measuring behavior conferences reflect the trends in the developments of social and behavioral sciences, as well as the changes in society and the environment. In this way the conference has become an indicator for the important issues in science and in real life.

The seventh MB conference features themes such as measuring behavior in different professional domains and specific social environments, such as measuring surgical behavior, the behavior of crime scene investigators, the potential conflicts in isolated groups during long-term missions, autistic behavior, sports performance, consumer behavior, etc. It is a trend that continues to develop that behavior is more and more studied a rich natural context in addition to the purely laboratory environment.

Other group of themes revolve around new methods and techniques for measuring specific aspects of animal and human behavior, such as new behavioral models in rats, mice and zebrafish; ambulatory recording of psycho-physiological signals; innovation in movement behavior analysis and tracking techniques; applications of GPS technology such as wildlife tracking.

The tutorials feature methods which are already established for measuring behavior. These include; analyzing behavior and interactions with THEME, analysis of behavior using operant conditioning methods; Pavlovian conditioned freezing; two tutorials on BIOPAC physiological data acquisition systems, ExpertEyes (an opensource eye-tracking system) and as well as tutorials on commercially avialblae eye trackers, behavioral research and digital video, and MotoRater (a new, system for quantification of locomotor impairment in rodents with CNS damage).

Measuring movements, gaze, facial expressions, for the purpose of understanding and prediction of emotions, and social conflicts are popular themes of the conference, present in the symposia, workshops, and the main program. The symposia that feature those are; The significance of voluntary exploration to monitor emotional behavior of rodents, Unveiling affective signals, Large and small scale physiological recordings in behavioral context, Monitoring of social interactions and initial signs of conflicts in isolated groups during long-term missions. Topical topics include measurement of intelligent technologies such as Measuring (ambient) persuasive technologies. We see that increasingly behavioral scientists do not just rely one one way of collecting their data, but that multiple modalities are combined to give more insight into the behaviors of the subjects.

Studies of measuring human and animal behavior have resulted in a wide range of results, some of which have inspired interesting and societal relevant applications. The novel applications in their turn pose the need for new measuring approaches and technological solutions. We hope that the discussions during the upcoming conference will help us cast a new critical look at existing methodologies and identify novel applications for existing measuring methods. The creative atmosphere that the conference creates will inspire new methods for measuring behavior that will be based on the discussions on the shortcomings in the existing methods and techniques.

MB2010 will take place in the city of Eindhoven, a Dutch center for Hi-tech industries and innovative technologies. Established industries such as, engineering, the automotive industry and electronics coincide with new sectors as industrial distribution, environmental technology, medical technology and information technology. The integration of research and development activities are the central engine

powering the development of the region. 50% of the total amount spent annually on research and development in the Netherlands is invested in the Eindhoven Region, which rightfully holds the slogan "Leading in Technology".

Emilia I. Barakova Andrew Spink

Chairs, Scientific Program Committee Measuring Behavior 2010

Keynote Lecture

Professor Naotaka Fujii

Laboratory for Adaptive Intelligence, Brain Research Institute RIKEN, Japan

About the Speaker

Naotaka Fujii was granted MD in 1991 and PhD in 1997 at Tohoku University School of Medicine. He started his career as Ophthalmologist after graduating medical school and later switched to Neuroscientist. His first post-doctoral position was appointed in Graybiel lab at MIT, USA. At MIT, he conducted a research revealing neural mechanism of sequential oculomotor behavior in monkeys. Then, he moved back to Japan in 2004 and joined Dr. Iriki's lab at RIKEN (Laboratory for Symbolic Cognitive Development) as deputy lab head and started studies of social brain function and development of interactive brain machine interface. He is a member of Society for Neuroscience, Japan society of neuroscience and Japan society of Physiology.

Naotaka is interested in Social Brain Function and Brain Machine Interface. He is trying to combine these two research topics to understand how we can make adaptive behavior. Especially development of interactive brain machine interface is the most important research theme for him. He is also interested in development of interactive virtual reality (VR) environment in which subject can make interactive communication as same as reality. In future, he is dreaming revealing human brain function by using interactive BMI and VR technologies.

Naotaka wrote several books. One of the books titled "SOCIAL BRAINS" was awarded The 63rd MAINICHI Publication and Culture Award in Natural Science Category.

Multi-Dimensional Recording in Social Primates: Method and Application

Social brain function is a neural mechanism of communication that enables us to make social context dependent behaviors. The function has to deal with complex parameters in environment and behavior of surrounding others to select socially correct behavior at the moment. Thus learning the function also requires recording and analyzing multidimensional data which conventional physiological methods couldn't handle.

To break such technical limitations in learning social brain function, we have developed multi-dimensional recording technique (MDR). MDR consists of two state of the art technologies, multi-electrode recording technique and motion-capture system. We can record neural activity from wide brain regions and precise motion data simultaneously as well as other biological parameters, like eye position, heart rate, blood pressure, skin conductance, muscle activity and etc, while the subject and other agents are freely behaving. I will introduce detailed information of MDR and explain how we could apply the technique in Japanese macaque during social task.

Keynote Lecture

Professor Kerstin Dautenhahn

Adaptive Systems Research Group, The University of Hertfordshire, School of Computer Science, College Lane, Hatfield, Hertfordshire AL10 9AB United Kingdom

http://homepages.feis.herts.ac.uk/~comqkd/

About the Speaker

Kerstin Dautenhahn pioneered research in robot social learning and imitation, and the study of robots in autism therapy. She has published more than 200 research articles and has been involved in several European research projects as Principal Investigator of her research team. Kerstin has authored and edited several books and frequently organizes international research workshops and conferences. She is Editor in Chief of the journal Interaction Studies, and Associate Editor of the journals Adaptive Behavior, the International Journal of Social Robotics, and IEEE Transactions on Autonomous Mental Development. She is a Fellow of the Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA).



Measuring Behaviour in Human-Robot Interaction Studies

For more than 10 years I have been involved in human-robot interaction (HRI) studies in a variety of application contexts ranging from assistive technology for children with autism to robots as assistive home companions. In addition to questionnaires, interviews, focus groups and other measures, behavioural analysis of human-robot interaction plays an important part in this research. Such analysis illuminates how people behave in the presence of and in interaction with robots, which may or may not be consistent with participants' subjective evaluation (e.g. the user experience).

The ultimate goal of such evaluations is to improve a robot's performance and social skills. Methodologically, we also investigated and developed new approaches towards user studies in HRI including the use of theatre. I will provide examples of such research and point out particular challenges for behavioural analysis and behaviour design for social robots in these domains.

Keynote Lecture

Scott Makeig, Ph.D.

Director, Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego

About the Speaker

Scott Makeig was born in Boston, MA, USA in 1946 and completed an honors Bachelors degree, 'Self in Experience,' at the University of California Berkeley in 1972. He received a Ph.D., 'Music Psychobiology,' from the University of California San Diego (UCSD) in 1985. After spending a year in Ahmednagar, India as an American India Foundation research fellow, he became a psychobiologist at UCSD, and then a research psychologist at the Naval Health Research Center San Diego. In 1999, he became a staff scientist at the Salk Institute, La Jolla, then moved to UCSD as a research scientist in 2002 to develop the Swartz Center for Computational Neuroscience, which he now directs. Recently, university teams he organized won large US Navy and Army research project grants to develop basic research principles for cognitive monitoring from mobile brain/body imaging (MoBI) data, an imaging modality he recently defined and is now working to develop.



Measuring What the Brain Does, What It Experiences, and What It Controls: Mobile Brain/Body Imaging

Cortical brain areas and dynamics evolved to organize motor behavior in our three-dimensional environment also support more general human cognitive processes. Yet traditional brain imaging paradigms typically allow and record only minimal participant behavior, then reduce the recorded data to single map features of averaged responses and compare the severely reduced behavioral and brain data measures. To more fully investigate the complex links between distributed brain dynamics and motivated natural behavior, we have proposed and are now pursuing development of wearable mobile brain/body imaging (MoBI) systems to continuously capture the wearer's high-density electrical brain and muscle signals, threedimensional body movements, audiovisual scene and point of regard, plus new data-driven analysis methods to model their interrelationships to each other and to cognitive context. The new imaging modality can give new insights into how spatially distributed brain dynamics support natural human cognition and agency.

Symposia

Symposium: Improving Sports Performance

Patricia de Cocq Experimental Zoology Group, Wageningen University, The Netherlands Patricia.deCocq@wur.nl

ABSTRACT

In top class sports, performances of athletes are very close. Small improvements in technique can make the difference between winners and losers. A combination of behavioral analysis and physiological data can play a key role in improving sports performance and developing innovative training tools. It is a challenge to incorporate modern measuring equipment into the daily training of athletes. The measurements should not disturb the athlete and coach and in the same time should record all data necessary. Measurement equipment should therefore be small, wireless and user friendly. Several sports face this challenge.

Author Keywords

Sports performance, behavioral analysis, physiological data, innovations

INTRODUCTION

In several sports, projects are carried out to improve performance by measuring behavioral and physiological parameters. Performance monitoring is becoming more important to achieve and maintain a top position in several sports. Furthermore, new insight in training techniques enables the development of innovative tools which can improve the sports performance, increase the maintenance of top positions and help to prevent injuries.

An example of an innovative tool that changed sports performance tremendously is the klapskate. The klapskate increases the skating speed with about 5% by increasing the external power of the skater. It took 10 years before the klapskate was accepted by the top skaters. This demonstrates the challenge researchers face when bringing their research to the athletes and the importance of a joined approach by researchers, athletes and coaches. Techniques that might seem usable in an experimental set-up, might not be usable in practice. Furthermore, a joint approach will fill the needs of the athletes better and barriers for potential users will be prevented or lowered.

In sports where speed is the measure for a good performance, a first approach in improving the performance lays in reducing air and water resistance. This could be a change in clothing, but also a change in the technique of the athlete. This can be a body position that minimizes friction, a more efficient turn or a faster, more efficient takeoff (swimming).

Performance in team sport can often been judged in the amount of goals that are scored. This performance is influenced by the tactic, the team play and the performance of individual athletes. These separate performances can also be monitored and behavioral and physiological data can lead to an improved performance. In the AC Milan lab these data streams are integrated using specific algorithms. In these algorithms the different parameters for sports performance are integrated.

Sports that are judged by juries, such as gymnastics and equine sports (dressage) face an extra challenge. It is hard to objectify the performance of these athletes. Performance is these sports is often specified as an specific combination of movements. Behavioral observation techniques can play an important role in the specification of the performance, especially in combination with sport analysis software. In equine sports, one of the athletes is a horse. When an animal is used in sports, special care should betaken to guard the welfare of the animal. The use of behavioral observations in combination with physiological parameters also play a role here.

This symposium demonstrates how these different sport disciplines take on the challenge to incorporate scientific tools into the training.

SYMPOSIUM CONTENTS

Performance Monitoring in Equine Sports

Patricia de Cocq (Wageningen University, The Netherlands) & Carolien Munsters (Utrecht University, The Netherlands)

Methodological Approach to Evaluate Interactive Behavior's in Team Games: an Example in Handball

João Prudente (University of Madeira), Júlio Garganta (University of Porto, Portugal) & M.Teresa Anguera (University of Barcelona, Spain).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

The Measurement of the Visual Search Behavior in Sport. Can it Be a New Avenue into Talent Identification and Development?

Geert Savelsbergh (Institute Move VU University, The Netherlands).

Faster Marathon Times by Measuring Human Performance.

Cees van Bladel (Sports and Technology, The Netherlands).

Performance Monitoring in Equine Sports

Patricia de Cocq Experimental Zoology Group, Wageningen University, The Netherlands Patricia.deCocq@wur.nl

INTRODUCTION

Training in equine sports is largely based on experience and intuition. To optimize the training of equine athletes, there is a need for objective measurements of biomechanical, sport physiological and behavioral parameters. These parameters can be used by trainers to improve their training strategy.

Thereby, performance of equine athletes is largely determined by the musculoskeletal apparatus of the horse. Thus far, studies on training have mainly focused on exercise physiology [7] or kinematics [1, 6, 16]. The incorporation of forces between rider and horse and behavioral observations would provide essential information when comparing riding techniques, rider levels and training programmes. This could lead to better training and increase the understanding of the rider-horse interaction.

Another important reason to develop objective measurement techniques for these parameters, related to the use of the horse, is the high incidence of musculoskeletal disorders, which rank first among the causes of wastage in performance horses. These musculoskeletal injuries are almost invariably caused by single or repetitive biomechanical overload.

Equestrian sports are mainly judged by juries and therefore face an extra challenge. It is hard to objectify the performance and welfare of these athletes. Performance is these sports is often specified as an specific combination of movements. Behavioral observation techniques can play an important role in the specification of the performance, especially in combination with sport analysis software.

We aim to enhance the understanding of the biomechanical interaction and physiological information of rider and horse in order to optimize training and performance. **Carolien Munsters**

Department of Equine Sciences, Utrecht University, The Netherlands C.Munsters@helicon.nl

Training of Horses

Training is essential for horses to compete effectively and safely. All equine athletes undergo some type of training regime to prepare them for the level of their competition. Training prepares the equine athlete for competition by inducing the physiologic adaptations necessary to perform at a high level with minimal risk of injury, and by providing the appropriate behavioral and psychological factors essential for effective training [11]. Field testing during training and competition of warmblood sport horses can give more optimal training programs to prepare them for their level of competition.

Aids and Forces During Riding

During the training of horses several signals (aids) are used. It is very important to understand the effect of the different aids. The following groups of aids are of interest: rein force, saddle force and force on the bridle. Next to these aids the effects of rider weight, rider leg pressure and rider balance should be evaluated. These aids are important in most equine disciplines (dressage, jumping, etc.)

The equestrian world is especially interested in the effects of training techniques that are considered less conventional, such as hyperflexion. The questions is if and how these techniques can be applied without jeopardizing the welfare of the equine athlete. Recent studies have provided information on hyperflexion, but it is still questioned whether this technique is acceptable [10, 18].

Welfare

Welfare is a very important issue in equine sports. The evaluation of welfare is mainly subjective. It would be a big step forward if more objective criteria for welfare could be developed. There is good reason to believe that force measurements can be of great help, as it is likely that there is a relationship between the forces exerted on the horse and the degree to which the welfare of the animal is affected.

Injuries in Horses

The focus lies on overloading of the musculoskeletal system of the horse and its prevention. Overloading can be related to incorrect riding technique, training, shoeing, surface characteristics, excessive rider's weight, or riding a too young horse.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Lameness is the most common cause of involuntary early retirement in performance horses [15]. Prevention and cure of lameness are therefore economically important for the horse industry. Loss of horses due to lameness is so frequent that insurance is expensive. There is an important welfare aspect too; lameness is becoming socially less acceptable because it affects the well being of animals.

MATERIALS AND METHODS

Analysis of horse and rider in locomotion is done by means of development of measurement equipment and experimental work. The experimental work focuses on the comparison of different riding techniques, levels of riding experience and training programmes for horses and riders. During the experimental work motion analysis, force measurements at contact points between rider and horse (saddle, girth, stirrups and reins), sport physiological measurements (heart rate and heart rate variability of horse and rider and lactate) and behavioral observations will be combined. The results of these experiments will be translated to the equine practise. This will lead to a correct balance between loading and the physical and mental fitness of both horse and rider.

Identify Variables

The measurement results will be compared for the different riding techniques, riding levels and training programs. Besides this the evaluation of different variables (magnitudes of forces, joint angles, force patterns) in order to identify variables that provide the best information on the difference in training and horse loading. This will create the possibility to develop a panel consisting of a restricted number of selected, well-validated, highly relevant variables for the use of measurement techniques in a practical setting (by trainers).

3D Kinematics

Motion of the horse and rider are recorded by high speed infrared cameras. The kinematics of the head and neck, back, limbs and centre of mass of both the horse and rider are evaluated [2, 3, 4, 5, 8, 9, 13, 14, 17].

For measurements of kinematics under field conditions, the use of high speed cameras is not practical. Inertial motion sensors might provide a user friendly alternative. For the application of inertial sensors in equine sports a combination of the data of the sensors with a biomechanical model of the horse has to be developed.

Force Measurements

The rider provides several sensory signals to the horse that lead after processing by the nervous system to a range of locomotor behaviors. In addition, the rider influences the musculoskeletal system of the horse also directly, for instance by his or her weight and by other forces exerted on the horse. The biomechanical effects of the rider on the horse can be examined by various measurements. We use high-speed video to quantify the 3D kinematics of the bodies of rider and horse, and will measure forces underneath the saddle, between the legs of the rider and the horse, in reins, stirrups leathers and in the girth using custom-built equipment.

Sport Physiological Measurements

Fitness of warmblood sport horses is measured by a standardized exercise test (sub maximal) in combination with lactate measurements. Using a heart rate monitor, the heart rate and heart rate variability (beat to beat intervals) of the horse can be measured. Thereby simultaneously the GPS system will measure distance and speed during the exercise test. To assess performance and fitness lactate concentration VLa4 (velocity at a lactate concentration of 4 mmol/L) can be used. As the fitness increases, VLa4 increases. Thereby V140, V170 and V200 (velocity at heart rate 140,170 and 200 bpm) and the recovery heart rate 5 and 10 minutes after training can be used to asses performance and fitness of the horse.

Behavioral Observations

The behavioral observations will be a combination of parameters characteristic for equine locomotion (e.g. gait, exercise) and stress-related behaviors in the ridden horse [12]. Stress-related behaviors are used because they give an indication of the welfare of the horse. These behaviors will be defined in an ethogram.

REFERENCES

- 1. Back, W., Hartman, W., Schamhardt, H.C., Bruin, G. and Barneveld, A. (1995a) Kinematic response to a 70 day training period in trotting Dutch warmbloods. *Equine Veterinary Journal, Supplement*, 18, 127–131. **1**
- Back, W., Schamhardt, H.C., Savelberg, H.H., van den Bogert, A.J., Bruin, G., Hartman, W. and Barneveld, A. (1995b) How the horse moves: 1. Significance of graphical representation of the equine forelimb kinematics. *Equine Veterinary Journal*, 27, 31-38.2
- Back, W., Schamhardt, H.C., Savelberg, H.H., van den Bogert, A.J., Bruin, G., Hartman, W. and Barneveld, A. (1995c) How the horse moves: 2. Significance of graphical representation of the equine hindlimb kinematics. *Equine Veterinary Journal*, 27, 39-45.3
- 4. Buchner, H.H.F., Obermüller, S. and Scheidl, M. (2000) Body centre of mass movement in the sound horse. *The Veterinary Journal*, 160, 225-234.
- 5. Buchner, H.H.F., Obermüller, S. and Scheidl, M. (2001) Body centre of mass movement in the lame horse. *Equine Veterinary Journal, Supplement*, 30, 122-127.
- 6. Clayton, H.M. (1993) The extended canter: a comparison of some kinematic variables in horses trained for dressage and for racing. *Acta Anatomica*, 146, 183–187.
- 7. Evans, D.L. (2007) Physiology of equine performance and associated tests of function. *Equine Veterinary Journal*, 39, 373-383.

- Faber, M.J., Schamhardt, H.C., van Weeren, P.R. and Barneveld, A. (2001) Methodology and validity of assessing kinematics of the thorocolumbar vertebral column in horses on the basis of skin-fixated markers. *American Journal of Veterinary Research*, 62, 301-306.
- Faber, M., Johnston, C., van Weeren, R. and Barneveld, A. (2002) Repeatability of back kinematics of horses during treadmill locomotion. *Equine Veterinary Journal*, 34, 235-241.
- 10. Gómez Álvarez, C.B., Rhodin, M., Bobbert, M.F., Meyer, H., Weishaupt, M.A., Johnston, C. and van Weeren, P.R. (2006) The effect of different head and neck positions on the thoracolumbar kinematics in the unridden horse. *Equine vet. J., Suppl.* 36, 445-451.
- 11. Hinchcliff K.W, Geor R.J, Kaneps A.J; (2008) Equine Exercise Physiology, Saunders Elsevier.
- Kaiser, L., Heleski, C.R., Siegford, J. and Smith, K.A. (2006) Stress-related behaviors among horses used in a therapeutic riding program. *Journal of American Veterinary Medicine Association*, 228, 39-45.
- Lida, H. and Yamamuro, T. (1987) Kinetic analysis of the centre of gravity of the human body in normal and pathological gaits. *Journal of Biomechanics*, 10, 987-995.

- 14. Lovett, T., Hodson-Tale, E. and Nankervis, K. (2005) A preliminary investigation of rider position during walk, trot and canter. *Equine and Comparative Exercise Physiology*, 2, 71-76.
- Perkins, N.R., Reid, S.W.J., Morris, R.S. (2004) Profiling the New Zealand Thoroughbred racing industry. 2. Conditions interfering with training and racing. New Zealand Veterinary Journal 53: 69-76.
- Rogers, C.W., Firth, E.C. and Anderson, B.H. (2005) Musculoskeletal responses of 2-year-old Thoroughbred horses to early training. 5. Kinematic effects. *New Zealand Veterinary Journal*, 53, 95-100.
- Schils, S.J., Greer, N.L., Stoner, L.J. and Kobluk, C.N. (1993) Kinematic analysis of the equestrian Walk, posting trot and sitting trot. *Human Movement Science*, 12, 693-712.
- 18. Sloet Van Oldruiterborgh-Oosterbaan, M.M, Blok, M.B., Begeman, L., Kamphuis, M.C.D., Lameris, M.C., Spierenburg, A.J. and Lashley, M.J.J.O. (2006) Workload and stress in horses: Comparison in horses ridden deep and round ('rollkur') with a draw rein and horses ridden in a natural frame with only light rein contact. *Tijdschrift voor Diergeneeskunde*, 131, 152-157.

Methodological Approach to Evaluate Interactive Behaviors in Team Games: An Example in Handball

João Prudente

prudente@uma.pt University of Madeira, Social Sciences Department of Physical Education and Sport, Funchal, Portugal

Júlio Garganta jgargant@fade.up.pt University of Porto- Faculty of Sport, Porto, Portugal

M.Teresa Anguera

tanguera@ub.edu University of Barcelona, Faculty of Psychology, Barcelona, Spain

ABSTRACT

In this paper we describe a method to evaluate interactive behaviors in Team Games, most precisely goalkeeper interaction with defenders. The aim of this study is to understand the performance of a handball goalkeeper, not only by the results of their own actions, but by integrating in the model the interactive actions between goalkeeper and defenders. The requirements of observational methodology are observed to gather data and we used sequential analysis and polar coordinates techniques to analyze data. The data are gathered from a natural environment where players have their behaviors, i.e., from handball matches in high level competitions context. Using an observational tool specially created and validated by an expert panel, the data was registered in a sequential way to make possible the analysis sequential with lags, prospective and retrospectively made.

The results show the goalkeeper performance is significantly associated with defenders actions.

Author Keywords

Observational methodology, sequential analysis, interaction, handball, goalkeeper.

INTRODUCTION

The evaluation of a goalkeeper performance in Handball is normally made considering only the balls received in relation with the balls saved by this player. This equation is very poor to translate goalkeeper's performance during a match. In fact, the goalkeeper makes actions according to the player who is throwing to the goal. But, this situation isn't always a simple and direct confrontation between goalkeeper and the attacker with the ball. Most times, goalkeeper makes actions and moves in a complex situation, where the attacker with ball and the defender(s) which made opposition are present and make necessary, to the goalkeeper, give attention to that. We can consider three kinds of different main situations when goalkeeper is defending the goal in Handball: 1) During a 7 m throw or in a 1x0 situation after a counter-attack; 2) when the throw is made with a defender between the attacker and the goal; 3) when the attacker is in the middle of defence or is trying to score a goal from the wing, with opposition.

Except for the first case, we always have a complex situation with interactions between goalkeeper, defender(s) and the attacker who is trying to score a goal.

The question is: How to evaluate the goalkeeper performance considering this reality? How can we analyze this performance without considering these interactions?

The Goalkeeper Cooperation with Defence: A Special Need and a Special Skill

The velocity of the ball and the distance, from where the throw takes place, make impossible to the goalkeeper to react and stop and save the ball. Because of the place from where the throw is made, near the goal, and the velocity of the ball, he needs to anticipate the attacker action and calculate the probability of the ball location on the goal. This behavior is a tactical one and, in order to make decisions correctly, he must consider all the signs he can perceive on advance from the ball, from the attacker and from defenders.

Considering the three different main situations related to goalkeeper when he is defending the goal: 1) During a 7 m throw or in a 1x0 situation after a counter-attack; 2) when the throw is made with at least a defender, between the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

attacker and the goal; 3) when the attacker is in the middle of defence or is trying to score a goal, with opposition, from the wing, we easily identify a direct confrontation in the situation 1 between goalkeeper and the shooter and a complex situation (case 2 and 3) where interactions between goalkeeper, defender and attacker are a reality [6].

In Handball more than 50% of shots made have a defender between the ball and the goal or the attacker who throws is in constraint by defenders actions or position, like we wrote before (points 2 and 3). That means the cooperation between goalkeeper and defenders is a necessary skill to both kind of handball players, defenders and goalkeeper, in order to maximize their actions and have efficacy in intent to protect their goal.

This cooperation is more effective when defenders "attack the attack", this means, when defenders are pro-active and create constraints to the attacker shot action, intervening on decision making concerning this offensive skill, decreasing the possibilities of choice concerning the placement of ball on the goal. Thus, defender gives to goalkeeper the possibility to anticipate the location of the ball on goal increasing his efficacy. However, the defender position is a positional constraint for the goalkeeper and also for the attacker with ball and as important as the defender actions against attacker and the distance between these two players. Having someone that it makes possible for the goalkeeper to anticipate actions, from an attacker with ball and understand what he is trying to do give him a possibility of wining the duel against the shooter. Considering all these reasons, doesn't make sense to evaluate the goalkeeper performance without analyzing goalkeeper/defender and defender/attacker interaction, when this interactive behavior is a significant part of goalkeeper behavior during a match.

Analyzing Goalkeeper/ Defence Interaction

To assess goalkeeper performance, considering the situations 2 and 3, we must analyze the defender's behavior and the influence he has on the goalkeeper, also considering the interaction between defender and shooter concerning actions and distance between players. We need to understand the influence of the defender positional and behavior constraint, related to the shooter and goalkeeper, in the defensive process.

Using observational methodology and considering multievents, we believe it's possible to find patterns that give the possibility to understand the goalkeeper performance with more accuracy. With this methodology we guarantee make possible a qualitative and quantitative analysis of data [1].

Types of Data and Variables

Multi-events (Multievent Sequential Data) [2], i.e., events that co-occurred, when registered sequentially give the possibility to gather data considering goalkeeper action related to the defender and attacker actions that cooccurred. The variables considered in this model are: 1) related to the goalkeeper position and movement on goal: initial position on the goal; moving to the left or right side anticipating the ball location; 2) position on the goal related to defender: behind defender; on the left side; on the right side; 3) goalkeeper action when he is defending the goal: move to the front; jump; stay on position; moves laterally on the line goal; get down; 4) related to defender during defence process: attack the man with ball giving him pressure; press the shooter and contact him; stay on defence position with arms down; stay on defence position with arms up; move laterally and intent to decrease free space to the attacker; stay, jump and try to block the ball; move in front and try to block the ball; 5) defender position on playfield in relation with attacker: near attacker; far from attacker; on the side of ball arm; on contrary side of ball arm; on middle position of the attacker body; 6) the attacker position on playfield when he tries to throw into goal: wing position on left or right side: in 6m zone: in central backward zone: in lateral backward zone on left or right side; 7) the result of the sequence: goal; goalkeeper save; ball throw out the field play; the defender block the ball; defender intercept the ball.

Analyzing Data with Sequential Analysis and Polar Coordinates

Sequential analysis with lags is one of the analysis techniques used on observational methodology that was developed by Bakeman & Gotman with support of Saket [3]. The goal of this technique is to detect patterns of behavior or regularities of behaviors when they are sequentially registered.

Using sequential analysis with lags we can detect and find regular association between behavior criteria and the events we have registered, on prospective and retrospective way. Using this technique analysis we can detect patterns of behavior that occurred during a competition, from goalkeeper in relation with defender and attacker. Looking prospectively from goalkeeper behaviors criteria to behaviors that happens on lag 1, 2 or 3 until the end of sequence, we know if the association between behaviors are significant, considering the level of significance we use (0,05). If the results on lag 1, 2 or 3 are significant, that means the probability of behavior criteria activate behaviors on lag 1, 2 or 3 is significant too.

Considering the events related with the result of defensive sequence, ball saves by goalkeeper for example, as criteria behavior and using retrospective sequential analysis, from this behavior backward, we can see behaviors that happens in lags -1,-2,-3 until the beginning of the sequence, looking for significant results.

A sequential analysis process is a particular kind of probabilistic process, which allows us to know the structure of behavior flow [4].

Other technique used on observational methodology is polar coordinate technique that has a double strategy: reduce data and give the possibility to represent, on vectorial way, the complexity of inter-relations established between different variables of an *ad hoc* system [5].

CONCLUSIONS

Using this model to evaluate goalkeeper performance we have more accuracy in the evaluation because we consider all the situations that can happen during a match and not only the situations of direct confrontation between goalkeeper and shooter. Considering interactive behaviors between goalkeeper, defender and shooter, we guarantee more similarity with goalkeeper behavior in her natural environment, in this case, on playfield during a match competition.

The results show how significant are positional and action constraint of a defender on shot action and on goalkeeper behavior trying to save the ball. Giving a different weight to the different situations according the match occurrences of events, considering direct confrontation and others different situations, we can obtain the final result of goalkeeper performance. Besides, if we are interested, we can isolated the goalkeeper performance in direct confrontation from de others that depends on cooperation and interactions.

REFERENCES

- 1. Anguera, M. T. (2001). *La observación*. Barcelona: Universidade de Barcelona.
- 2. Bakeman, R., & Quera, V. (1996). Análisis de la Interacción-Análisis Secuencial con SDIS y GSEQ. Madrid: RA-MA Textos Universitários.
- Bakeman, R., & Gottman, J. (1989). Observación de la interacción: introducción al análisis secuencial. Madrid: Ediciones Morata.
- 4. Hernández Mendo, A., y Macias, M. (2002). *Cómo usar la observación en la psicologia del deporte:principios metodológicos*. <u>http://www.efdeportes.com.</u>
- Hernández Mendo. (1999). Observación y Deporte. In M. T. Anguera (Ed.), Observación en Deporte y conducta cinésico-motriz: Aplicaciones. Barcelona: Edicions Universitat de Barcelona.
- Prudente, J., Garganta, J., & Anguera, T. (2004). Desenho e validação de um sistema de observação no Andebol. *Revista Portuguesa de Ciências do Desporto*, 4(3), 49-65.

The Use of Video for Improving Swim Start Performance

Sander Schreven

InnoSportLab De Tongelreep, Eindhoven, The Netherlands and Research Institute MOVE, FBW, VU University Amsterdam, The Netherlands info@escritosport.nl

Roald van der Vliet

InnoSportLab De Tongelreep, Eindhoven, The Netherlands r.vandervliet@me.com

PAPER WITHDRAWN

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

The Measurement of the Visual Search Behavior in Sport. Can It Be a New Avenue into Talent Identification and Development?

Geert J.P. Savelsbergh

Faculty of Human Movement Sciences, VU University, Amsterdam, The Netherlands GSavelsbergh@fbw.vu.nl

ABSTRACT

Competitive sports like cricket, football, and tennis require the players to catch, intercept or return a fast moving ball. The key to such a successful action is to meet the ball precisely at the right place at the right time. This type of successful coordinated performance requires skill in perception as well as the efficient and accurate execution of movement patterns. Therefore, the contribution of visual information is equally important as the motor skill. In other words, skill in perception in conjunction with an efficient execution of movement patterns is paramount. In the last decade, several researchers examined the skill of perception extensively. Common method in such experiments is to ask participants to predict the end result of video clips showing more or less predictive information from the opponent's body or the ball's flight path. Their findings show that experts have superior anticipatory skills compared to novices (e.g. Savelsbergh et al., 2002). In fact, research shows that an important difference between experts and novices appears to be the capability to pick up advance information from some visual sources (Abernethy & Russell, 1987; Savelsbergh et al., 2002, 2005). For instance,

with respect to football, players have developed an extensive football-specific knowledge base that enables them to recognize meaningful associations between the positions and movements of players in game situations (Savelsbergh et al., 2006; Williams et al., 1994). In other words, not so much the visual search strategy itself, but how the expert athlete can make use of this information is essential (Savelsbergh et al., 2006). In the presentation the registration of gaze behaviour (visual information) and movement behaviour in several different sports on the court or field will be used to illustrate that experts are superior to novices because they much better in "reading the game". Since experts and novices do differ in visual search strategies they employ, it has been assumed that experts have developed an ability to recognize advanced visual cues to 'buy' time, i.e. they are better anticipating future events. The important question is: can we use visual search strategies as an indicator for talent? In addition, ideas about visual attention training, that is, special practice to speed up the visual earning process of talented players will be discussed.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Faster Marathon Times by Measuring Human Performance

Cees van Bladel

Sports and Technology Theo Koomenlaan 7, 5644 HZ Eindhoven, The Netherlands cees@des.nl

ABSTRACT

This article describes a study to performance monitoring and the variables that influence performance of runners during the marathon of Eindhoven. In 2009 core temperature and cooling was investigated. In this research we have not found a relation between core temperature and running pace. But we do have new insights on how the human body reacts on running a marathon. In 2010 the test group will be expanded to 100 athletes and the number of variables will be expanded.

Author Keywords

Marathon, heat, running pace, dehydration.

INTRODUCTION

The marathon of Eindhoven, situated in the south of the Netherlands, is not only one of the most fastest in the world but also one of the most innovative marathons. The region of Eindhoven is known for the concentration of knowledge institutes and companies focused on technology innovation. Combining this with one of the largest sports events of the region, one created a once a year field lab focused on measuring on marathon runners. These measurements leads to the understanding of the human body and eventually into products that support or improve physical performance.

The first project started in 2007 on the development of the RunAlyser, a device with which the foot pattern throughout the marathon could be monitored. This foot pattern gives information about the running technique which affects the performance, effectiveness and risk of injuries of the specific athlete. During the performance the athlete received real time feedback on his running technique which improved his performance.

For 2009 and 2010 projects in the Eindhoven Marathon are based on the themes heat and dehydration. These themes are the main cause of serious medical problems during the marathon. In cool weather, there are few problems, but in extreme situations in the Netherlands 1/3 of the athletes did not finish. In such cases, a significant proportion of the athletes need medical help. These projects focus on preventing such problems.

MATERIALS AND METHODS

Heat and Running Pace

In 2009 one started a project which investigated the effect of external cooling of the body during the running performance. The core temperature of an athlete can rise above 40 degrees during a marathon. Therefore, in this project athletes are cooled by running through a mist shower. The hypotheses in this research is:

By external cooling of the human body, core temperature decreases and physical performance (running pace) will increase [1].

During the marathon fourteen runners are equipped to measure the following variables: heart rate, respiration, and skin temperature. These variables are measured with a Hidalgo Equivital system. Core temperature is measured by means of a pill that is ingested three hours before the start of the marathon. Running pace is measured with a GPS tracker (MYlaps). Based on age, weight and maximal oxygen uptake (VO2 max), the runners are divided into two groups (n=7). The intervention group runs, during the marathon course, twice through a mist shower and drinks water with a temperature of 15° . The mist shower is a 10 meter tunnel which sprays water on the runners. The control group drinks water with a temperature of 35°. The data focus is on the data in the range of 2 KM before the mist shower compared to two KM after the mist shower in both groups.

RESULTS

Based on a t test (sign. p < 0,05) in case of significance the following differences are notified:

- Pulse Frequency (± 3 beats);
- core temperature (± 0.1 ° C);
- Skin temperature (± 0.5 ° C).

Also the trend line of running pace in the range of two KM before de mist shower was compared to running pace with the trend line of two KM after the mist shower.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Average running pace of the intervention and control group during the marathon.

There are no significant result on average running pace between the intervention group and control group at the finish of the marathon (figure 1).

Both in the intervention group as in the control group the running pace decreases at the 20 KM point. Both groups are able to sprint to the finish. There was no significant difference in heart rate, or core temperature between the intervention and control group. The average core temperature of the runners increased after the start in the first 10 KM to 39 ° C. In both groups the heart rate and core temperature decreases after 30 KM. There was a significant difference in skin temperature (see figure 2).

The intervention group has a significant higher skin temperature then the control group after the first 10 KM of the marathon. For both groups skin temperature decreases after the start and increases again after 30 KM. An important fact that may have influenced these results is that almost all runners were faced with rain during the final part of the marathon.



Figure 2. Average skin temperature of the intervention and control group during the marathon.



Figure 3. Average speed in a interval 2 KM before first cooling intervention and after.

When we focus on the interval 2 KM before the mist shower, there was no significant difference between the intervention group and the control group (see figure 3 and 4). The results of this research are summarized in table 1.

| 1: increase → : no change Intervention Group ↓: decrease | | Group | Control Group | Control Group | |
|--|----------------|----------------|----------------|----------------|--|
| | Intervention 1 | Intervention 2 | Intervention 1 | Intervention 2 | |
| Speed | | ** | ** | ** | |
| Hart rate | ** | * | | 4 | |
| Core temperature | ↔ | \$ | ↔ | ** | |
| Skin Temperature | - St | 1 | ÷. | 4 | |

Table 1. Results of the research on heat and cooling.



Figure 4. Average speed in a interval 2 KM before second cooling intervention and after.

CONCLUSION AND DISCUSSION

In this research we rejected the hypothesis that cooling decreases core temperature and increases running pace. Level et al. [2] gives three possible explanations for a missing effect. First, core temperature is not the only variable that influences running pace. Also hydration level, motivation and temperature are important variables [3]. Second, also during the marathon outside temperature was only 12.8 °C to 16.7 °C and during the end both groups are cooled by rain which also minimizes the difference between the groups. And third, all test runners did not have a good pacing strategy due to inexperience.

However, we only presented the objective results of the research. In a questionnaire among the intervention group responded very positive on the question how they experienced running through the tunnel.

If this research is repeated we recommend the following:

- A higher frequency of cooling interventions;
- Trained athletes with a stable running pace;
- A larger research group.

Further Research Marathon Performance Dehydration

In 2010 the focus of the research is repeated with additional variables, especially hydration level. The human body loses a lot of water during the marathon this could lead to dehydration; the blood thickens. The athlete's heart must work harder to send blood with the necessary oxygen and energy to his muscles. This creates an increased heart rate,

which in most cases the athlete is forced to reduce the intensity and slows down.

For this year there are 100 test runners (n=50). The intervention consists of a 3 months training program. The hypothesis is:

Specific advice ensures optimal hydration level among runners. With this intervention runners are better able to regulate heartbeat so the runner can keep pace. It is expected that the group with a hydration advice have the best individual performance.

Besides the variables that are measured in 2009 we also measure hydration level before and after the marathon, and the amount of fluid the runner drinks during the marathon.

REFERENCES

- 1. Daanen, H.A., Van Es, E.M. et al. Heat strain and Gross efficiency during endurance exercise after lower, upper, or whole body precooling in the heat. *International Journal of Sports Medicine*, 27 (5) (2006), 379-388.
- 2. Levels, K., Tiemessen, I. Koude douche tijdens de marathon. *Sportgericht*, 64 (3) (2010), 36-39.
- 3. Cheuvront, S.N., Haymes, E.M. Thermoregulation and marathon running. Biological and environmental influences. *Sports Medicine* 31 (10) (2001), 743-762.

Symposium: Large and Small Scale Physiological Recordings in Behavioural Context

Gernot Riedel g.riedel@abdn.ac.uk **Bettina Platt** b.platt@abdn.ac.uk

School of Medical Sciences, University of Aberdeen Foresterhill Aberdeen, AB25 2ZD, United Kingdom

ABSTRACT

Coordinated rhythmic activity of neural populations gives rise to oscillatory local field potentials and large scale electroencephalograms at a broad range of frequencies. Synchronous rhythms are likely to reflect relevant information and frequency shifts may underlie experienceor behaviour-dependent functional interactions between neuronal assemblies. While the most common mode of synchronisation may be established through the local cohesive discharge of neighbouring cells or neurones with direct synaptic contacts, global synchronisation may also take place to establish widespread assemblies of disparate neural populations. To detect and analyse such global and local electrophysiological traits is one of the current challenges in basic and translational neuroscience.

This aim is made even more difficult when attempting to correlate electrophysiological data with stage- or taskrelated behaviour or even cognitive processes. It requires cohesive and linked recordings of physiological, spatial and behavioural responses synchronised and time-stamped in real time. Once aligned in the spatio-temporal domain, analysis needs to implement novel sorting strategies for correlational analysis applying linear and/or non-linear algorithms.

In this symposium, we seek to review some methodical progress focussing on large scale (global EEG) as compared to small scale (single unit) recordings in clearly defined behavioural paradigms in rodents. Speakers are selected because they utilise different technical products for physiological measurements (using cable, transmitter or microchip) and video-observation software of differing specification. We seek to generate intense discussion highlighting both the advantages but also the limits of each system and intend to foster a more intense interaction between manufacturer and scientist for product enhancement.

SYMPOSIUM CONTENT

Epidural EEG Recording Using Microchips in Behavioural Context

Bettina Platt, Andrea Pano, Amar Jyoti & Gernot Riedel (University of Aberdeen, Scotland).

Timed Behaviors in Mice

Valter Tucci & Glenda Lassi (Italian Institute of Technology, Italy) & Patrick M. Nolan (Mammalian Genetics Unit, Harwell, UK).

Route Finding in a Complex Maze in Wild-Type and CA1 NR-1 KO Mice: Hippocampal Local Field Potentials, Single Units and Relationship with Behaviour

Francesco Battaglia (SILS - Center for Neuroscience, Amsterdam, The Netherlands).

Use of Behavioral Outcome to Assess Cognitive State

Robert Hampson (Wake Forest University School of Medicine, USA).

Simultaneous Measurement of Brain Activity, Physiology & Behavior in Large Animals

Nadine Reefmann (Agroscope Reckenholz-Tänikon Research Station, Switzerland), Thomas Muehlemann & Martin Wolf (ETH and University Zürich, Switzerland), Beat Wechsler & Lorenz Gygax (Federal Veterinary Office, Switzerland).

Platform for Ambulatory Assessment of Psycho-Physiological Signals and Online Data Capture

Jürgen Stumpp & Panagiota Anastasopoulou (Karlsruhe Institute of Technology, Germany).

Experimental Design for Sternocleidomastoid Muscle Stress Measurement

CheeFai Tan, Wei Chen & Matthias Rauterberg (Technical University of Eindhoven, The Netherlands).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.
Peripheral Arterial Tone as an Index of ANS Trade-Off

Stas Krupenia (Thales Research and Technology, The Netherlands), Eldad Yechiam & Maya Arad (Israel Institute of Technology).

Using EEG Recordings to Examine the Relationships Between Sustained Attention and Types of Background Music in Individuals with ADHD

Chelsea Liang Ru Chew (Nanyang Technological University, Singapore).

Psychophysiological Data Collection in an Organizational Setting: Studying Interaction Between the Manager and Subordinate During Performance Review Discussion

Mikko Salminen, Pentti Henttonen & Niklas Ravaja (Center for Knowledge and Innovation Research, Finland), & Mikael Saarinen (Sensitiva Inc, Finland).

Extracellular Multi Unit Recording in Fear Conditioning in Mice Using a Telemetry Approach in an Automated Home Cage (DualCage) Environment

René F. Jansen, Anton W. Pieneman, Andries Ter Maat, (VU University Amsterdam, The Netherlands) Oliver Stiedl & Manfred Gahr (Max Planck Institute for Ornithology, Germany).

Epidural EEG Recordings Using Microchips in Behavioural Context

Bettina Platt b.platt@abdn.ac.uk Andrea Plano a.plano@abdn.ac.uk

Amar Jyoti a.jyoti@abdn.ac.uk Gernot Riedel g.riedel@abdn.ac.uk

School of Medical Sciences, University of Aberdeen Foresterhill, Aberdeen AB25 2ZD, Scotland

ABSTRACT

behavioural and electrophysiological Traditionally. measurements in animals are performed separately; however their combined use has considerably enriched our understanding of, for example, memory-related processes. Especially single unit recordings paved the way and highlighted performance-specific firing characteristics of individual neurones (place cells, phase cell, head direction cell - see other contributions to this symposium). Less well explored is the correlation between behavioural activity and global activity recorded brain through the electroencephalogram (EEG), particularly in mice. This may be due to technical limitations both in terms of hardware and software, such as time-stamping of events and complex analytical issues.

We here present the first EEG recordings from freely moving mice equipped with wireless microchips (Neurologger – NewBehavior). Unique features of the devices include 4 recording channels, an accelerometer to determine 3-D activity, and an infrared sensor for timestamping of external events. Devices were tested in two behavioural conditions in which previous use of videoobservation systems has provided behavioural indexes widely used to define and explain task performance.

Author Keywords

Video observation, subcranial EEG recording, sleep, disease models, cognition.

INTRODUCTION

Recording of brain activity is pivotal to our understanding of brain function and malfunction, and has recently experienced a revival for translational medicine and the development of powerful computational tools for complex data analyses. The (EEG) determines global electrical activity of the brain and is generally recorded from the scalp at multiple recording sites in humans. Corresponding animal studies utilize both surface and depth EEG individually or simultaneously, and these may be recorded at rest or as event-related potentials.

EEG traces are oscillations with characteristic frequency ranges (1-50Hz) that depend on the vigilance state (e.g. wakefulness or sleep), the location of the recoding sites, and the general health status of the subjects. It comprises the summation of synchronous dendritic events of large neuronal populations, favored by the radial orientation of cortical dendrites [1].

Compared with other modern imaging tools such as PET and MRI, EEG also is non-invasive in humans, comes at considerably lower costs and the relative ease of recording, brain activity is directly measured, and the temporal resolution is excellent. EEG permits determination of essential physiological processes, and EEG spectra can be utilized to determine e.g. sleep stages, ageing, cognition, CNS disorders and drug actions, to name but a few. At the same time, the spatial resolution is poor as activity recorded is by and large from the most superficial layers of the brain.

In rodents (especially mice), typical EEG recordings comprise the surgical implantation of surface/depth electrodes under stereotaxic control, which are then anchored to the skull in a small head-stage. To discriminate between sleep and activity patterns, additional EMG Teflon-coated silver wires are implanted into either neck- or eyelid muscles [2,3]. Animals are then returned for recovery and housed individually to prevent interference with the head-stage during social contacts and aggressive episodes. For recordings, the head-stage is either connected transmitter/transponder to а wire-less positioned subcutaneously in a skin pocket, or flexible lightweight cables that are counterbalanced through a weight on a string running over roller-bearing pulleys [4].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

This is a delicate set-up and restricts movement of mice spatially and can impinge on vigilance outcome by interference with sleep pattern (dangling cables will keep animals awake [5]. Therefore, the absence of tethering in a telemetric recording set-up would be preferred. At the same time, EMG silver wires undergo necrosis so that reliable recording of muscle activity will eventually cease after several weeks. This precludes repeated measurements of animals in a within-subject design during ageing studies. As a result, the vast majority of EEG studies are conducted as short-term experiments in between-subject, crosssectional study designs. We here set out to refine and optimize longitudinal recording techniques by using a novel untethered recording device and provide proof of principal that such studies can be used for a range of scientific research questions including ageing, preclinical diseasemodel comparison and in behavioral situation with cognitive relevance.

METHODS

Subjects

Mice (female and male) were derived from different sources and in some experiments carried various transgenes to generate Alzheimer or schizophrenia-like phenotypes. They were maintained either on a C57B1/6 x C3H background, or as C57BL/6 x CBA hybrids. All mice were group housed until surgery (3-5 per standard Macrolon cages), and kept in a controlled holding environment with a 12-hour light-dark cycle (lights on at 7am). The procedures concerning animal care and treatment were in accordance with international standards on animal welfare and Home Office (UK) regulations.

After surgery and at least 10 days recovery, mice were singly housed and assigned to the different experiments (see below), their recording quality tested and animals with poor signals were excluded from the study.

Surgery

Anesthesia was induced with 3% isoflurane in medical grade oxygen and maintained at 1.5% after the mouse was placed in a stereotaxic frame (Stoelting, US). The skull was exposed and epidural gold screw electrodes placed into bur holes at the following coordinates relative to Bregma [6] above i) medial prefrontal cortex (PFx: AP +2 mm/close to midline); and ii) parietal cortex and dorsal hippocampi (bilateral, AP -2 mm /1.5 mm lateral to midline). Surface recordings at this position are dominated by coherent hippocampal discharges [7] and will thus be referred to as the hippocampal recoding site. Reference and ground electrodes were placed at neutral locations superficial to parietal and occipital cortex respectively. Electrodes were assembled into a 7 pin adaptor and fixed to the skull by a mixture of Durelon dental cement and tissue glue. Once the cement had dried, animals were removed from the stereotaxic instrument, injected with 0.5 ml saline (intraperitoneal) and 0.01 µl analgesic (Temgesic; subcutaneous), returned to their home cages and placed in a heating cabinet for 60 minutes. Mice were weighed daily to monitor their recovery and at least 10 days recovery allowed before experiments commenced.

Home Cage Activity Monitor

Circadian activity was recorded in PhenoTyper home cages (Noldus, The Netherlands) through video observation techniques and XY coordinates recorded over 7 consecutive days as described previously [8]. PhenoTyper cages (30cm x 30cm x 35cm) have clear Perspex walls with a plastic floor covered with sawdust, fixed feeding station and water bottle attached to the front wall and an overhead infra-red sensitive camera connected a computer-assisted video tracking software (EthoVision 3.0, Noldus Information Technology, Wageningen, The Netherlands) for automatically recording of activity. Video tracking was performed at a rate of 12.5 samples /second. Ambulatory activity (distance moved) was extracted using the in-house 'Mnimi' software package with automatic exclusion of artifacts. Path length was averaged from a 5-day recording (excluding initial 2 days of habituation) into 1 hour bins and group means (+/- SEM) calculated for a 24hr period separated into light (sleep) and dark (activity) phases.

Object Recognition

Object recognition followed the protocol implemented by Good and Hale [9] with minor modifications. The apparatus was a white Perspex cylinder (50cm diameter; 50cm wall height), in which the animal's movement was recorded by an overhead camera and digitized to a PC-observation system (EthoVision Pro 3.1). Objects were tall to prevent mice from climbing onto them (15-20cm). Behavioural testing consisted of: 1) Habituation (2 days, 2 trials, 5 min, ITI: 2 min). During trial 1, the arena was empty; during trial 2, a single object was placed in its centre. 2) Object novelty: Presentation of two identical objects A (sample phase); objects were replaced by one identical (A) and one novel (B) object (test phase). 3) Spatial novelty: two novel objects C and D (sample phase); displacement of one object (test phase). Locomotor activity (path length) was determined and object exploration recorded (% time within in-object-zone (4cm)). Exclusion criteria: object bias during the sample phase or no exploration of objects.

Wireless EEG Recording and Analysis

EEG recordings commenced for 24 hours on day 3 of PhenoTyper recording. Wireless Neurologger (NewBehavior, Zurich, Switzerland) microchips weighing < 3g (approximately 10% of the body weight) were used for EEG registrations from freely behaving mice. Movement was detected via a built-in accelerometer. Neurologgers were set to record at 200 samples per second with a high pass filter of 0.25 and low pass filter of 70Hz, recordings were downloaded offline to a PC using USB plug-in docking stations.

Data retrieved were transformed with Matlab 7 (The MathWorks Inc., Natick, USA) and imported into SleepSign (Kissei Comtec Co. Ltd, Nagano, Japan) for

vigilance staging and extrapolation of spectral power (Fig. 1B). Vigilance stages (wakefulness, NREM and REM sleep) of 4 sec epochs/bins were identified based on combined Fast Fourier Transform (FFT; delta / theta ratio from hippocampal EEGs) and accelerometer activity (body movement). Automated staging was followed by visual inspection and corrections excluding any movement-related artifacts from spectral analyses. FFTs were finally calculated for each epoch with a resolution of 0.77 Hz, Hamming window smoothed and averaged. The EEG power spectra (1-20 Hz, PFx and hippocampus) for each vigilance state were normalized relative to the absolute maximum power over all frequency bands and averaged for each genotype/age-group for hippocampus and PFx (spectral bands: delta: 0.5-5 Hz, theta: 5-9 Hz, alpha: 9-14 Hz and beta: 14 - 20 Hz).

For object recognition tests, vigilance staging was not required as animals were exploring the environment continuously. Recording with EthoVision was set such that both objects were surrounded by virtual in-object-zones. Entry of the animal into such zones triggered a TTL pulse for time-stamping of the event through infra-red lights onto the Neurologger. Averaging of power spectra was conducted dependent on spatial distribution and FFT analysis followed outline above. Spatial distribution pattern were averaged and compared in all stages of the test.

Histology

After recordings, mice were terminally anaesthetized and intra-cardially perfused with saline followed by 4% paraformaldehyde in 0.1M phosphate buffer (PFA). Brains were removed, post-fixed in PFA overnight, photos taken of brain surfaces to determine electrode/screw locations based on indentation marks, wax embedded and sectioned (5µM).

Statistical Analysis

PhenoTyper and object recognition data are expressed as group means (+/- SEM) and statistically assessed using analysis of variance (ANOVA, 2-way or 1-way, with repeated measure where appropriate) with group (age and genotype) and time of the 12 h light/dark cycles as factors, followed by planned paired comparisons (t-test with Bonferroni correction) to determine the source of reliability.

For EEG power spectral analysis, a 2-way factorial ANOVA was conducted using group (genotype/age) and frequencies as discriminators. Post-hoc planned paired comparisons and frequency specific analyses were carried out on preselected frequency bands to determine effects between genotypes. The significance level in all calculations was set to P<0.05.

RESULTS

Experiment 1: Both activity and EEG were recorded by Neurologger microchips for at least 24 hours in PhenoTyper (Noldus) home cages. Circadian activity was measured as ambulation by EthoVision and accelerometer-based determination of movement. This enabled assessment of

sleep patterns, vigilance staging and quantification of stagespecific EEG power. Recordings are sensitive to age- and genotype specific variations and can thus be used for a range of studies.

Experiment 2: Mice equipped with EEG recording devices were tested in an object recognition paradigm. Video observation (EthoVision) generated spatial distribution maps in predefined zones (with familiar of novel objects), that form behaviourally relevant endpoints. These maps were space-stamped onto the continuous EEG to enable quantitative analysis of global EEG according to spatial distribution patterns. Fast Fourier Transformation of spectra confirmed alterations in spectral power specific to the spatial location and/or familiarity of the object.

CONCLUSION

The combination of video-observation with quantitative cable-free EEG recording provides a major step towards a combined psycho-physiological approach needed to improve translational tools in neurosciences. Together, they deliver powerful multidisciplinary attempt to provide both systems and cellular information, and they are applicable to a wide range of research projects. Wireless devices such as the Neurologger enable long-term home cage and training-induced EEG observations in a withinsubject design. Such a longitudinal study design is favourable for preclinical studies of experimental models that monitor different stages of disease progression. Such studies are also relevant in terms of reduction of animal numbers and cost savings for extensive drug studies.

- 1. Buzsaki, G. *Rhythms of the brain*. Oxford University Press, 2006.
- 2. Qu, W-M, Xu, X-H, Yan, M-M, Wang, Y-Q, Urade, Y, Huang, Z-L. Essential role of dopamin D2 receptor in the maintenance of wakefulness, but not homeostatic regulation of sleep, in mice. *Journal of Neuroscience*, *30*, 12 (2010), 4382-4389.
- 3. Nakano, H, Saito, K, Suzuki, K. Chronic implantation technique for monopolar G monitoring of epileptic seizures in mice. *Brain Res. Bull.*, *35* (1994), 261-268.
- Grenwis, J.E. Recent advances in telemetry promote further progress in reduction and refinement. NC3Rs #20, Mar 2010 1; <u>www.nc3rs.org.uk</u>.
- 5. Tang, X., Orchard, S.M., Liu, X., Stanford, L.D. Effect of varying recording cable weight and flexibility on activity and sleep in mice. *Sleep*, *27* (2004), 803-810.
- 6. Franklin, K.B.J., Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates*. Compact 3rd Edition: Academic Press, Sydney, 2008.
- 7. Megevand, P, Quairiaux, C, Lascano, AM, Kiss, JZ, Michel, CM. A mouse model for studying large-scale neuronal networks using EEG mapping techniques. *Neuroimage*, 42 (2008), 591-602.

- 8. Riedel, G, Fadda, P, McKillop-Smith, S, Pertwee, R.G., Platt, B, Robinson, L. Synthetic and plant-derived cannabinoid receptor antagonists show hypophagic properties in fasted and non-fasted mice. *Br. J. Pharmacol.*, *156* (2009), 1154-1166.
- 9. Good, M, Hale, G. Impaired visuospatial recognition memory but normal object novelty detection and relative familiarity judgments in adult mice expressing the APPswe Alzheimer's disease mutation. *Behav. Neurosci.*, 119 (2005), 884-891.

Timed Behaviors in Mice

Valter Tucci

valter.tucci@iit.it

Glenda Lassi

Department of Neuroscience and Brain Technology, Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy

ABSTRACT

The mouse has an extensive capacity to modulate timed responses and arm movements according to different environmental conditions and different genetic background may affect this ability. Recently, we have characterized a number of quantitative and qualitative meaningful endophenotypes related to mouse timing mechanisms and different motor actions. We have further investigated timing and motor behaviors in mice by acute experiments and long-term monitoring. Novel technological advances allowed us to simultaneous monitor brain and muscle electrical activity, spontaneous and conditioned behaviors as well as metabolic profiles during 24-hours in testing and home-cage environments. We have used these new technologies to screen for novel mutants and to characterize existing mouse lines.

Author Keywords

Timing, reaching and grasping, motor behaviors.

INTRODUCTION

A number of functions in organisms require accurate temporal processing of environmental inputs in order to orchestrate the appropriate output. Organisms can process temporal information and generate responses across a wide range of timescales (microseconds to years). The circadian system is not the only biological clock in the organism. Timing is central to many cognitive functions and motor actions. Many neuropsychiatric and neurological disorders present with timing deficits that lead to cognitive or motor symptoms. This shows the fundamental role of our ability to localize ourselves in time. Studies on conditioned behaviors in humans and animals have demonstrated that subjects retain a cognitive biological clock, which allows them to store memories of the duration of an inter-event interval, and subsequently to recall this information to determine the time of the conditioned response. In this respect, a crucial integrative property of the acting brain is to execute goal- directed motor actions with a proper timing and with a meaningful value.

METHODS

Recently, we have developed two main automated instruments that allow us to study timing and motor control in mice under the manipulation of biological (e.g. gene) and behavioral (e.g. risk assessment) variables.

Timing Screen

In this task, conditioning is directly evident in the timing of the conditioned response (peak procedure test). In particular, we have trained mice to collect a reward (a food pellet) at a specific time interval following a light event (20 seconds) in the home cage. We then use the timing of the mouse's response during a probe trial, in which no food reward is given, as a measure of learning. By using an automated system, we were able to test the mice at night, their normal period of activity, thus not effecting there natural sleep and/or daily rhythms.

Electrophysiological Correlates of the Timing Switch Task in Home-Cage Environment

This test has been developed by Balci et al. (Balci et la., 2008) in mice. Here mice are trained to perform in a timing task. Respect to other timing tests (e.g. the peak procedure test), the switch task requires for the mouse to judge when it is the optimal time to switch from one feeder to another within the cage. In particular, mice are asked to anticipate the appearance of a reward at 1 of 2 locations at each trial. On a fraction of the trials, the reward pellet comes in location "A" after a short latency (e.g. 3 seconds). On the other trials, the pellet is obtained after a long latency (e.g. 9 seconds) at the location "B". It has been shown that mice learn very fast that it is convenient to wait at the shortlatency location. If and when they reckon that the short latency has passed, they decide to leave the location "A" and to move (switching) to the location "B" (Balci et al., 2008). If they switch too soon or too late for those longlatency trials, the mice end up with no payoff. Interestingly, by changing the probability of short trials, the optimal

Patrick M. Nolan MRC, Mammalian Genetics Unit, Harwell, Didcot, Oxfordshire OX11 0RD, United Kingdom

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

switch time changes. The test allows us to test the timing learning of a mouse but also to quantitatively assess how this is performed under risk. The optimal performance at this test implies that the animal has an accurate representation of both endogenous and exogenous uncertainty of the temporal environment. We are interested to test the electrophysiological correlates under temporal uncertainty in mice. Furthermore, we test the relation between timing learning and sleep homeostasis. In doing that, we monitor EEG, muscle and temperature over long periods of time. For this study we use a wireless monitoring system (DSI) consisting of implantable transmitters that measure in mice physiological parameters such as: EEG, EMG, activity and body temperature. The telemetry system allows mice to move freely with no constrictions due to a cable. In fact this strategy allows us to subject mice to automated training in the home cage without any interference from the cable.

Home-Cage Automated Phenotyping

Because the existing protocols in timing learning are all very time-consuming and with a considerable variability due to the repeated daily sessions for the mouse in a different environment compared to its home-cage, we decided to invest efforts in refining the procedure. To respond to this necessity, we have developed a novel behavioural device which allows testing mice within their home-cage.

Reaching and Grasping Motor Behaviors

Reaching and grasping are fundamental goal-directed movements across many species, included mice. However, the use of rodents to investigate the mechanisms of reaching and grasping has been limited in genetics. Recently, we have characterized a number of quantitative and qualitative fine motor phenotypes that constitute significant properties of the mouse motor control system. In addition, in our lab we are exploring the possibility of testing specific aspects of arm movement and to study their significance when testing cognitive functions in mice.

Subjects

We decided to subject a number of mutant lines (circadian and cognitive mutants) and wild-type mice to cognitive timing tests and reaching and grasping tests to explore the possibility that gene mutations affects the timing learning ability of mice at a different timescale.

RESULTS

Within our studies we characterised a number of timing phenotypes in novel mouse models (e.g. within ENU mutagenesis screens) and in existing mouse mutants. In particular, we created a novel mouse model (*Nosy*) with a cognitive timing delay. We have identified phenodeviants for timing phenotypes in the G3 progeny of ENU mutagenised mice. These mice show an abnormal delay for response in the probe trials (in which no food reward is given), such that the mutant has a 50 second delay,

compared to the 20 second delay seen in wildtype mice (3 out of 32 mice had the phenotype). Initial inheritance testing of this pedigree, indicates a recessive mode of inheritance with incomplete penetrance.

For what regards the motor phenotype issue, we have developed a test, the MoRaG test, that has proven to be successful and rapid for detecting abnormalities/differences in inbred strains, genetically-modified and aged mice (Neuroscience. 2007 Jul 13;147(3):573-82.). Moreover, the test has well served to the characterisation of novel mouse models for hereditary and motor neuropathy (Dis Model Mech. 2009 Jul-Aug;2(7-8):359-73.) and for pathology affecting the strength in limb movement (PLoS One. 2010 Feb 9;5(2):e9137).

CONCLUSION

To our knowledge, the mutation carried by the *Nosy* line, would be the first genetic locus found to underlie timing learning in mice. Whether such an internal timing mechanism resembles a clock-like oscillator (Church, 1984) or is distributed over a cortical network (Karmarkar and Buonomano, 2007) is still debatable. A few studies have investigated the relationship between circadian clock mechanisms and timing learning, concluding that the two clock systems are independent (Cordes and Gallistel, 2008). However, the question has not well been addressed yet at the molecular level.

By using our Automated Nose Poking prototype we have observed a remarkable reduction of the time of training for each animal. In particular by using our set-up, mice reach a *steady state performance* within 10-15 days of training compared to a 1-month standard protocol.

The detection of nose poking behaviour in conditioned behaviour paradigms has become a common procedure to study cognitive processes in mice. Our study represents an optimisation of this procedure in mice.

Regarding the reaching and grasping experiments, we have observed that attention plays a role in the performance of arm movements in mice improving/diminishing (according to specific compounds) the accuracy to reaching behaviours. An ongoing *in vivo* electrophysiological investigation is now focusing on timed combination of muscle activation as a requirement for developing an appropriate motor learning in mice.

Ethical Statement

An authorized ethical committee (by the Italian law) has approved the experiments.

REFERENCES

 Achilli, F., Bros-Facer, V., Williams, H.P., Banks, G.T., AlQatari, M., Chia, R., Tucci, V., Groves, M., Nickols, C.D., Seburn, K.L., Kendall, R., Cader, M.Z., Talbot, K., van Minnen, J., Burgess, RW., Brandner, S., Martin, J.E., Koltzenburg, M., Greensmith, L., Nolan, P.M., Fisher, E.M. An ENU-induced mutation in mouse glycyl-tRNA synthetase (GARS) causes peripheral sensory and motor phenotypes creating a model of Charcot-Marie-Tooth type 2D peripheral neuropathy. *Dis Model Mech.*, 2, 7-8 (2009), 359-373.

- Balci, F., Papachristos, E.B., Gallistel, C.R., Brunner, D., Gibson, J., Shumyatsky, G.P. Interval timing in genetically modified mice: a simple paradigm. *Genes Brain Behav.*, 7, 3 (2008), 373-384.
- 3. Church, R.M. Properties of the internal clock. *Ann. N. Y. Acad. Sci.*, 423 (1984), 566-82.
- Cordes, S., Gallistel, C.R. Intact interval timing in circadian CLOCK mutants. *Brain Res.*, 28, 1227 (2008), 120-127.

- Cordes, S., Gallistel, C.R. Intact interval timing in circadian CLOCK mutants. *Brain Res.*, 28, 1227 (2008), 120-127.
- Miller G., Neilan M., Chia R., Gheryani N., Holt N., Charbit, A., Wells, S., Tucci, V., Lalanne, Z., Denny, P., Fisher, E.M., Cheeseman, M., Askew, G.N., Dear, T.N. ENU Mutagenesis Reveals a Novel Phenotype of Reduced Limb Strength in Mice Lacking Fibrillin 2. *PLoS ONE*, 5, 2 (2010), e9137.
- 7. Tucci, V., Achilli, F., Blanco, G., Lad, H.V., Wells, S., Godinho, S., Nolan, P.M. Reaching and grasping phenotypes in the mouse (Mus musculus): a characterization of inbred strains and mutant lines. *Neuroscience*, *147*, 3 (2007), 573-82.

Route Finding in a Complex Maze in Wild-Type and CA1 NR-1 KO Mice: Hippocampal Local Field Potentials, Single Units and Relationship with Behavior

Francesco P. Battaglia

SILS - Center for Neuroscience Science Park 904 1098XH Amsterdam F.P.Battaglia@uva.nl

ABSTRACT

Spatial navigation relies on the brain ability to combine information from multiple streams in order to generate a cognitive map, which is then used, for example, to find routes to a goal. This takes place in a network of brain structures centered on the hippocampus. Hippocampal place cells, cells that activate only when the animal is in a particular location in the environment combine sensory information with the estimate of animal's position updated by self-motion signals. Synaptic plasticity is important for these processes: Transgenic mice with a disfunctional NR-1 receptor in the CA1 hippocampal subfield are impaired in a route finding task on a star-shaped maze. We use a combination of neural ensemble electrophysiology and behavioral video tracking to analyze place cells in these mice and wild-type controls, showing how they adapt to the strategy adopted by the animal for route finding.

Author Keywords

Hippocampus, spatial behavior, place cells, NMDA, transgenic models, mouse.

INTRODUCTION

Navigation in a complex environment can rely on several different strategies: for example, subjects may decide on the route to take based on external landmarks (allocentric strategies), or memorize a well-known route as a sequence of body movements (egocentric strategy). The hippocampus is at the center of the network of brain structures supporting spatial memory and route computations, embodying the Cogntive Map function that have been proposed in the '40s by Tolman (1). In rodents, interference with hippocampal function with lesions studies, see e.g. (2), pharmacological see e.g. (3) and genomic means (4), impairs the animals' ability to navigate to a goal. On the other hand, when a route can be computed in terms of a simple set of body movements (e.g. right/left

turns), the role of hippocampus seems less important (5). Interestingly, different hippocampal subregions appear to have different involvement in the acquisition of spatial memories (4), with CA1 disruption being more effective than intervention on CA3. For these reasons, the availability of specific genomic manipulation that affect only one of the hippocampal subfield have been particularly fruitful in order to dissect the neural circuitry of spatial navigation. In particular, a series of knockout models have targeted the NR-1 subunit of the NMDA receptor specifically in CA1 (6), CA3 (7) an the dentate gyrus (8), offering an unprecedented chance to explore how this receptor activity affects synaptic and system plasticity in each of these structures.

In addition to these behavioral effects, the implication of the hippocampus in spatial processing is critically supported by the correlation between space and the activity of hippocampal cells: in all hippocampal substructures place cells have been found, cells that activate each one as the animal traverses a different place in the environment. Taken together, place cells activities form a veritable map of the environment, which can be used to localize the animal. Place cells activity is the result of a computation involving multiple inputs. Most importantly, the hippocampus has to combine inputs related self motion, or path integration (9) and external cues. In order to dissociate these different contributions, and different navigational strategies, Rondi-Reig and co-workers. Rondi-Reig and co-workers (10) tested wild-type and CA1 NR-1 KO mice in a complex, star-shaped maze, where animals had to find the one (out of five) rewarded arm. In standard trials, mice always started from the same departure arm. NR-1 KO mice were found to be impaired in acquiring this task. In probe trials, mice started from a different arm, so that, if they followed a body-turn based strategy, they should end up in an arm different from the goal arm. If however, they based their route on external cues, they should still be able to reach the goal arm. Thus, NMDA receptors in CA1 are important for this type of route learning. However, nothing is known about how hippocampal cells encode routes computed according to different strategies.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

We recorded neural activity in the CA1 hippocampal subfield, extracted instantaneous mouse position by video tracking and found that a large proportion of neurons there were place cells. Interestingly, during probe trials, place cells shifted their place cells according to the strategy used by the animal: in "egocentric" trials place fields rotated as if the new departure arm corresponded to the standard one. In "allocentric" trials, the map was instead preserved. These results suggest that the hippocampus is fully involved in the system responsible for path finding. We are currently analyzing the differences between KO and wild-type mice, also with respect to Local Field Potential (LFP) activities and behavioral learning curves.

MATERIALS AND METHODS

Animals

NR-1 KO specific for CA1 mice (C57Bl6 background) were bred in the lab of Dr. Rondi-Reig at the Université de Paris 6. They were transported in the Amsterdam lab upon weaning, in a litter contaning both homozygous (i.e. with an active mutation) and heterozygous subjects acting as controls. Behavioral training on the star-maze was performed for 10 sessions in 5 days. After that, a pair of subject composed by one control and one KO mouse was selected for surgical implant. Because the KO mutation remains specifically expressed in CA1 only until 2-1/2 months of age (11), all experimental procedures were performed up to that age limit. All experiments were carried our in accordance with the Dutch law, and upon approval of the institutional Committee for Animal Usage in Research (DEC).

Surgical Procedures

A microdrive containing 6 independently movable tetrodes, weighing less than 2 grams (12) was implanted over the right dorsal hippocampus (bregma AP: -2.0 mm ML: 2.2 mm) and anchored to the skull with micro-screws, under isoflurane-buprenorphine anesthesia. Tetrodes are bundles of 4 13 μ m nichrome wires, which allow, by triangulation, discrimination of up to ~12 cells per tetrode. In general, our setup allows yields of 10-20 simultaneously recorded hippocampal cells.

After surgery, tetrodes were gradually lowered in order to reach the pyramidal layer of the CA1 hippocampal subfield. Electrode location was first assessed by distinctive signs in the LFP (theta rhythm, sharp wave/ripples compleces) and confirmed *ex-post* by histological examination.

Electrophysiological Techniques, Data Pre-Processing

Tetrode signals were amplified 2000x, bandpass filtered (600-6000 Hz) and digitized at 30 kHz sampling rate by the Neuralynx Cheetah data acquisition system (Neuralynx Inc., Bozeman MT U.S.A.). Spikes from the surrounding cells were detected by a threshold mechanism and a 1 ms snippet was recorded on disk for each spike. Additionally, the same tetrode signals were low pass filtered (475 Hz cutoff) and acquired as LFP.



Figure 1. The star-maze, the behavioral apparatus used for this experiment.

After recording, spike signals were sorted and assigned to putative emitting neurons by means of a semi-automated clustering algorithm, KlustaKwik (13), further refined manually in the Klusters user interface (14).

Behavioral Monitoring

Correlating neural signals with animals position requires precise tracking of the mouse position. We accomplished this with the Ethovision XT software package, which was analyzing signal form a color video camera placed directly over the center of the maze. Synchronization of video and electrophysiological signals was obtained by recording timestamps signals from Ethovision (TCAP signals) in one of the Cheetah system's digital inputs. Custom-written MATLAB routines decoded the TCAP signals, providing a lookup table between the Ethovision and Cheetah timelines.

Behavioral Task.

The behavioral task took place in the starmaze (10), a maze formed by 5 alleys arranged in a pentagon, and 5 further arms connecting radially to the vertex of the pentagon, on its exterior (Figure 1). The maze was located in a room lined with black curtains, with a number of large geometrical cues hanging from the curtains. We placed mildly food-deprived mice in a departure (D) arm, and had to find a sugar pellet reward in another arm, the target (T) arm. In order to avoid that mice could find the reward by an odor trace, all arms are baited with sugar pellet, but the pellets are covered by a Plexiglas grill that makes them inaccessible in all arms but the T arm.

In probe trials, which occur every 3-5 normal trials, the mouse is placed in a different arm (P arm) at the beginning of the trial. In general probe trials are classified as allocentric if the mouse enters the T arm, egocentric if it ends up in the arm that bears the same spatial relationship with the P arm as the T arm does with the D arm. Trials with different outcomes are classified as error trials.

Mice are trained for 5 days (two 15 trials sessions per day). At the end of that period, animals animals that did not learned the task are discarded, and two of the remaining ones (1 mutant, 1 control) are implanted. The pair of implanted mice undergoes recording sessions during the



Figure 2. Place field map for a cells rotating its spatial correlates during an egocentric probe trial.

same days, with a counterbalanced schedule. Recording sessions entail 2 20-30 minutes sleep episodes before and after task performance.

RESULTS

After pre-training, both control and mutant mice were able to consistently find the T arm. However, consistently with Rondi-Reig et al. (2006), mutant performance was significantly lower (p < 0.05 2-way ANOVA), using two specifically designed behavioral measures (15). During probe trials, both mutant and control mice were about equally likely to use a allocentric or egocentric strategy.We recorded from 5 control and 5 mutants mice (1 control mouse had to be excluded from analysis due to poor signal quality). From these mice, a total of 369 cells were recorded and so far analyzed. Data analysis is now under way: in our initial findings, there is a slight impairment in the spatial selectivity of CA1 place cells in mutants. One aspect of interest in the data is whether strategy selection affects



Figure 3. A second example analogous to Figure 2.

place cells firing. In fact, in egocentric probe trials, place cells tended to rotate their place field, maintaning a consistent relationship with the departure point. In contrast, during allocentric trial, different populations of place cells encode the visited arms (which differ from the arms traversed during regular trials; Figures 2-3). We are at the moment quantifying possible differences in this behavior between the mutants and the controls.

CONCLUSIONS

Our combination of neural ensemble electrophysiology in the mouse, genomic manipulations, and behavioral tracking offers novel possibilities for dissecting the neural basis of behavior. In addition to the mentioned results, we are now in the process of analyzing the detailed temporal structure of neural ensemble activity and its relationship with LFP oscillations.

- 1. O'Keefe, J., L. Nadel. *The hippocampus as a cognitive map*, Oxford University Press, Oxford, 1978.
- Morris, R., P. Garrud, J. Rawlins, J. O'Keefe. *Nature*, 297 (1982), 681-683.
- 3. Eichenbaum, H., C. Stewart, R.G. Morris. J. Neurosci., 10 (1990), 3531-3542.
- Nakazawa, K., T.J. McHugh, M.A. Wilson, S. Tonegawa. *Nat. Rev. Neurosci.*, 5, (2004), 361-372.
- 5. Packard, M.G., J. L. McGaugh. Neurobiology of Learning and Memory, 65 (1996), 65-72.
- Tsien, J.Z., D.F. Chen, D. Gerber, C. Tom *et al. Cell*, 87 (1996), 1317-1326.
- 7. Nakazawa, K., M.C. Quirk, R.A. Chitwood, M. Watanabe, *et al. Science*, 297 (2002), 211-218.
- McHugh T.J., M.W. Jones, J.J. Quinn, N. Balthasar *et al. Science*, 317 (2007), 94-99.
- McNaughton, B.L., F.P. Battaglia, O. Jensen, E.I. Moser, M.B. Moser. *Nat. Rev. Neurosci.*, 7 (2006), 663-678.
- 10. Rondi-Reig, L., G.H. Petit, C. Tobin, S. Tonegawa *et al.* J. Neurosci. 26 (2006), 4071-4081.
- Fukaya, M., A. Kato, C. Lovett, S. Tonegawa, M. Watanabe. *Proc. Natl. Acad. Sci. USA*, 100 (2003), 4855-4860.
- Battaglia, F.P., T. Kalenscher, H. Cabral, J. Winkel et al. J. Neurosci. Methods., 178 (2009), 291-300.
- Harris, K.D., D.A. Henze, J. Csicsvari, H. Hirase, G. Buzsáki. J. Neurophysiol., 84 (2000), 401-14.
- 14. Hazan, L., M. Zugaro, G. Buzsáki. J. Neurosci. Methods, 155 (2006), 207-216.
- 15. C. Fouquet, G. H. Petit, A. Auffret, E. Gaillard *et al.*, *Neurobiol Aging* (2009).

Use of Behavioral Outcome to Assess Cognitive State

Robert E. Hampson

Dept. of Physiology & Pharmacology Wake Forest University School of Medicine Winston-Salem, NC 27157-1083, USA rhampson@wfubmc.edu

ABSTRACT

Historically neurological, psychological, and physiological measures have been used to assess the neural status or "brain state" corresponding to cognitive performance and behavior. We have recently demonstrated a means by which behavioral assessments can begin to assess cognition and brain state by searching for distinct patterns of neural activity and observing the behavioral results of those patterns. This presentation will explore the use of mean neural activity derived across specific behavioral states, then examining the neural patterns on single trials/events to determine how that behavioral or cognitive state may be altered on a trial-by-trial basis. Within this context it is then possible to determine how a given state arises from the preceding state, and further assess the effects of cognitive workload, stress, and pharmacological manipulation. Studies from rodent and nonhuman primate reveal correlations between past and present neural state with behavior that can be used to predict future cognitive state along with behavioral outcome. Facilitation or impairment of the animal's ability to perform mnemonic tasks can thus be produced by pharmacologically manipulation to confirm the role of specific neural circuits as well as neurotransmitters and neuromodulators within those circuits. Mathematical modeling of the relationship between recorded neural patterns and underlying cognitive state is currently underway to yield algorithms and devices that may eventually allow restoration of impaired cognitive function due to neural damage. These studies have concentrated on three primary brain regions within which it is possible to correlate specific neural firing patterns with behavior and vice versa. The first of these areas is the prefrontal cortex or PFC. which is responsible for executive function, and some forms of short-term memory. The second area is the basal ganglia, particularly the striatum, which mediates behavioral motivation and reward. The

third area is the hippocampus and medial temporal lobe, which mediate memory, association and cognitive decision. Current theories view prefrontal cortex, basal ganglia and medial temporal lobe as a functional circuit that allocates specific types of cognitive processing (i.e. memory, rule implementation, response execution) proportionally across each but that each has neuronal representation that can independently or conjointly correlate with specific behavioral events.

Prefrontal cortical function has been the emphasis of several new theoretical notions which integrate it with other frontal cortical regions and also structures in the basal ganglia. The more traditional notion of the dorsal prefrontal cortex subserving "executive function" as prominent in human clinical literature, appears to have evolved into the role of "rule planning and use." Neurons in the prefrontal cortex (ventral and dorsal) appear to be "most" responsive to task features that are unrelated to particular stimulus features or response components of the task, but rather rules or strategies that must be employed to satisfy experimental contingencies. Thus PFC neural firing patterns frequently correlate with specific behavioral events that signal context switching or application of different behavioral "rules." Likewise experimental design that emphasize a switching context, increasing cognitive workload, or application of various response rules can reveal the status of prefrontal neural activity on the basis of behavioral measures. Specifically we have shown that prefrontal neurons correlate with cognitive workload in a behavioral task. At the same time we have measured the frequency of pupil dilation, an independent measure of cognitive workload, that correlates with behavioral outcome as well as response latency (Hampson et al. Behav. Brain Res., 212 (2010), 1-11). Thus in this instance, the behavioral outcome reveals as much about prefrontal neural activity, as the neural activity reveals about the behavior.

Neurons recorded in the basal ganglia often encode the reward or outcome properties of behavioral tasks that feature rule-based learning and performance. Single neurons in the putamen showed a progressive increase in firing with learning. Dorsal striatal neurons appear to function in the role of "error detection" within a behavioral task, while ventral striatal neurons respond to salience and magnitude of reward. In addition, striatal neurons

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

differentially respond to drug versus natural rewards (Opris et al. *Neuroscience*, *163* (2009), 40-54) as well as which type of reward the subject will select. Thus, while basal ganglia play a major role in controlling responsiveness, the response to a behavioral stimulus can in turn be used to predict the underlying neural state, thereby predicting future behavioral responses.

For more than 50 years, the hippocampus and medial temporal lobe have been linked to learning and memory. Although there has been continual refinement of theories of hippocampal function, it is clear that damage to the hippocampus and associated areas impair spatial, as well as nonspatial memory. It has become apparent from lesion studies that the hippocampus is essential to representing not just position or "place", but relationships between stimuli (especially spatial stimuli). A novel role for hippocampus has been demonstrated in a visual delayed-match-to-sample task for nonhuman primates in which we have identified patterns of task-specific firing in the hippocampus that (1) fire differentially depending on the phase of the DMS trial, and (2) encode aspects of the stimuli that facilitate retention of object specific information over the interposed delay period. This latter characteristic was unexpected, in that hippocampal neurons demonstrated a "cognitive rule" by responding to specific categories of visual stimuli which could quickly discriminate between distractor stimui and identify the appropriate stimulus which matched prior encountered daily within the task (Hampson et al., Proc. Nat. Acad. Sci. USA, 101 (2004), 3184-3189). In another example of behavior revealing the underlying cognitive

processing, it was possible to predict stimuli to which hippocampal neurons would respond, on the basis of specific visual features of the stimulus.

This laboratory has taken advantage of the identification of discrete spatial and temporal patterns of hippocampal neural activity that can predict behavioral outcome in rodents performing a delayed match to sample task. Hippocampal neural activity within the task is analyzed via nonlinear systems analysis to identify the fine temporal structure and predict not only behavioral outcomes but the subsequent status of neural activity within hippocampus. Further correlation of sequences of behavioral trials with each new neural and cognitive state have yielded a model of the interaction between detection of a stimulus, cognitive decision-making, behavioral response, and the underlying cognitive state on top of which subsequent processing will occur. In this model we find that the behavioral history modulated neural encoding on any given trial. Subsequently the neural encoding on each trial predicts the behavioral outcome on that trial. This neural-behavioral circuit illustrates the importance of behavioral measures as a means of assessing underlying cognitive state on par with the use of neural cognitive recording to assess behavioral performance. In addition future development of prosthetic devices intended to replace or restore damaged brain areas will rely heavily on behavioral validation in studies such as those cited here in which neural recordings inform behavioral studies and behavioral outcome reveal the underlying brain activity.

Simultaneous Measurement of Brain Activity, Physiology & Behavior in Large Animals

Nadine Reefmann¹

Centre for Proper Housing of Ruminants and Pigs, Federal Veterinary Office Agroscope Reckenholz-Tänikon Research Station ART, Tänikon, 8356 Ettenhausen, Switzerland Nadine.Reefmann@hmh.slu.se

Beat Wechsler

Thomas Muehlemann

Martin Wolf

Molecular Imaging and Functional Pharmacology, Institute for Biomedical Engineering, ETH and University Zürich, Wolfgang-Pauli-Strasse 27, 8093 Zürich, Switzerland

Lorenz Gygax

Centre for Proper Housing of Ruminants and Pigs, Federal Veterinary Office Agroscope Reckenholz-Tänikon Research Station ART, Tänikon, 8356 Ettenhausen, Switzerland

ABSTRACT

Assessing an animal's neuro-physiological and behavioral reaction to external stimuli simultaneously, continuously, and non-invasively is a major aim of applied animal welfare research to interpret a situation from the animal's perspective. The purpose of this study was to assess responses of large animals (sheep) to 3 model situations with positive, intermediate and negative emotional valence. The following reactions of up to 19 sheep were recorded: brain activity using functional near-infrared spectroscopy, facial expressions (ear postures and eye aperture), electrocardiogram, respiration rate, and electro-dermal activity. All variables were successfully recorded in the sheep and are promising to assess stress and well-being in sheep and potentially other large animals. The combination of several methods resulted in a comprehensive picture of the animals' reactions to environmental stimuli, initiating innovative animal research that combines neuroimaging, physiological monitoring and behavioral observations.

Author Keywords

animal welfare, cardiac activity, respiration, electrodermal activity, emotion, facial expression, neuroimaging, sheep

INTRODUCTION

Concurrently assessing neuro-physiological and behavioral changes in animals is highly useful for a comprehensive

picture of how an animal reacts to its environment. Psychophysiological research in humans has often focused on assessing both behavioral and physiological reactions at the same time [2]. Observations of facial expressions were combined with cardio-respiratory and electro-dermal activity. The measurement of brain activity or associated hemodynamic oxygenation changes has become increasingly applied in humans with the availability of corresponding techniques [1]. In animals, however, these methods have not been widely applied, and measuring brain activity remains a particular challenge in non-sedated animals. Here, we report concurrent assessment of hemodynamic changes in the brain, cardio-respiratory activity and facial expressions in large mammals. Electrodermal activity was assessed in a similar setup. Our aim was to evaluate which of these methods are most promising to assess the animals' positive and negative emotional reactions. In addition, we investigated correlations between physiological and behavioral variables with regard to the possibility of substituting the latter with the former, and if any of those variables could be assessed at shorter time periods (10 seconds) in order to decrease the measurement duration and effort of signal analysis.

METHODS

Animals and Testing

Nineteen female sheep were used as experimental animals and housed at the Agroscope Reckenholz-Tänikon Research Station ART, Tänikon in Switzerland. The animals' reactions to external stimuli, which were very likely to elicit different emotional reactions, were assessed in an experimental area within their home pen. Being equipped with the measurement devices described below (see Figure 1), they were exposed to the following three exemplary situations ranging from positive to negative emotional

¹Also Swedish University of Agricultural Sciences, Dept. of Animal Environment and Health, Box 7068, 75007 Uppsala, Sweden.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

valence: i) positive: animals were voluntarily groomed by a familiar human, ii) intermediate: animals were not groomed and free to move in the experimental area, iii) negative: separation from other sheep. The study was licensed by the Cantonal Veterinary Office.

Neuroimaging

Functional changes in cerebral oxygenation were assessed using a specifically designed wireless near-infrared spectroscopy sensor [3]. A halter held the small sensor (35mm x 22mm x 80mm) on the sheep's head. Light from four near-infrared sources (760nm and 870nm each) was emitted through the intact skull into the brain, and two detectors collected the transmitted light. Data were wirelessly transmitted to a nearby stationary receiver and mean changes in oxygenated and deoxygenated hemoglobin were compared for up to 9 repeated stimulations of grooming. Data were analyzed for differential reactions with respect to the localization of hemodynamic changes, i.e. right and left hemisphere, rostral and caudal as well as shallow and deep brain oxygenation. For more details see [4].

Physiology

The electrocardiogram of an animal was continuously recorded using three electrodes by means of a human Holter recorder (Modular Digital Holter Recorder, Lifecard CF, DelMar Reynolds GmbH). Using the Pathfinder software provided by the same company, mean heart rate (beats per minute) and heart rate variability (root mean square of successive difference, RMSSD) were calculated and compared across the three situations. For more details regarding all the physiological measurements, see [7].

A commercial logger (MSR145W, Modular Signal Recorder Electronics GmbH) was fixed on the animal's depilated skin with a breathable plaster in order to detect the relative body-surface humidity [%] and the body-surface temperature [°C] on the animals' skin of the rump once per second. These data were later analyzed as mean



Figure 1. Sheep wearing measurement equipment (adapted from [7]: a) electrocardiogram holter and b) electrodes, c) electro-dermal activity device, d) respiration rate device, e) eye camera, f) battery pack, g) fNIRS sensor.

temperature, humidity and variance of both measures over a given time period.

An extendable belt (1132 Pneumotrace II, UFI) in front of the hind legs around the abdomen of a sheep generated a continuous signal for the relative extension of the belt during inspiration and exhalation. The signal was saved at a rate of 10 Hz by the logger used for electro-dermal recordings, and the respiration rate was determined from these recordings by a smoothed spectrogram of the signal's time series in S-PLUS (Version 7.0 for Windows).

Behavior

In analogy to human facial expressions, the animals' ears and eyes were observed. A number of different ear postures and movements were recorded in continuous direct observations (see [5] for more details). Two finger cameras (DV-2000B Weatherproof, CCD B&W finger camera, Conrad Electronics) placed on the sheep's head recorded the pictures for analyzing the eye aperture and amount of visible eye white of the sheep. The size of the eyes were electronically measured on the digitized images with onscreen mouse clicks in R, and compared across the situations.

RESULTS & DISCUSSION

All measures were simultaneously assessable in sheep using the devices as described. Cerebral changes in relative tissue oxygenation could be assessed during the positive compared to the intermediate situation (please see [4] for details). Motion artifacts in the negative situation were too strong for reliable data analysis. Nevertheless, the fNIRS sensor is the only device that can be used to measure hemodynamic oxygenation changes non-invasively in nonsedated and moving animals.

The device to assess cardiac activity was well suited to record the animals' electrocardiogram. The analysis tool is programmed for humans and special adaptation of the software to detect R-peaks specifically for other species would be desirable in the future to analyze the data in less time. The respiration rate was measurable in sheep using the Pneumotrace belt. Raw data should be pre-processed to remove movement artifacts before automatically calculating mean respiration rate based on the signal's time series during a given time period. Assessing the relative electrodermal activity posed an easy way to monitor body surface humidity and temperature of the animals, which are related to sympathetic nervous system activity.

Observations of ear postures presented the most noninvasive way to externally monitor an animal's reaction to a situation. The number of ear-posture changes or, for example, forward ear postures (Figure 2), were higher in the negative than the positive emotional situation (for more results see [5]). The ear postures can be recorded in direct observations or video-recorded and later analyzed using commercially available software for behavioral analysis.



Figure 2. Sheep's behavioural and physiological responses to separation from group members (SP, negative valence), standing in the feeding area (ST, intermediate valence), and grooming (G, positive valence); adapted from [7].

Observations of the eyes of animals using the robust and water-proof finger cameras were also possible. Nevertheless, the development of eye trackers for animals would be highly appreciated by researchers since these observations are not possible in direct observations and would contribute greatly to assess sympathetic physiological responses in animals.

The number of ear-posture changes and the relative eye aperture both correlated strongly negative with heart rate variability, indicating that behavioral observations could replace the monitoring of cardiac activity (see [7] for more details). Given that some physiological reactions were reflected in the behavioral responses, future experiments do not have to assess all measures so that the required time and effort can be reduced but the animal's reaction to environmental stimuli is still fully assessed.

With respect to the measurement duration, cardiac activity is regulated very rapidly and can be assessed within tensecond time windows. This short interval also seemed to be sufficient for assessing eye aperture (Figure 2). All other physiological and behavioral measures should be assessed continuously within longer time windows to reliably assess the animals' reactions. Ear postures [5] and respiration rate, for example, are best monitored within at least 30-second time windows.

CONCLUSION

This study demonstrates that cerebral, physiological and behavioral responses can be concurrently, continuously and non-invasively assessed in sheep. Innovative behavioral measures like ear postures may even substitute physiological reactions, and measuring hemodynamic changes is possible in unsedated and freely moving animals, so that a whole new range of possibilities opens up to assess subjective states in animals in response to their environment.

- 1. Bunge, S.A., Kahn, I. and Larry, R.S. Cognition: an overview of neuroimaging techniques. In *Encyclopedia* of *Neuroscience*, Academic Press (2009), 1063-1067.
- 2. Hubert, W. and de Jong-Meyer, R. Psychophysiological response patterns to positive and negative film stimuli. *Biological Psychology 31*, 1 (1990), 73-93.
- 3. Muehlemann, T., Haensse, D. and Wolf, M. Wireless miniaturized in-vivo near infrared imaging. *Optics Express 16*, 14 (2008), 10323-10330.
- 4. Muehlemann, T., Reefmann, N., Wechsler, B., Wolf, M. and Gygax, L. Bad mood increases cerebral pleasure response in sheep studied using near-infrared light. Submitted to *NeuroImage*.
- 5. Reefmann, N., Buetikofer Kaszàs, F., Wechsler, B. and Gygax, L. Ear and tail postures as indicators of emotional valence in sheep. *Applied Animal Behaviour Science*, *118*, 3-4 (2009), 199-207.
- 6. Reefmann, N., Buetikofer Kaszàs, F., Wechsler, B. and Gygax, L. Physiological expression of emotional reactions in sheep. *Physiology & Behavior, 98*, 1-2 (2009), 235-241.
- 7. Reefmann, N., Wechsler, B. and Gygax, L. Behavioural and physiological assessment of positive and negative emotion in sheep. *Animal Behaviour*, *78*, 3 (2009), 651-659.

Platform for Ambulatory Assessment of Psycho-Physiological Signals and Online Data Capture

Jürgen Stumpp Juergen.Stumpp@kit.edu Panagiota Anastasopoulou Panagiota.Anastasopoulou@kit.edu

Karlsruhe Institute of Technology (KIT) Fritz-Erler-Straße 1-3, 76133 Karlsruhe, Germany

ABSTRACT

Over the last years there has been an increasing interest in finding new methods for capturing psychological, behavioral and physiological data in real-time using infield data acquisition systems. Within our research group a system for ambulatory assessment of psycho-physiological signals for real-time data capture has been developed. The system is based on Smartphones which are equipped with software to record contextual and subjective data from participants and have that information transmitted wirelessly to a central online database. An online database offers the possibility of collecting, storing and analyzing all study related data in a central point. This paper provides an overview of this system.

Author Keywords

Platform, Ambulatory Assessment, Smartphones, Selfreport method, Field Studies, online database

ACM Classification Keywords

J.4 Social and Behavioral Sciences: Psychology

INTRODUCTION

Self-report methods for understanding everyday human behavior in form of diaries or questionnaires are a very common procedure in the field of medicine, psychology and sociology. In the past, studies that include self-report have mainly been done via paper and pencil; participants were given questionnaires after the study and were asked to fill them with their thoughts or feelings. Such procedures have the disadvantage that the participants can only estimate the events in the past and thereby make false statements. Paper diaries try to solve this problem by letting the participant track the events during the study [1]. But the participants often fill out the diary at the end of the day which also leads to false estimation. An electronic diary can exactly track the times when an entry was made which

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

makes it possible to better characterize the statement of the participant.

In the last years, a wide range of hardware and software solutions, focusing on multichannel systems for physiological data acquisition and hand-held computer based experience sampling systems has been developed [2]. However, the systems that have been developed up to now do not provide one solution that combines all the key features and that would set a standard platform for studies in different research fields.

The main goal of this paper is to present all components of a mobile system that collects physiological, psychological and contextual data (e.g. the geographical position) and stores them in a central online server. The mobile data collecting software is designed to run on Smartphones. These devices are not only a very good platform for mobile data acquisition they also provide multimedia features such as digital photo or video capture. Once the data is recorded, using the Smartphones connectivity to the internet, the data can be transmitted to a central online database. This feature provides the researchers with the ability of having real-time information regarding the ongoing study.

SYSTEM OVERVIEW

Mobile Sensors

Smart mobile sensors are recording physiological signals like the electrocardiogram, the activity or the galvanic skin response. From these physiological signals, important features are extracted. These features are transmitted to the



Figure 1. Schematic overview of the system.



Figure 2. Components of the data collecting platform.

Smartphone over a wireless body area network. For this wireless technology, Bluetooth has been the best solution due to its technical benefits and also because many vital sensors are already outfitted with the necessary radio hardware. Depending on the study either a continuous stream of extracted features is transmitted to the Smartphone or, if this consumes too much power, events can be transmitted as soon as features reach a defined threshold.

Smartphone

On the Smartphone we use the open-source software MyExperience that runs on Windows Mobile devices. Besides the synchronous recording of the features received from the mobile sensors MyExperience also collects self-reports of the participant to gather quantitative and qualitative data on human behaviors in the field [3]. MyExperience is based on a three tier architecture of sensors, triggers and actions. Triggers use sensor event data to conditionally launch actions. E.g. Self-reports can be questionnaires which are triggered by features from mobile sensors [4].

We extended MyExperience to transmit the collected physiological, psychological and contextual data to an online server.

Server

In order to achieve a modular, flexible and reusable online platform for study-relevant data collection and visualization, there are several things to focus on. First of all it must be a system capable of handling any kind of content. It also needs to be expandable and configurable to meet the special needs of each study. Another important thing is that it is expandable and not be limited to solve a specific problem. A Content Management System (CMS) is an application to collaboratively create, edit, and review various kinds of digital media. We analyzed the possibilities of storing study relevant content such as questionnaires, cognition tests, location tracks or physiological data in Content Management Systems. For this task we compared different open-source CMS which accomplishes this task best.

Because of its modular and extendable architecture the CMS "Drupal", which is also called a Content Management Framework [5], fits the requirement of handling different data and different studies.

Drupal accomplishes this task by abstraction. It provides systems and tools that allow assembling of different modules that can be combined to achieve a completely new functionality. The content managed by Drupal is stored in so called nodes. To achieve an abstraction the main content types are a variation of a "node"-Object. Nodes hold structured data (such as Title, Content, Author and Date). It is possible to create custom content types that extend a node by additional fields (e.g. the content type "personal data" could have the fields: name, gender and age).

A module extends Drupal features. It is possible to write custom modules by using the programming language PHP and interact with Drupal via an Application Programming Interface (API).

With this flexible architecture it is possible to use modules to build a server specific for a study that collects, processes, stores and visualizes relevant data. The modules can be shared with the community so that other researchers can use them in their studies. There are already lots of contributed modules developed and shared by the Drupal community. For example there are several modules available that can be used to visualize results of a study in a graph.



Figure 3. Modular Architecture of Drupal-Server.

Personal Computer

The study supervisor can monitor the study over a browser on his personal computer and view data collected in realtime. Additionally, he or she can download the collected data for further data fusion and analysis.

STUDIES

Although the system is still being developed, it has already been applied to a number of studies. The participants were all students of the Karlsruhe Institute of Technology (KIT) and the following studies have been carried out in the year 2009:

- Mobile acquisition of vital parameters and geographic position [6].
- A study investigating the mental state of 100 students during lectures.
- A study carried out in cooperation with the Department of Sociology where the student's campus life was investigated.
- An investigation into the cognitive ability of students during lectures.

CONCLUSION

In conclusion, within this article all components of a mobile online monitoring system have been described. The advantages of this system are that no more data allocation or conditioning is necessary because all the data is stored in a central point. Further studies will be carried out in order to generalize the system and to establish an integrated solution that can be used in a wide range of research studies. This research establishes a basis for further developments in the field of real-time psycho-physiological monitoring.

- 1. Carter, S. and J. Mankoff, "When participants do the capturing: the role of media in diary studies," *Proceedings of the SIGCHI conference on Human factors in computing systems* (2005), 908.
- 2. Ebner-Priemer, U.W. and T. Kubiak, "Psychological and Psychophysiological Ambulatory Monitoring," *European Journal of Psychological Assessment*, 23 (2007), 214-226.
- 3. Froehlich, J., M.Y. Chen, S. Consolvo, B. Harrison, and J.A. Landay, "MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones," *Proceedings of the 5th international conference on Mobile systems, applications and services*, (2007), 70.
- 4. Froehlich, J., T. Dillahunt, P. Klasnja, J. Mankoff, S. Consolvo, B. Harrison, and J.A. Landay, "UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits," *Proceedings of the 27th international conference on Human factors in computing systems*, (2009), 1043–1052.
- Rawtani, M.R. and S.S. Chidambaram, "Drupal: The Open Source Content Management System Software Suit For Library With Library 2.0 Features", 2009.
- Stumpp, J., U. Großmann, J. Ottenbacher, and S. Hey, "Mobile acquisition of vital parameters and geographic position," *Abstracts of the 1st Biennial Conference of the Society for Ambulatory Assessment*, Greifswald, Germany (2009), 59.

Experimental Design for Sternocleidomastoid Muscle Stress Measurement

CheeFai Tan¹ c.f.tan@tue.nl

Wei Chen w.chen@tue.nl Matthias Rauterberg g.w.m.rauterberg@tue.nl

Designed Intelligence Group, Department of Industrial Design, Eindhoven University of Technology Den Dolech 2, 5612 AZ Eindhoven, The Netherlands

ABSTRACT

In this paper we assess the sternocleidomastoid muscle stress with two experimental studies to evaluate the developed smart neck support system. The first study involving head rotation angle and sternocleidomastoid muscle (SCM) electromyography (EMG) activity. This (descriptive) study extends efforts from different authors to assess the isometric strength of SCM, in which the experiment is carried out in relation to time. This first study functions as a calibration experiment which provides the input for the second study as a validation experiment. The validation experiment was conducted inside the aircraft cabin simulator. The SCM of participant was evaluated with electromyography. The smart neck support system is to reduce the SCM muscle stress of aircraft passenger adaptively during air travel.

Author Keywords

Calibration experiment, validation experiment, smart neck support system.

ACM Classification Keywords

C.4. Computer Systems Organization: Performance of Systems: Measurement Technique.

INTRODUCTION

Aircraft passengers may experience some degree of physiological and psychological discomfort, as well as stress during air travel [1]. The aircraft passenger comfort depends on different features and the environment during air travel. Seat comfort is a subjective issue, because it is the customer who makes the final determination. Customer evaluations are based on their opinions having experienced the seat [2]. The aircraft passenger seat has an important role to play in fulfilling the passenger comfort expectations. The seat is one of the (most) important features of the vehicle and in which the passenger spends most of their time during air travel.

Several studies [3,4,5,6] conducted the experiment to measure the relationship between electromyography (EMG) activity of sternocleidomastoid (SCM) muscle in relation to head rotation. Ylinen et al. [3] evaluated the isometric rotation strength in both neutral and different pre-rotated positions of the neck. Maximal axial rotation strength of the neck muscles was measured in a neutral position and bilaterally at 30° and 60° rotation using the isometric neck strength measurement system. Isometric maximum voluntary contractions of the neck muscles in flexion and extension were tested. The neck strength measurement (system was) used to measure different rotation angles, isometric flexion and extension.

Moon et al. [4] studied EMG signal from neck muscles to estimate face direction angle. The EMG signal of the SCM muscle concerned in the head movements is measured. The clavicle region was selected as the reference point for the EMG signal measurement because the clavicle is the bone nearest to the neck and it has no muscles. During the experiment, when the head was rotated to the right or left, the EMG signal was measured from the SCM at the opposite side.

Bexander et al. [5] conducted an experiment to investigate the effect of the eye position on neck muscle activity during cervical rotation. In the study, the root-mean-square EMG amplitude was measured for one second during the period in which the position of the neck and head was held statically in each rotation angle (0° , 15° , 30° , 45°). Recordings were made from three dorsal neck muscles on the right using electrodes and from the left SCM muscle with surface electrodes.

Lin and Huang [6] investigated the changes of neck muscle activities when using different pillows in a time series and different kinds of pillows. Each participant was investigated for neck muscle activities in sitting and supine position for 30 minutes in three conditions: with a neck support pillow, a standard pillow, or without using a pillow in a random order.

¹Also Universiti Teknikal Malaysia Melaka, Melaka, Malaysia.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

In this paper, we describe two experimental setup to investigate the relationship between head rotation angle and EMG value. The output from the experiment is used for the validation of the developed smart neck support system. Sternocleidomastoid (SCM) muscle was selected for the experiment because it is related to the head rotation activity [7]. In order to objectify the EMG value of SCM muscle at a pre-defined head rotation angle, a self designed and self built apparatus was used to perform the experiment. The results would provide information about the relationship between head rotation angle and SCM muscle EMG value. In addition, an aircraft cabin simulator was built to conduct the validation experiment on the developed smart neck support system [8,9,10].

METHODS

Calibration Experiment

Subject

Four participants with no neck pain in the last three months were recruited in this experiment. The group consisted of two females and two males.

Experimental Setup

The calibration experiment was designed to evaluate the SCM EMG value in relation to different pre-defined head rotation angles, time and gender. They were informed regarding the experiment which involved sitting inside the special designed head rotation angle apparatus, wearing a special designed head set with laser beam, video recording and recording of electromyography on SCM muscle (right and left). The location of experiment was the simulation lab in the main building of Eindhoven University of Technology. For video recording, we used a closed circuit camera (CCTV) that was installed to record the activities of the participant throughout the experiment.

Apparatus and Data

The hardware used in the experiment is as follows:

- MP150 Biopac system with EMG module
- Head rotation angle measurement apparatus
- Headset with laser beam
- Laptop
- CCTV

The EMG data of right SCM and left SCM were measured with the head rotation. The SCM muscle was related to the major neck flexors [11]. The skin surface of SCM was cleaned with alcohol before the surface electrodes were applied [12]. A pair of surface electrodes (Ag/AgCL electrodes; EL504-10; 10 mm diameter on a 25 mm square backing; Biopac Systems, Inc., USA) was placed in parallel with the muscle fibers of SCM with 20 mm inter-electrodes distance. The electrodes were placed at lower 1/3 of the line connecting sternal notch and mastoid process [13]. Figure 1 shows the location of electrodes.



Figure 1. The location of electrodes on left SCM muscle.

The EMG signals were recorded at 1000 Hz sampling rate, band-pass filtered between 20 Hz and 350 Hz, full-wave rectified, and smoothed with a low-pass filter [14]. The high-pass cut-off frequency at 20 Hz was used to reduce motion artifacts and electrocardiography (ECG) artifacts with minimal impact on the total power of EMG [12]. Data were continuously recorded for 10 minutes with Biopac MP150 and analyzed with AcqKnowledge 3.9.1 (Biopac System Inc., USA).

Experiment Procedure

We started the experiment with 30 minutes of briefing of the participants and attachment of electrodes on SCM muscles. The participant performed maximal voluntary contraction (MVC) of the SCM by rotation of the head to the left hand side and right hand side for 10 seconds. After that, we positioned the participant in the head rotation angle measurement apparatus while wearing the headset. Figure 2 shows the participant inside the head rotation angle measurement apparatus and in LR0° head rotation.

Both right and left SCM muscles were evaluated in two conditions (with and without support) and 7 rotation angles (L45°, L30°, L15°, LR0°, R15°, R30°, and R45°). The participant was tested in each angle for 10 minutes. A rest period of two minutes was implemented between each angle to minimize the effect of fatigue [14]. After the measurement, all the electrodes were detached from the participant and the participant was debriefed.

Validation Experiment

Subject

Three participants (1 female and 2 male), which had no neck pain in the last three months, were recruited. They were informed regarding the procedure of experiment such as sitting inside the aircraft cabin simulator [16, 17,18] for



Figure 2. The participant sitting inside the head rotation angle apparatus.

one hour, video recording and attachment of EMG electrodes on their SCM muscle.

Experimental Setup

The validation experiment was conducted in the aircraft cabin simulator. The location of experiment was the simulation lab in the main building of Eindhoven University of Technology. The smart neck support system was installed at the economy class aircraft seat.

There were two CCTVs used to monitor the participant activity. One CCTV was located in front of each participant and another CCTV is located above the head of the participant. There was one additional CCTV that monitored the overall activity in the simulator.

Apparatus and Data

The hardware used in the experiment was as follows:

- MP150 Biopac system with EMG module
- Aircraft cabin simulator
- Smart neck support system
- Computer
- CCTV

Two EMG modules of MP150 Biopac system were used for each participant. The aircraft cabin simulator was designed and built to simulate the average economy class cabin. Three smart neck support systems were installed in each aircraft seat. The computer was used for data logging and video recording. The CCTVs were installed in front as well as above the participant. The acquisition of EMG signal and procedure were the same as in the calibration experiment.



Figure 3. The setup of participants in the aircraft cabin simulator.

Experiment Procedure

We started the experiment with 30 minutes of briefing to the participants and attachment of the electrodes on SCM muscles. The participant performed maximal voluntary contraction (MVC) of the SCM by rotating the head to the left hand side and the right hand side for 10 seconds. After that, we positioned the participant in the economy class aircraft seat. The aircraft seat sitting position was classified as aisle seat, center seat and window seat. Subsequently, the light in the aircraft cabin was dimmed and the participant was advised to rest during the one hour experiment. The EMG signals for participants were monitored and recorded in parallel with system logging and video recording. Figure 3 shows the view in the aircraft cabin simulator during the experiment.

CONCLUSION

In conclusion, this paper describes two experimental setups for the evaluation and validation of a smart neck support system. The calibration experiment is used to gather the SCM EMG value information in relation to different predefined head rotation angle, time and gender. A specially designed head rotation angle apparatus and a headset were modified for the calibration experiment. For the validation experiment, the participants sat inside the aircraft cabin simulator to validate the smart neck support system. Each participant was attached with EMG electrodes and the activities of the participant were observed with CCTV. The activity of each participant was recorded in real time. The research integrated different fields of study such as control, physiology and ergonomic for the development of the smart neck support system. Future studies are suggested to provide more insight into the experimental design process which could lead to the better objectification of the smart system.

REFERENCES

1. Kalogeropoulos, S. Sky rage. Flight Safety Australia (1998), 36-37.

- Runkle, V.A. Benchmarking seat comfort. SAE Technical Paper, No. 940217 (1994).
- 3. Ylinen, J., Nuorala, S., Häkkinen, K., Kautiainen, H. and Häkinnen, A. Axial neck rotation strength in neutral and prerotated postures. Clinical Biomechanics, *18* (2003), 467-472.
- Moon, I., Kim, K., Ryu, J. and Mun, M. Face directionbased human-computer interface using image observation and EMG signal for the disabled. *Proceeding of the 2003 IEEE Internationa Conference* on Robotics & Automation (2003), 1515-1520.
- 5. Bexander, C.S., Mellor, R. and Hodges, P.W. Effect of gaze direction on neck muscle activity during cervical rotation. *Exp Brain Res*, *167* (2005), 422-432.
- 6. Lin, Y.-H. and Huang, W.-Y. Cervical postures and electromyographic activities of related neck muscles when using a neck support pillow. *Journal of Biomechanics*, 40, 2 (2008), 414.
- 7. Sommerich, C.M., Joines, S.M.B., Hermans, V. and Moon, S.D. Use of surface electromyography to estimate neck muscle activity. *Journal of Electromyography and Kinesiology*, *10*, 6 (2000), 377-398.
- Tan, C.F., Chen, W., Liu, H. and Rauterberg, M. Adaptive framework and user preference modelling for economy class aircraft passenger seat. *Proceeding of* 2009 Third UKSim European Symposium on Computer Modeling and Simulation, Athens, Greece (2009), 66-69.
- Tan, C.F., Chen, W., Kimman, F. and Rauterberg, M. Sleeping in sitting posture analysis of economy class aircraft seat. *Electronic Engineering and Computing Technology, Series: Lecture Notes in Electrical Engineering*, Ao, Sio-long; Gelman, Len (Eds.), 1st Edition, Springer Publication, 60 (2010), 703-713.
- 10. Tan, C.F. Chen, W. and Rauterberg, M. Development of adaptive aircraft passenger seat system for comfort improvement. *Proceeding of International Conference*

for Technical Postgraduates, Kuala Lumpur, Malaysia, (2009), paper no. CC02-05.

- 11. Conley, M.S., Meyer, R.A., Bloomberg, J.J., Feeback, D.L. and Dudley, G.A. Noninvasive analysis of human neck function. *Spine*, *20*, 23 (1995), 2505-2512.
- Hermens, H.J., Feriks, B., Disselhorst-Klug, C. and Rau, G. Development of recommendations for SEMG sensors and sensor placement procedures. *Journal of Electromyography and Kinesiology*, 10, 5 (2000), 361-74.
- 13. Falla, D., Dall'Alba, P., Rainoldi, A., Merletti, R. and Jull, G. Location of innervation zones of sternocleidomastoid and scalene muscles – a basis for clinical and research electromyography applications. *Clinical Neurophysiology*, 113, 1 (2002), 57-63.
- 14. Cheng, C.H., Wang, J.L., Lin, J.J., Wang S.F. and Lin, K.H. Position accuracy and electromyographis responses during head reposition in young adults with chronic neck pain. *Journal of Electromyography and Kinesiology*, doi:10.1013/j.jelekin.2009.11.002, (2009).
- 15. Clancy, E.A., Morin, E.L. and Merletti, R. Sampling, noise reduction and amplitude estimation issues in surface electromyography. *Journal of Electromyography and Kinesiology*, *12*, 1 (2002), 1-16.
- 16. Tan, C.F., Chen, W. and Rauterberg, M. Interactive aircraft cabin testbed for stress-free air travel system experiment: an innovative concurrent design approach. *Proceeding of International Conference on Advances in Mechanical Engineering*, Shah Alam, Malaysia, (2009), 137.
- 17. Tan, C.F., Chen, W. and Rauterberg, M. Design of aircraft cabin testbed for stress free air travel experiment. *Proceeding of 5th International Conference on Planning and Design*, Tainan City, Taiwan, (2009), 157.
- 18. Tan, C.F., Chen, W. and Rauterberg, M. Total design of low cost aircraft cabin simulator. *Proceeding of Design* 2010, Cavtat, Croatia, (2010), 1721-1728.

Peripheral Arterial Tone As an Index of ANS Trade-Off

Stas Krupenia Thales Research and Technology, Delftechpark 24, 2628XH, Delft, The Netherlands stas.krupenia@d-cis.nl Tel: +31 15 2517 874 **Eldad Yechiam** yeldad@techunix.technion.ac.il Maya Arad maya.ohad@gmail.com

Israel Institute of Technology Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa 32000, Israel

ABSTRACT

The present study investigated the possibility of a tradeoff between generalized and sympathetic autonomic arousal. We examined participants' Heart Rate (HR) and Peripheral Arterial Tone (PAT) during completion of vigilance Go/No-Go task. PAT is considered to be a relatively pure measure of sympathetic arousal. The two measures tended to be consistent within each participant. However, as predicted by the notion of Individual Response Specificity, across individuals there was a negative correlation between the two measures. Additionally, each measure had a different correlation with the performance indices, suggesting that the sympathetic and parasympathetic branches of the Autonomic Nervous System (ANS) modulate different cognitive strategies. These results challenge the view that measures of the autonomic system are indicators of the same psychological construct.

Author Keywords

Peripheral arterial tone, heart rate, autonomic nervous system, cognitive load.

INTRODUCTION

A variety of ANS indices are used to asses mental load, including Heart Rate (HR), Pupil Dilation (PD), and Galvanic Skin Response (GSR). However, the ANS has two branches, sympathetic and parasympathetic, which affect physiological indexes differently. For example, HR is modulated by sympathetic and parasympathetic affecters [2] whereas GSR, PAT, and the Pre-Ejection Period (PEP) are mainly affected by sympathetic affecters [16]. The relationship between the different ANS measures and behavior is unclear [1] and the few studies that have examined this report mixed findings [7, 9].

Studies of Individual Response Specificity (IRS) suggest that individuals tend to have a typical autonomic response that is more elevated than other responses across diverse conditions [11]. IRS implies a negative correlation across performers between different measures because different people consistently have different reactivity patterns [1]. However, this research direction was not taken up and subsequent studies did not replicate the negative correlation [7]. Furthermore, [12] found a low positive correlation across participants between HR and PEP.

The idea of IRS also predicts a positive correlation between measures within individuals: consistent with the view of the two ANS branches being complementary affecters of "fight or flight" vs. "rest and digest" [2]. Yet as far as we know, the inter-individual correlation between measures has not been examined. Another under-studied line of investigation concerns the functional difference between the two systems. Studies focusing on PAT have shown that pure sympathetic channels tend to be more sensitive to increases in arousal than general measures. For example, PAT was found to be a better indicator of cognitive load than HR in memory [6] and complex perceptual motor tasks [5]. Additionally, sympathetic and parasympathetic branches are differently sensitive to psychological stimuli of different valence [10]. Similar findings were observed using PAT as an index of sympathetic activation and PD as a general index [4]. This is consistent with literature suggesting that parasympathetic activity is highly related to stress responses [3].

The main goal of the present study was to re-examine the correlation between general and sympathetic measures of arousal using the Peripheral Arterial Tone (PAT; a sympathetic measure that is not a direct component of heart activity) and HR (a general measure of ANS activation). In the current study we used a passive effort vigilance Go/No-Go task to enable simultaneous examination of the focus on positive and negative events (decision bias) and performance level (decision accuracy). The use of the vigilance task enabled the continuous monitoring of relationships between the arousal measures over a longer time periods with the participant's task being homogeneous throughout. Moreover, this task is particularly sensitive to mental load investment [18], allowing for the comparison

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

of the functional significance of general and sympathetic measures of arousal.

The results revealed that within each participant, HR and PAT correlated, whereas across participants the measures were negatively correlated. The results suggest that the two branches of the ANS modulate different arousal strategies that vary across participants.

METHODS

Participants

Forty undergraduate students from the Israel Institute of Technology (Technion; 17 females) volunteered to complete the experiment. Their average age was 23.6, ranging from 18 to 28. Reimbursement ranged from 30 to 90 New Israeli Shekels (NIS) depending on task performance.

Go/No-Go Task

Participants completed a single block computerized Go/No-Go task that involved the presentation of 100 two-digit Go (98 of 100 trials) and No-Go cues (2/100) randomly presented between trials 20 and 100. The meaning of each cue was told in advance and this information was continuously presented using a cue card placed next to the computer. Participants were rewarded NIS 10 for hits (correct response to Go cues), penalized NIS 20 for misses (failure to respond to Go cues), and penalized NIS 5 for false alarms (incorrect response to No-Go cues). The payment asymmetries were designed to highlight the importance of the rare stimuli.

Stimuli were presented for 1.5 seconds with a response window of two seconds (SOA = 19 s). For 17.5 s following the removal of the stimulus, the computer screen was blank. Participants were tested individually using a PC (Intel ® Core 2 CPU 6400 @ 2.13 GHz, 1.98 GB of RAM). The numbers representing Go and No-Go cues ("21" and "32", or "22" and "31") were counterbalanced between participants. Physiological measures. PAT and HR were measured throughout the experiment using a Site-Pat200 device (Itamar Medical, Ltd, Israel). This system is a finger cup shaped photo-cell sensor plethysmograph that is placed at the end of the first finger of the non-dominant hand. The cup envelops the finger up to and beyond its tip with uniform pressure so that only arterial volume is measured and the aggregation of the venous blood is prevented (Lavie, Schnall, Sheffy, & Shlitner, 2000). The device includes a light-emitting diode and a photo-sensitive cell that increases in sensitivity by unloading the arterial wall tension and allowing arterial volume to vary with each pressure wave. The arterial tone affects the translucence of the light through the finger tissue, which, in turn, affects the output measures. The amount of light transmitted through the finger is measured in Volts (V). When the artery is constricted (sympathetic activation) the blood is more concentrated and less light is transmitted through the finger. Thus, sympathetic activation (arterial constriction) leads to

lower PAT. The system records PAT for each heart beat and thus HR is also measured. Because higher amounts of light represents lower levels of arousal, we used the level of vasoconstriction (-V) as a positive index of mental effort according to the level of PAT. This index is referred to as PAT Vasoconstriction, or PATV.

Procedure

Participants were given general information about the experiment's approximate length and possible payment. They were seated in front of the experiment computer, connected to the PAT system, and provided with a written information and consent form. After signing the form, the instructions were presented verbally and written. Participants were instructed to remain attentive to a series of numbers to be flashed on the screen in front of them. They were told that two types of numbers existed, "friends" and "foes" and that they must respond only to foes by clicking the left mouse key while ignoring the friends. Participants were then told which numbers denoted friends and foes and were informed of the points gained or lost following a response to each cue type. Participants were also given a card repeating this information.

Participants were provided with an initial sum of NIS 70 and their updated gains or losses were continuously displayed on the screen. Participants were given ten practice trials at the beginning of the experiment to become familiar with the task. The total duration of the experiment was approximately 50 minutes. At the end of the experiment the participants were paid according to the total amount of points earned.

RESULTS

Behavioral Data

The amount of correct and incorrect responses (and SE) from each type of stimuli was as follows; hits= 1.4 ± 0.42 , misses= 0.6 ± 0.42 , false alarms= 1.12 ± 0.14 , and correct rejections= 96.88 ± 0.14 . Misses were more frequent then false alarms. Task performance was not at floor or ceiling levels and there was room for error.

Physiological Data

PATV and HR values were calculated using a 10 second window from 5 s before to 5 s after the stimulus averaged across participants. Repeated measures analysis shows a significant decrease in HR (F(1,85) = 9.41, p < 0.001) and a non-significant increase in PATV across successive blocks (F(1,78)=1.98, p=0.15). Additionally, there was a significant decrease in PATV from pre-task to block 1 (t(39)=-2.44, p < 0.05). PATV was also significantly higher pre- compared to post-stimuli (t(39=4.91, p < 0.001). These findings thus suggest discrepancies between the changes of the two measures during the task. The differences were further examined using a correlation analysis.

Correlations Between Physiological Measures

First, the average Pearson correlation between PATV and HR within each participant across the 100 trials was examined. The results showed that the average withinparticipant correlation between the two measures was 0.18 pre- and 0.17 post-stimulus presentation. In both cases the correlations were significantly above zero (t(39)=4.29), p < 0.01 and t(39) = 4.08, p < 0.01). For correct rejections the average correlation was approximately the same as the overall correlation between PAT and HR. Due to the low number of hits, misses, and false alarms, the within participant correlation between measures could not be calculated for these events. For 75 % of participants there was a positive correlation between PATV and HR. Thus, it appears that within individuals the two measures correlated. Second, the association between PATV and HR across individuals was examined. The pre-task correlation taken 5 s before the start of the experiment was not significant (r= -0.004, p=0.49). During the task, however, there was a negative and approaching significance pre- (r=-0.27). p=.09), and post- (r=-0.26, p=.10) stimulus correlation that was not driven by extreme values. Finally, there were no differences between genders in the correlation between the two measures (-0.32 for females and -0.23 for males).

Correlations with Performance

Because the number of hits had only three levels, a Spearman Correlation was used for this analysis (Table 1). There were conflicting correlations between the HR, PATV, and performance. There was a negative pre- and post-cue correlation between PATV and misses. The lack of correlation between HR and performance suggests dissociation between the measures at the functional level.

DISCUSSION AND SUMMARY

The results showed that within each participant there was a positive correlation between PATV and HR. However, across participants there was a negative correlation between the two measures. The results thus demonstrate the emergence of Individual Response Specificity. This can be interpreted as due to differences in people's arousal system with different individuals scoring higher on a particular index, or to different individuals using different arousal strategies.

A result that is more consistent with the second explanation is the change in these measures across time. The

| | PATV | | HR | |
|--------------|--------|--------|--------|-------|
| | Before | After | Before | After |
| Misses | -0.23+ | -0.25* | 0.17 | 0.17 |
| False Alarms | -0.06 | -0.06 | 0.14 | 0.13 |

Table 1: Spearman Correlation between PATV and HR from 5 seconds pre- and post-stimuli type with measures of performance; $^{+} = p < 0.1$; $^{*} = p < 0.05$.

generalized measure (HR) stabilized faster than the sympathetic measure (PATV) throughout the task. This suggests that the processes tapped by these two measures may also have different dynamics. Examining the functional significance of the two measures supports their distinction because they correlated differently with performance. One possibility for the low association between HR and performance is that the task involved too little active effort with only two No-Go trials. Another possibility is that the task was too long to maintain high arousal.

Previous studies have shown that PAT, an autonomic measure of sympathetic activation, is a better predictor of performance in cognitive tasks than HR [5]. This was argued to be due to the greater sensitivity of the PAT to fast paced changes in cognitive load [6]. The current findings suggest a complementary explanation, showing that in fact PAT and HR are not measures of the same construct. Rather, within participants, the two measures are positively correlated but only to a limited extent (about 0.2). Moreover, across performers the two measures are negatively associated. This negative association is also accompanied by different functional significance of the two measures, which suggests that the two measures represent different cognitive/physiological arousal strategies.

The current findings are somewhat consistent with the few previous studies examining correlations between general and sympathetic arousal. Consistent with the negative association between measures *across participants*, [9] found that during an auditory odd-ball and mental arithmetic task, HR reactivity was associated with minimal GSR while HR non-reactive subjects showed more skin conductance. For example, [7] reported a low positive correlation within participants however this was not statistically significant.

However, our results contrast [12] who reported a positive correlation between PEP and HR *across* participants during a delayed matching task. A possible explanation for the contradiction between results of the current study and those of [12] is that they did not examine the outcome of a particular strategy but studied the gross effect of effort on a simple task where effort improves performance. It may be that in this situation the two arousal subsystems work complimentarily. Still, this does not contradict the idea that the parasympathetic and sympathetic systems are associated with different cognitive strategies.

An explanation for the dissociation and even tradeoff between these measures must consider energetic differences between the measures. What PATV represents is a change in the allocation of bodily resources, mediated by the flow of blood, from the periphery to the center, thus enabling a re-organization of resources more effectively according to the task. In contrast, HR increases represent an increase in the bodily resources in general. It appears that individuals can exhibit both reactions, a re-organization and a general increase, simultaneously. However, some individuals responded to a novel task by increasing their resources, whereas others respond by re-organizing them. In other words general arousal (HR) is a gross performance enhancing mechanism. By increasing HR, more oxygenated blood is available to the brain than at lower HR values. In contrast, sympathetic arousal is a more strategic performance enhancing mechanism. Similar to general arousal increases, by constricting peripheral arteries, more oxygenated blood is available to the brain. However general arousal increases oxygenated blood flow throughout the body, whereas sympathetic arousal increases oxygenated blood flow to specific parts of the body (e.g. the brain). Sympathetic arousal thus appears a more efficient arousal mechanism for increasing task performance over the long term. Evidence for this claim is that participants in the current studies tended to exhibit an increase in either HR or PAT. For the task used, general arousal increases may be sufficient for maintaining performance at a target rate for the duration of the task. However, the gradual decrease in HR across trials suggests that maintaining higher levels of general arousal is challenging. Contrasting this, sympathetic arousal appeared to maintain steady (or increase slightly) during task completion. Thus for the tasks examined, a more efficient strategy appears to increase sympathetic activity rather than general arousal. For short duration tasks, either general or sympathetic activation is sufficient for increasing oxygenated blood flow. For long duration tasks, however, ANS activation is less efficient and more difficult to sustain than sympathetic arousal.

Despite the internal and external consistency of the results reported, there are several limitations. The lack of correlation between physiological measures and performance may be because the task involved few opportunities for variability; only two No-Go stimuli were presented. Future studies would benefit from more data points; however, [12] reported a correlation between measures of arousal and performance on early but not later trials of an attention task suggesting that arousalperformance correlations may be found with few data points. Given the argument presented above relating to the use of superior coping strategies by some participants as defined by differences in the physiological recordings suggests that the PAT may be used as a biofeedback training tool to improve performance. Recent research by [13] showed that ANS components can be controlled by directed training involving both internal cognitive and physiological monitoring. This is a recommended area of future research.

ACKNOWLEDGEMENTS

This research was supported in part by United States Airforce Grant FWR20060084H and in part by the European Commission Integrated Project IP-SKILLS-35005.

- 1. Cacioppo, J.T., and Tassinary, L.G. Inferring psychological significance from physiological signals. *American Psychologist*, 24, 1 (1990), 16-28.
- Cacioppo, J.T., Tassinary, L.G., and Berntson, G.G. Handbook of Psychophysiology (2nd ed.). Cambridge: Cambridge University Press, 2000.
- 3. Heponiemi, T., Ravaja, N., Elovainio, M., Näätänen, P., and Keltikangas-Järvinen, L. Experiencing positive affect and negative affect during stress: Relationships to cardiac reactivity and to facial expressions. *Scandinavian Journal of Psychology*, 47 (2006), 327-337.
- 4. Hochman, G., Glockner, A., & Yechiam E. (in press). Physiological measures in identifying decision strategies. In A. Glockner and C.Witteman (Eds.), *Foundations for Tracing intuitions, challenges, and methods*. London: Psychology Press & Routledge.
- 5. Iani, C., Gopher, D., Grunwald, A.J., and Lavie, P. Peripheral arterial tone as an on-line measure of load in a simulated flight task. *Ergonomics*, 7 (2007), 1026-1035.
- 6. Iani, C., Gopher, D., and Lavie, P. Effects of task difficulty and invested mental effort on peripheral vasoconstriction. *Psychophysiology*, *4* (2004), 789-798.
- Kettunen, J., Ravaja, N., Näätänen, P., Keskivaara., and Keltikangas-Järvinen, L. The synchronization of electrodermal activity and heart rate and its relationship to energetic arousal: A time series approach. *Biological Psychology*, 48 (1998), 209-225.
- 8. Lavie, P., Schnall, R.P., Sheffy, J., and Shlitner, A. Peripheral vasoconstriction during REM sleep detected by a new plethysmographic method. *Nature Medicine*, 6 (2000), 606.
- 9. Lawler, E.E. *Strategic Pay*, Jossey-Bass, San Francisco, CA, 1990.
- 10. Löw, A., Lang, P.J., Smith, J.C., and Bradley, M.M. Both predator and prey: Emotional arousal in threat and reward. *Psychological Science*, *19* (2008), 865-873.
- Marwitz, M., and Stemmler, G. On the status of individual response specificity. *Psychophysiology*, 35 (1998), 1-15.
- Richter, M., and Gendolla, G.H.E. The heart contracts to reward: Monetary incentives and pre-ejection period. *Psychophysiology*, 46 (2009), 451-457.
- 13. Tang, Y., *et al.* Central and autonomic nervous system interaction is altered by short-term meditation. *Proc. Nat. Acad. Sci. USA, 106,* 22 (2009), 8865-8870.
- 14. Wallin, B.G. Sympathetic nerve activity underlying electrodermal and cardiovascular reactions in man. *Psychophysiology*, *18* (1981), 470–476.

Using EEG Recordings to Examine the Relationships Between Sustained Attention and Types of Background Music in Individuals with ADHD

Chelsea Chew

Nanyang Technological University, National Institute of Education 1 Nanyang Walk, Singapore chelsea.chew@nie.edu.sg

ABSTRACT

The purpose of this exploratory study was to use Electroencephalogram (EEG) recordings to examine the relationships between the maintenance of sustained attention across various academic tasks and the presentation of different background music amongst individuals diagnosed with Attention Deficit (Hyperactivity) Disorder (ADHD). Specifically, the present research sought to explore the differences in sustained attention and ease of paying attention and enjoyability across Reading, Reading Comprehension, Reading Comprehension with recall (with working memory load), Numerical Operations, and the presentation of silence, classical music, pop music, and sounds from television programs. Differences in sustained attention were measured using EEG recordings while the participants rated the ease of paying attention and enjoyability across the different sound conditions on the different tasks. Twelve individuals with ADHD participated in this study. Individuals with ADHD were found to be better able to sustain attention and enjoy tasks with pop music in the background in tasks that did not have a working memory load.

Author Keywords

EEG, ADHD, music, sounds.

INTRODUCTION

Attention-Deficit/Hyperactivity Disorder (ADHD) is an enduring mental disorder that characterized by an inability to pay attention and/or hyperactivity (American Psychiatric Association, 2000). The prevalence rates for this disorder range from 7 to 12 percent (Barkley, 2009). Without effective treatment, children with ADHD are at greater risk of developing poor academic, behavioural, social, and mental health outcomes (Antshel, Macias, & Barkley, 2009). It is thus imperative that efforts be directed towards helping students with ADHD achieve academically. Unfortunately, ADHD students typically find it very challenging to focus and filter distractions (both internal and external) long enough to be successful academically.

It is conventional wisdom that silence promotes engagement in academic tasks. However, findings from a few studies suggest that the brains of individuals with ADHD require stimulation in the form of music (Zentall, 2006), to function more effectively on academic tasks. According to Barkley (2009), students with ADHD have a greater than average need for stimulation. Thus, students with ADHD attempt to increase their levels of stimulation, consciously or not, in various settings and tasks, and this need for stimulation extends to academic settings (Zentall, 2005). This belief, nonetheless, has only been examined anecdotally.

With the advent of Electroencephalogram (EEG) devices, a tool that measures electrical activity originating in an area of the brain called the cerebral cortex, anecdotal information regarding the effect of music on engagement in academic tasks can be verified objectively.

ADHD is really a neurological disorder and the symptoms of ADHD reflect an underlying dysfunction of the brain which is related typically to increases in slow waves in the frontal and central cortical and subcortical regions (Mann, Lubar, et al., 1991). Electrical detectors, which, when placed on the scalp, are able to detect small changes in electrical activity allow the relationships between music and task engagement to be examined.

METHOD

Participants

I invited children and youth at a clinic to participate. Twenty-two potential participants who ranged in age from 9 years to 15 years agreed to take part with their parents' permission. All the potential participants' parents rated them on the Conner's Parents Rating Scale-Revised: Long Form (Conners, 1997). The CPRS consists of 80 items which reflect the construct of ADHD as stated in the Diagnostic and Statistical Manual of Mental Disorders

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

(DSM-IV-TR, American Psychiatric Association, 2000). To be eligible for participation, the potential participants had to obtain at least a T-score of 65 on the ADHD Index. All participants were assessed by an educational psychologist. None of the participants were taking psychostimulant medication for ADHD at the time the experiment was conducted.

All the potential participants were administered the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV, Wechsler, 2003). Potential participants who scored below 85 (M=100, SD=15) on any of the four subscales of the WISC-IV were excluded in order to exclude children with possible cognitive problems.

Finally, the researcher hooked the potential participants up to an EEG machine and assessed their brain waves for 10 mins. The researcher chose participants who produced slow frontal brain waves at the middle of their heads (Cz).

The ADHD, IQ assessment, and slow wave requirement plus the time commitment to participate in the study reduced the sample from 22 to 12.

Setting

I tested the participants individually in a bare room. I placed the relevant tasks on the desk. I also placed a computer and EEG equipment on the desk.

Procedure

I asked the participants to engage in one of the following tasks: reading, reading comprehension, reading comprehension with recall (with working memory load), and numerical operations.

I modified the questions for use for these tasks from test questions in the Wechsler Individual Achievement Test-Three (WIAT-III), a test none of the participants had taken. The reading comprehension with recall (with working memory load task) was the same as the reading comprehension task except that I did not allow the participants to refer to the passage they had just read and expected them to give the answers by trying to recall the information from the passage.

The year levels of the tasks chosen were 2 years below the participants' chronological ages (if there were enough questions to permit that) so that presumably the tasks were easy enough for the participants. I instructed each participant to engage in these 4 tasks and the tasks were administered in a counter-balanced fashion to the extent possible for 3 min each to the participants. I gave the participants a 2-min break after the 3-min engagement in each task regardless of the number of questions they had answered.

While the participants were completing the tasks, I administered one of 4 sound conditions: silence, classical music, pop music, or sounds from television programs. I

played these sounds in the background without the use of headphones.

While the participants were engaged in the tasks with the various sound conditions, I measured their brainwaves at the middle of their heads (Cz) with EEG equipment. Immediately at the end of each task, I asked each participant to rate, on a 10-point scale, how easy it was for them to pay attention while attempting each task under each sound condition, and how enjoyable it was for them to perform each of the tasks under each sound condition.

RESULTS

Test results were analysed using SPSS version 17.0 for Windows. The MANOVA method was employed to compare the means in slow wave activity at the middle of the head, and the responses to the two questions (ease of paying attention and enjoyability) across the different sound conditions. On the reading and reading comprehension tasks, participants were able to lower their slow wave activity in the pop music condition. A significant number of the 12 participants rated the popular music condition as easy to pay attention to and enjoyable.

DISCUSSION

The results show that the pop music condition was rated most highly by the participants for engagement in tasks. It is possible that the pop music condition proved to be more effective as the participants are mostly pre-teens or teens. This study, being exploratory in nature, has a rather small sample size. To be more certain about the findings, the study will have to be replicated with a larger sample which would include adults with a variety of tastes in music.

It has also been claimed that binaural beats (produced by playing 2 slightly different tones into each ear) can help to increase focus (http://www.i-dose.us/dose/en/21/Focus+2/0/). Using EEG technology, this claim can be explored and verified.

CONCLUSION

Understanding the conditions that promote task engagement in ADHD students is highly important as ADHD students, much more than other students, require a high level of stimulation and support. To a parent or a teacher, the ways in which these students seek out stimulation can certainly be very confusing or frustrating. If parents or teachers are able to understand the ways in which students with ADHD seek appropriate stimulation while trying to complete their academic tasks, they would be better able to foster these optimal conditions and thereby help to promote better academic outcomes on the part of the ADHD students.

REFERENCES

1. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (4th ed.) Text Revision. Washington, DC, USA, 2000.

- Antshel, K., Macias, M., & Barkley, R. A. The child with attention deficit hyperactivity disorder. In R.B. David, J. B. Bodensteiner, D. E. Mandelbaum, & B. J. Olson (Eds.), *Child Pediatric Neurology* (pp. 525-540). New York: Demos Medical Publishing, 2009.
- Barkley, R. A. Attention deficit hyperactivity disorder. In Ingram, R. E. (Ed.), *The International Encyclopedia* of Depression (pp. 40-43). Springer, NY, USA, 2009.
- 4. Binaural Beats.

http://www.i-dose.us/dose/en/21/Focus+2/0/

- Conners, C. K. Conners' Rating Scales revised technical manual. Multi-Health Systems, NY, USA, 1997.
- Mann, C., Lubar, J., Zimmerman, A., Miller, C., & Muenchen, R. Quantitative analysis of EEG in boys with attention-deficit-hyperactivity disorder: Controlled study with clinical implications. *Pediatric Neurology*, 8 (1991), 30-36.
- 7. Wechsler, D. *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: Psychological, 2003.
- 8. Zentall, S. S. *ADHD and education: Foundations, characteristics, methods, and collaboration.* Pearson Merrill Prentice Hall, Upper Saddle River, N.J., USA, 2006.
- 9. Zentall, S. S. Theory-and evidence-based strategies for children with attentional problems. *Psychology in the Schools*, 42 (2005), 821-836.

Psychophysiological Data Collection in an Organizational Setting: Studying Interaction Between the Manager and Subordinate During Performance Review Discussion

Mikko Salminen mikko.salminen@hse.fi Pentti Henttonen pentti.henttonen@hse.fi

Aalto University School of Economics, Center for Knowledge and Innovation Research (CKIR), PL 21255, 00076 Aalto, Finland

Niklas Ravaja

Aalto University School of Economics, CKIR Center for Knowledge and Innovation Research (CKIR). PL 21255, 00076 Aalto, Finland niklas.ravaja@hse.fi

Mikael Saarinen

Sensitiva Inc., Sensitiva Oy. Perämiehenkatu 5 D 65 00150 Helsinki, Finland ms@sensitiva.fi

ABSTRACT

In this paper we present a method for collecting synchronized psychophysiological signals and video stream in a real work-life setting, during performance review discussion between the manager and subordinate. A split screen video showing the both discussants is prepared, including the visualized psychophysiological signals. The video is used in coaching sessions organized to the managers. The goal of the coaching sessions is to train the managers to better perceive their own and the discussing partner's emotional state and thus improve their interaction skills. The described method can be used also in the study of emotions during dyadic interaction in other contexts.

Author Keywords

Psychophysiology, emotions, social interaction, management, visualization.

INTRODUCTION

Hogan and Kaiser [5] argue that leadership is perhaps the single most important issue in the human sciences because good or bad leadership can have huge effects in the performance and well-being of organizations. Leadership has been studied from various different aspects, emotional intelligence being one of them [see, for example 8]. Various studies have shown that skills and attributes that are related

to emotional intelligence have an important role in effective leadership [e.g. 4, 10]. We have studied the role of emotional intelligence in effective leadership in a 2-year project, and as a part of the project we have created a method to provide the managers feedback on their interaction with the subordinates during a performance review discussion. The psychophysiological activity of the both discussing partners is visualized in real time in the video of the discussion. The idea behind this is to make visible the emotional responses of the discussants so that by reviewing the video material the manager could increase his or her understanding of his or her own emotional reactions and also those of the discussing partner. We hypothesize that this leads to increased emotional and interaction skills.

Psychophysiology and Emotions

In psychological research emotions can be seen as biologically based action dispositions that have an important role in the determination of human behavior [e.g., 6]. Emotions can be considered to be constituted of three separate but related reactive systems: (a) expressive and evaluative language, (b) physiological response, and (c) behavioral response [6]. According to the dimensional theory of emotion, all emotions can be placed in a twodimensional space, defined by valence and arousal [e.g., 6]. The valence dimension varies from unpleasant to pleasant and the arousal dimension defines the level of bodily activation related to the current emotional experience and ranges from calm to excited state.

The psychophysiological method uses physiological signals, such as facial electromyography (EMG), electrodermal activity (EDA), and electrocardiography (ECG) for investigating different psychological processes

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

[3]. By measuring bodily activities, inferences about emotions, attention, and motivation can be made. Facial EMG has often been used to study the hedonic valence [e.g., 2]. Increases in the activation of the cheek (zygomaticus major) muscle area have been associated with positive emotions, whereas increases in the activation of the brow (corrugator supercilii) muscle region have been associated with negative emotions [11]. Periocular (orbicularis oculi) muscle area activity appears to be particularly high during positive-valenced high-arousal emotions [see for example, 9]. For the measurement of arousal, electrodermal activity (EDA) is an important index. EDA has been shown to correlate with self-reported emotional arousal in studies where affective pictures have been used as stimuli [e.g., 7].

METHODS

Subjects and the Setting

We have until preparing this paper recorded 112 discussions from 40 managers. The subjects were from five different organizations (two public, three commercial) and the mean age of the managers was 44 years and for subordinates 45 years. The discussions were held at conference rooms of each participating company and the structure of the discussions was decided freely by the manager and the subordinates. The topic, however, was fixed to be a performance review discussion. In some organizations this type of discussions were held annually, whereas in other (more sale oriented) organizations it was almost a monthly custom.

After the placement of the electrodes a 5 minute baseline was recorded before the actual discussion could start. The discussions lasted between 40 minutes to 2 hours, with an average of about one hour and each manager had a discussion with two subordinates, with at least 30 min break in between the sessions. The discussing partners were seated by a table and a stand for two web-cameras was placed in between them for the recording of facial gestures.

Psychophysiological Recordings

The physiological signals were recorded from participants with two Varioport-B portable recording systems (Becker Meditec, Karlsruhe, Germany). Facial EMG activity was recorded from the left corrugator supercilii, zygomaticus major, and orbicularis oculi muscle regions, using surface Ag/AgCl electrodes with a contact area of 4 mm diameter (Becker Meditec, Karlsruhe, Germany). Electrodes were filled with Synapse conductive electrode cream (Med-Tek/Synapse, Arcadia, CA). The raw EMG signal was sampled at 1024 Hz, amplified, and frequencies below 57 Hz and above 390 Hz were filtered out, using the analog filter built in the Varioport device. Electrodermal activity (EDA) was recorded with Varioport 16-bit digital skin conductance an amplifier (input range = $0-70 \text{ }\mu\text{S}$) that applied a constant 0.5 V across Ag/AgCl electrodes with a contact area of 4 mm diameter (Becker Meditec), sampling at 32 Hz. Electrodes were filled with TD-246 skin

conductance electrode paste (Med Assoc. Inc.) and attached to the middle phalanges of the ring and little fingers of the subject's left hand after hands were washed with soap and water. Electrocardiogram (ECG) was recorded with three electrodes, one placed on the neck over the vertebra; second placed on the left side of the ribcage over the second lowest rib; and the third placed over the uppermost part of the center line of the rib cage.

Video Recordings

Conversations were recorded from a distance with two standard DV-cameras attached on tripods and separately with two web-cameras (Logitech Quickcam 9000 Pro) attached to a stand that was placed between the discussing partners. A custom Python program utilizing OpenCV library was used to record the signal from the web-cameras to a standard Windows XP laptop. Resulting videos were converted (720x576 / 640x480, 25 fps, 5Mbps / 192Kbps / 48Hz VBR) with a H.264 codec for final use. Audio was recorded from the internal microphones of the web-cameras and also from the DV-camera using a "shotgun"-pattern microphone attached to the camera directed to manager.

Synchronization Issues

Synchronization between the audiovisual recording system and the Varioport data-acquisition devices was handled using a custom Python program communicating via Bluetooth with both Varioports' marker channels, sending a digital marker when a video recording was started and stopped. Synchronization was also assessed by manually inserting markers before and after the discussion into the two Varioport devices using a digital Y-cable with the button presses clearly visible on the video recordings. This method allows the comparing of the psychophysical data samples to the video/audio frames and scaling the latter to match the internal clock of the recording devices. Our recording/synchronization program allows for an arbitrary rate of sent digital markers to Varioport devices, thus allowing a very fine-grained synchronization for the recordings.

Analyses of the Psychophysiological Signals

Data were analyzed using Matlab (v.7.7.0) software with Anslab Professional (v.2.4) toolbox. ECG signals was analyzed using 400 Hz sampling rate and a band of 0.5-40Hz. R-peaks were detected algorithmically. Artefactual values (+/- 3 standard deviations from physiologically realistic mean calculated with values in the range of 400-1300ms) were removed and missing values were interpolated. Heart rate signal (bpm) was obtained and smoothed. EMG signals were filtered (high pass: 28Hz, low pass: 15Hz and notch: 50Hz .The signal was rectified and smoothed using a 50ms moving average. The processed signals were saved with 4Hz sampling rate (10 Hz for EMG). EDA signal was decomposed to phasic driver and tonic components using Ledalab (V3.2.3) Matlab program [1] and saved with a sampling rate of 10 Hz.

Visualization of the Psychophysiological Data on the Video

Videos of the participants were edited and psychophysiological data were combined with them using Avisynth (V2.5) frameserving scripting program (Figure 1) Video tracks were arranged as a split screen in the top half of the resulting video. Psychophysiological visualization signals were derived with a sampling rate of 1Hz. Heartrate (bpm) and electrodermal activity were visualized as continuous graphs at the bottom half of the screen. EMG and EDA were visualized using a "smileyface" emoticon superposed on the corresponding videos. Orbicularis oculi and Corrugator supercilii EMG activity was represented as different degrees of smile and frown. EDA was represented as different degrees of blushing and sweatdrops appearing next to the face. The degrees of activity were determined using data clustering methods in Matlab. The resulting video was encoded to various formats (H.264 and WMV) using Adobe Premiere Pro CS4.

CONCLUSIONS AND FUTURE DIRECTIONS

The method we describe here can be also used in other contexts, where it is important to collect continuous, synchronized, psychophysiological signals and video from two interacting partners. It is also possible to increase the number of the observed persons. Our current set-up enables mobile in vivo data collecting, easy adding of more cameras and psychophysiological recording devices, accurate wireless synchronization between recording devices and an automated, reproducible way of editing and compiling video material. In future recordings we plan to use Noldus FaceReader or equivalent application to automatically detect emotional reactions from the face. We will thus test if the automatic detection of facial expression could be used instead of facial EMG electrodes in detecting the emotional valence. The facial EMG electrodes are most obtrusive for



Figure 1. Picture showing the final visualization of data, provided to the managers to be used during the coaching sessions. On the left side is manager's data and on the right side the subordinate's data. The emoticons visualize facial EMG and EDA activity. The graphs below show heart rate and EDA activity, respectively, for the whole discussion. the subjects of all the used devices and a compensatory method would be needed for a more ecologically valid setting.

ACKNOWLEDGEMENTS

This work has been supported by the Finnish Work Environment Fund.

- 1. Benedek, M. & Kaernbach, C. (in press). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*.
- 2. Cacioppo, J.T., Petty, R. E., Losch, M. E., & Kim, H.S. Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, *50*, 2 (1986), 260-268.
- Cacioppo, J.T., Tassinary, L.G. & Berntson, G.G. Psychophysiological science. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *The Handbook of Psychophysiology* (2nd ed.). New York: Cambridge University Press, (2000), 3-26.
- 4. Carmeli, A. The relationship between emotional intelligence and work attitudes, behavior and outcomes. An examination among senior managers. *Journal of Managerial Psychology*, *18*, 8 (2003), 788–813.
- 5. Hogan, R. & Kaiser, R. B. What we know about leadership. *Review of General Psychology*, 9, 2 (2005), 169–180.
- 6. Lang, P.J. The emotion probe Studies of motivation and attention. *American Psychologist*, 50, 5 (1995). 372–385.
- Lang, P.J., Greenwald, M. K., Bradley, M.M., & Hamm, A. O. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30, 3, (1993), 261–273.
- Mayer, J.D. & Salovey, P. Emotional intelligence. *Imagination, Cognition and Personality*, 9, 3 (1989–90), 185–211.
- 9. Ravaja, N. Contributions of Psychophysiology to Media Research: Review and Recommendations. *Media Psychology*, 6, 2 (2004), 193-235.
- 10. Spencer, L.M. The economic value of emotional intelligence competencies and EIC-based HR-programs. In: Cherniss, C. & D. Goleman (eds.) *The emotionally intelligent workplace: how to select for, measure and improve emotional intelligence in individuals, groups and organizations*. San Francisco: Jossey-Bass/Wiley (2001), 45–82.
- 11. Witvliet, C.V.O., & Vrana, S.R. Psychophysiological responses as indices of affective dimensions. *Psychophysiology*, *32*, 5 (1995), 436–443.

Extracellular Multi Unit Recording in Fear Conditioning in Mice Using a Telemetry Approach in an Automated Home Cage (DualCage) Environment

René F. Jansen¹

Anton W. Pieneman¹

rene.jansen@falw.vu.nl

¹Behavioral Neuroscience Group, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, The Netherlands

Manfred Gahr²

ABSTRACT

Fear conditioning is an important test in behavioral neuroscience to investigate the neural systems and molecular basis of various aspects of emotional learning across a wide range of species. Dysfunction of the fear circuits is assumed to underlie mechanisms of affective disorders and is frequently investigated in rodent models. While generally the scientific focus has been on the assessment of behavioral readouts such as freezing, we extended to autonomic measures [1] and now include electrophysiological measures in freely moving mice. We developed a radio-telemetry approach to perform extracellular recordings of neurons in freely moving birds that has now been adapted to mice. Experiments were performed in our novel fully automated DualCage (see Abstract by Stiedl et al., presented at this conference) environment to explore the dynamics neural responses in fear conditioning in male C57BL/6J mice without human interference. We used auditory trace fear conditioning to explore phasic responses to an explicit (tone) cue. Auditory trace fear conditioning occurred with five tone-shock pairings with tone and shock separated by 15-s trace intervals to render this behavior tests hippocampusdependent [2]. Electrical extracellular activity was recorded using a modified version of the transmitter used in zebra finches [3]. This transmitter uses a single channel to record extracellular electrical activity at frequencies between 100 and 15000 Hz. The transmitter weighs ~1.1 gram (batteries included) and this low weight ensures that it interferes minimally with the normal activities of the animal. The

Andries Ter Maat²

²Department of Behavioural Neurobiology, Max Planck Institute for Ornithology, Seewiesen, Germany

Oliver Stiedl¹

main modification is that the transmitter now runs for 5 days on a single battery as opposed to 24 hr for the original design. This extended battery life allows us to run extended training/retention test protocols without the need to handle the animal for battery changes. As in the original design, electrodes are glued into place using dental cement and can, therefore, not be moved. This procedure provides for the necessary long-term mechanical stability, albeit at the expense of the flexibility of adjustments of the recording position using drive-mounted electrodes. Preliminary experiments indicate that auditory trace fear conditioning in the DualCage produces learning-specific changes in fast extracellular field potential oscillations in the CA1 region of the hippocampus of male C57BL6/J mice.

- Stiedl, O., Jansen, R.F., Pieneman, A.W., Ögren, S.O., Meyer, M. Assessing emotional states through the heart in mice: implications for cardiovascular dysregulation in affective disorders. *Neurosci. Biobehav. Rev.*, 33 (2009), 181-190.
- Misane, I., Tovote, P., Meyer, M., Spiess, J., Ögren, S.O., Stiedl, O. Time-dependent involvement of the dorsal hippocampus in trace fear conditioning in mice. *Hippocampus*, 15 (2005), 418-426.
- 3. Schregardus, D.S., Pieneman, A.W., Ter Maat, A., Jansen, R.F., Brouwer, T.J., Gahr M.L. A lightweight telemetry system for recording neuronal activity in freely behaving small animals. *J. Neurosci. Methods.*, *155* (2006), 62-71.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Symposium: Measuring Surgical Behavior

Gabrielle J.M. Tuijthof Dept. of BioMechanical Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands g.j.m.tuijthof@tudelft.nl

Medical technology is evolving quickly. This is also true in the operating theatre, where the introduction of minimally invasive surgery has brought many benefits for the patient: less morbidity, and quicker recovery. Despite these advances in technology, patient safety cannot be guaranteed for 100%. Several factors can be summed that contribute to this:

- New surgical devices are not always sufficiently evaluated. In the operative setting, it should be demonstrated that the usefulness and usability of a new device is superior to conventional techniques.
- The quality and skills of surgeons have been considered to be at a constant competent level. In reality, this is not the case as every surgeon is different and his/her competency can fluctuate over time.
- The traditional training methods of surgical skills have not been adjusted according to the advances in technology. The result is long learning curves and increased risk of surgical errors.
- Emergency and critical situations occur in the operating theatre, which make it a highly dynamic environment. Lack of protocol and proper communication can increase the risk of surgical errors.
- All patients are unique. This implies that basically each operation requires a new learning experience.

As the demands from government and society increase for high quality and affordable medical care, initiatives are taken to guarantee predefined surgical quality and reduce the risk of surgical errors. To achieve this, measurements are performed in the operating theatre to document surgical behavior. Physically this is not an easy task, as patient care has number one priority, patient privacy and the sterile operating zone should be respected, and the operating theatre cannot be transformed into an experimental set up. Besides these practical issues, interpretation of recorded data is another huge challenge. How do you define surgical performance or surgical error, and how do you interpret video images or spoken language? A growing number of research initiatives attempts to address these challenges.

• First, analysis of the current operative setting is performed to document surgical actions, communication and instrument usage for identification of key problems and constructing base line reference. Video and sound recordings with multiple cameras are usually performed



Figure 1. Example video recording equipment in the OR and a software interface that enables simultaneous recording of two video images and data signals.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

for this purpose, with off-line time-action analysis.

- Second, assessment of surgical performance receives huge attention. A wide range of methods is applied to determine this: checklists, global rating scales, motion analysis, force measurements, video-assessment and (virtual reality) simulators. The variety of methods and application purposes are not fully explored.
- Third, development of a digital operating room assistant has been started, which should automatically register deviations from a routine operation. It warns the operating team to prevent errors and increase patient safety. Automated image detection and Hidden Markov models are a few examples that are applied to achieve this.

This symposium aims to give an overview of different methods as developed by researchers in the surgical field. They will all discuss their own efforts to meet the challenge of measuring surgical behaviour. Finally, we can point out a direction towards a future ultimate registration tool.

SYMPOSIUM CONTENTS

Teaching Arthroscopy: Analysis of Verbal Communication in The Operating Room

Gabrielle Tuijthof (Delft University of Technology / Academic Medical Centre, The Netherlands), Alexander Vos, Inger Sierevelt, Mattias Schafroth & Gino Kerkhoffs (Academic Medical Centre, The Netherlands).

Methods for Automatic Statistical Modeling of Surgical Workflow

Tobias Blum, Nassir Navab & Hubertus Feußner (Technische Universität München, Germany).

In-Vivo Measuring Surgical Workflow Activities in the OR

Loubna Bouarfa (Delft University of Technology, The Netherlands), Laurents P. S. Stassen (Academic Hospital Maastricht, The Netherlands), Pieter. P. Jonker & Jenny Dankelman (Delft University of Technology).

Learning Curve Assessment and Identification of Surgical Pitfalls of a New Hip Prosthesis Using Time-Action Analysis

Jakob van Oldenrijk & Elisa Rijk (Academic Medical Center Amsterdam, The Netherlands), Wouter Runne (Onze Lieve Vrouwe Gasthuis Amsterdam, The Netherlands) and Cees van Egmond (Isala Klinieken Zwolle, The Netherlands).
Teaching Arthroscopy: Analysis of Verbal Communication in the Operating Room

Gabrielle J.M. Tuijthof

Dept. of BioMechanical Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands g.j.m.tuijthof@tudelft.nl

Inger N. Sierevelt

C. Niek van Dijk

Alexander G. Vos

Orthotrauma Research Center Amsterdam, dept of Orthopedic Surgery, Academic Medical Centre, Amsterdam, The Netherlands

Matthias U. Schafroth

Gino M.M.J. Kerkhoffs

Orthotrauma Research Center Amsterdam, dept of Orthopedic Surgery, Academic Medical Centre, Amsterdam, The Netherlands

ABSTRACT

Objective evaluation of training performance remains a challenge in assessing the effectiveness of new training methods in the operating room. A classification method for verbal communication was adopted to analyze typical communication patterns in teaching arthroscopy and to verify its potential as evaluation tool. Four residents who were supervised by either one of two participating surgeons performed 18 operations that were recorded and classified. Typical communication patterns were identified, where types relevant for active learning ('questioning' and 'feedback') were infrequently used. More research is required to confirm if documentation of verbal communication is suitable to assess training performance objectively.

Author Keywords

Arthroscopy, communication, training, performance, classification.

INTRODUCTION

Arthroscopy (minimally invasive surgery in joints) as an alternative to open operative treatment is a beneficial option for patients [1-3]. Arthroscopic operation techniques require high surgical skills and show long learning curves [4-6]. In modern medicine, with the overall demand for high quality care in contrast to a reduced period of time for residents to

develop their arthroscopic skills, changes in training methods are initiated [7;8]. From this perspective, it is worthwhile to optimize the learning effect per operation. Objective evaluation of training performance remains a challenge in assessing the effectiveness of new training methods in the operating room. Verbal communication reflects the interaction between the surgeon and the resident. Recently, an objective classification method was introduced based on the type and the content of the verbal communication during surgical training [9]. The goals were to adopt this classification method for an arthroscopic setting, to analyze typical communication patterns and to verify its potential use in objective assessment of training performance.

METHODS

Within a period of two times 3 months, 18 arthroscopic knee procedures were recorded with a special capturing system consisting of two video cameras – one from the

| athus VS Standard - [Analyse] | | | | | | | -126 |
|--|---|--|--|------------------------------|---|----------|------------|
| a factored ferrentiane facility of the second ferrentiane formation for the second ferrentiane for the | | e ki | | | | | - 4 3 |
| Depres 21 | | Domatic 21 Alexander Vedical Communication | 641 | _ | | 30 | (gonetnem |
| | | lde colimit | | | er unkonne ming manding fonna here ellanenca | | |
| | | H: Dat D 41 0011.8 00 42 0011.5 00 43 0011.5 00 44 0012.5 00 45 0011.0 00 46 0012.00 00 47 0013.15 00 48 0013.20 00 49 0012.20 00 50 0013.5 00 51 0013.40 00 52 0013.40 00 53 0014.00 00 54 0013.91 00 50 0013.91 00 51 0014.00 00 54 0014.01 00 50 0014.01 00 50 0014.01 00 50 0014.01 00 | Censo 12.36 Censo ve. 12.36 Censo ve. 15.52 Sept and 15.52 Sept and 15.54 Kept and 15.54 Kept and 15.54 Kept and tour Tim. 15.54 Kept and tour Tim. 15.25 Kept and tour Tim. 15.25 Kept and tour Tim. 15.26 Kept and tour Tim. 15.27 Maior mod. 15.28 Kept and tour Tim. 15.29 Kept and tour Tim. 15.20 Kept and tour Tim. 15.21 Maior mod. 15.22 Kept and tour Tim. 15.24 Censo tour Tim. 15.25 Kept and tour Tim. 15.26 Censo tour Tim. 15.27 Kept and tour Tim. 15.28 Censo tour Tim. 15.29 Censo tour Tim. 16.24 Conso tour Tim. | | Denotweing | | |
| | - | No Bestand T means211 a 2 oprama22 a | () at 0.000 | 5kg 0.0017718 0.000541 | Opværindelsen. | | _ |
| General General | - | | | | | int or 1 | 155 (2 |

Figure 1. Screenshot of software used to analyze verbal communication.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

| | Total | Explaining | Commanding | Questioning | Feedback | Miscellaneous |
|--|--------|------------|------------|-------------|----------|---------------|
| Total | 100.0% | 38.8% | 27.4% | 5.7% | 10.6% | 17.4% |
| Operation Method | 4.2% | 3.6% | 0.1% | 0.1% | 0.5% | 0.0% |
| Anatomy and pathology | 17.7% | 14.8% | 0.0% | 2.4% | 0.5% | 0.0% |
| Instrument handling and tissue interaction | 35.7% | 13.0% | 14.2% | 2.0% | 6.4% | 0.1% |
| Visualization | 24.9% | 7.4% | 13.1% | 1.2% | 3.2% | 0.0% |
| Miscellaneous | 14.0% | 0.0% | 0.1% | 0.1% | 0.0% | 13.9% |
| Indefinable | 3.5% | 0.0% | 0.0% | 0.0% | 0.0% | 3.5% |

Table 1. Crosstabs for type (upper row) and content (left column) categories as percentage of total events.

arthroscopic camera and one of the hands of the residents (digital CCD camera, 21CW, Sony CCD, Tokyo, Japan) and a tie-clip microphone (ECM-3003, Monacor, Bremen, Germany) that was mounted on the supervising surgeon. The video images were combined by a colour quad processor (GS-C4CQR, Golden state Instrument co., Tustin, USA) and digitized simultaneously with the sound by an A/D-converter (ADVC 110, GV Thomson, Paris, France). Four residents who were supervised by either one of two participating surgeons performed the operations. The four communication types, previously introduced, were adopted Blom et al. [9]: explaining, questioning, commanding and miscellaneous (Table 1). As this study specifically focuses on training, one category was added: feedback which concerns judgment of the teaching surgeon on the actions of the resident. Categories for communication content were discussed and refined by analyzing two operations performed by different residents until mutual agreement in scoring behavior was achieved by two observers. This resulted in six domains: operation method (that has an accent on steps that have to be taken in the near future e.g. start creating the second portal), anatomy and pathology, instrument handling and tissue interaction (e.g. open punch, reposition instrument, stress joint, increase portal size, push meniscus backwards), visualization (e.g. move scope, irrigation, focus), miscellaneous (general or private), and indefinable (Table 1). All digitized videos of the operations were analyzed with Utilius VS 4.3.2 (Campus-Computer-Center, Leipzig, Germany), where one observer marked all communication events and assigned them into proper categories (Fig. 1). Communication on one type and one content was marked as one event independently of its length. The frequency of events as percentage of total events in each of the categories was determined (Table 1). Subsequently, a multivariable linear regression analysis was performed to determine if the teaching surgeon and the experience of the residents significantly influenced the frequency of communication events per minute (p < 0.05).

RESULTS

On average 6.0 (SD 1.8) communication events took place every minute. The communication types 'explaining' and 'commanding' show considerable frequency compared to 'questioning' and 'feedback' (Table 1). The explaining events were primarily on 'anatomy and pathology' followed by 'instrument handling and tissue interaction'. The commanding events were primarily on 'instrument handling and tissue interaction' and 'visualization', which in general were the most frequent communication content categories (Table 1). A difference in mean events per minute was found between both teaching surgeons (p < 0.05). No significant correlation was found between the frequency of events and the experience of the residents.

CONCLUSION

The results highlight distinctive communication patterns. The relative high frequency of the types 'explaining' and 'commanding' as opposed to 'questioning' and 'feedback' is noticeable as the latter two stimulate active learning in general. Additionally, explaining on the contents 'anatomy and pathology' and 'instrument handling and tissue interaction' is considerable. These items are particularly suitable for training outside the operating room. If trained so, more options are left to focus on other learning goals. As a clear difference was present between the frequency of events per minute amongst the surgeons and no correlation was found for the experience of residents, we cannot confirm that this method is suitable as objective evaluation tool for new training methods. Additional research is recommended with a larger group of residents to minimize the effect of outliers.

- 1. LaPorta, G., Turrisi, J.S. The history and clinical application of arthroscopy. *Clin. Podiatr. Med. Surg.*, *4*, 4 (1987), 829-834.
- 2. Pedowitz R.A., Esch, J., Snyder, S. Evaluation of a virtual reality simulator for arthroscopy skills development. *Arthroscopy*, *18*, 6 (2002), E29.

- Scholz, J., Kuhling, T., Turczynsky, T. Advantages of arthroscopy of the knee joint. *Biomed. Tech. (Berl)*, 37, 1-2 (1992), 11-13.
- Mabrey, J.D., Cannon, W.D., Gillogly, S.D., Kasser, J.R., Sweeney, H.J., Zarins, B., *et al.* Development of a virtual reality arthroscopic knee simulator. *Stud. Health Technol. Inform.*, 70 (2000), 192-194.
- 5. Meyer, R.D., Tamarapalli, J.R., Lemons, J.E. Arthroscopy training using a "black box" technique. *Arthroscopy*, 9, 3 (1993), 338-340.
- Morris, A.H., Jennings, J.E., Stone, R.G., Katz, J.A., Garroway, R.Y., Hendler, R.C. Guidelines for privileges in arthroscopic surgery. *Arthroscopy*, 9, 1 (1993), 125-127.

- Bridges, M., Diamond, D.L. The financial impact of teaching surgical residents in the operating room. *Am. J. Surg.*, 177, 1 (1999), 28-32.
- Howells, N.R., Gill, H.S., Carr, A.J., Price, A.J., Rees, J.L. Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. *J. Bone Joint Surg. Br.*, 90, 4 (2008), 494-499.
- Blom, E.M., Verdaasdonk, E.G., Stassen, L.P., Stassen, H.G., Wieringa, P.A., Dankelman, J. Analysis of verbal communication during teaching in the operating room and the potentials for surgical training. *Surg. Endosc.*, 21, 9 (2007), 560-566.

Methods for Automatic Statistical Modeling of Surgical Workflow

Tobias Blum blum@in.tum.de

Nassir Navab navab@in.tum.de

Technische Universität München Computer Aided Medical Procedures, Boltzmannstr. 3, 85748 Garching, Germany

ABSTRACT

An increasing amount of data can be recorded automatically during surgeries. This has led to a recent interest in methods that allow statistical modeling of surgical workflow from this data. One main challenge of these methods is how to automatically generate a statistical model from a set of recorded surgeries. In this paper we will review different methods and show advantages and drawbacks.

Author Keywords

Surgical workflow, statistical models of time series.

INTRODUCTION

Trough the introduction of new technology into the operating room it is now possible to record an increasing amount of signals during a surgery. Video images can be obtained, especially for laparoscopic surgeries, position and orientation data is available from tracking systems or robotic systems. Motion sensors, like accelerometers, are becoming cheap and small and technologies like RFID might be used in the future to identify instruments and estimate their presence. Also training systems that use e.g. haptic feedback devices are an important source of data.

As this kind of information becomes available, there is recent interest in methods to make use of this data. Possible applications are analysis of behavior and workflows, assessment of surgical skill and context-sensitive systems that can detect the current work step of a surgery. All these applications require some sort of statistical model of the workflow. Automatically building such models is an important topic in this domain. We will review and compare different methods for this task.

Hubertus Feußner

Technische Universität München Department of Surgery, Klinikum Rechts der Isar, Ismaninger Str. 22, 81675 München, Germany feussner@chir.med.tu-muenchen.de

METHODS

The most common method for modeling surgical workflow is to use Hidden Markov Models (HMM). HMMs model time series by a directed graph. Each node or state in the graph has transition probabilities representing the probability of advancing to other states. Each state also has an observation symbol distribution that represents the probability of observing a certain observation symbol in that state. For most applications in the field of surgical workflow modeling one tries to build a model from a set of training data such that the model represents the training data well. We will discuss several methods to automatically build HMMs. We use data that represents the use of instruments during ten laparoscopic cholecystectomies. Each surgery was split into 14 phases. The models are trained independently for each phase. For each method, the capabilities to automatically detect the current phase of a surgery and to generate models that are understandable by humans are discussed. Results are summarized in Table 1.

The most common method to train a HMM is to initialize it with random transition and observation probabilities as for example done in [1]. Next expectation maximization (EM) is used to adapt the HMM to the training data. This procedure is repeated multiple times until a good HMM is found. Many extensions of this method have been proposed e.g. to estimate the number of states that are required to model a certain procedure. These methods work, are easy to implement and do not require long computation time. Furthermore they have shown good results for automatically detecting the phase of a workflow. However the resulting model is usually hardly understandable to humans, which makes them practically useless for analysis of human behavior.

In the case of discrete observations, like the use of instruments, Markov Chains are an alternative. They are similar to HMMs and have been used for workflow modeling in [2]. While states of an HMM can represent the use of different instruments, a Markov Chain state represents only one instrument. Markov Chains can be built simply by counting transitions from one observation symbol to another. They give a relative good impression of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



| Model | Detection | Speed | Understandable |
|-----------------|-----------|-----------|----------------|
| Random | 6,7% | fast | bad |
| Markov Chain | 5,3% | very fast | medium |
| SSS | 12,5% | medium | good |
| Merging | 6,4% | slow | good |

Table 1. Results from different methods.

RESULTS

We have compared different aspects of the models. One interesting application is to build context-aware system. In order to estimate whether the methods are suitable for this application, we tested whether it is possible to detect the current phase of a running surgery. This measure also indicates if the model is a good statistical representation of the workflow. All methods but SSS have shown to deliver reasonable results here. Another important metric is whether a certain method can create models that reflect the workflow in a human understandable way. This is true for the data-derived methods. However, from our experience it has shown that these methods are harder to implement and unlike for standard methods, there are no toolkits or libraries available where these methods are already implemented.

CONCLUSION

Automatically building statistical methods is important, especially as a huge amount of data is becoming available, which makes manual modeling difficult. Using the right training methods, HMMs can be built that are suitable for automatic recognition systems and for human understandable models. Challenges in the future will be to extend these models such that they work on huge datasets and to incorporate different kinds of sensors at the same time. Future models should also take into account patient or surgeon related information.

REFERENCES

- Leong , J., Nicolaou, M., Atallah, L., Mylonas, G., Darzi, A., Yang, G-Z. HMM Assessment of Quality of Movement Trajectory in Laparoscopic Surgery. *MICCAI* (2006), 335-346.
- 2. Mackel, T.R., Rosen, J., Pugh, C. M. Markov Model Assessment of Subjects' Clinical Skill Using the E-Pelvis Physical Simulator. *IEEE Trans Biomed Eng*, *5*, 12 (2007), 2133-2141.
- Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., Hager, G. Data-Derived Models for Segmentation with Application to Surgical Assessment and Training, *MICCAI* (2009), 426-434.
- 4. Blum, T., Padoy, N., Feußner, H., Navab, N. Modeling and Online Recognition of Surgical Phases using Hidden Markov Models. *MICCAI* (2008), 379-386.

Figure 1. Markov chain representing one phase of a laparoscopic cholecystectomies.

the frequency of instrument use and are much easier to understand for humans then randomly initialized HMMs. Also automatic recognition rates are good. However they do not model the temporal order of actions within one phase are do therefore not model repetitions well. An example can be seen in Figure 1.

The methods that showed best results for generating human-understandable models are data-derived HMMs. One example is successive state splitting (SSS) [3]. This method starts with a model consisting only of one state and splits states until a good model is found. So e.g. if we have one state that represents that first one instrument is used and next another, this method will do a temporal split so that two successive states are used instead. If one state represents alternating use of two instruments it will do a contextual split into two states. While this method delivers models that are intuitive to understand, as can be seen in Figure 2, for our data it failed to deliver models that statistically well represent the workflow and are therefore not suited for automatic recognition systems.

Another data-derived method that works the other way round is model merging [4], where an initial very big model is built that represents every single action that occurred in any of the training surgeries. Iteratively, a smaller model is generated by merging similar nodes. As SSS, model merging well represents the temporal order of actions and is also well suited to detect workflow steps. The main drawback of this method is that it takes several hours to train it, even for a relative small training set consisting only of nine surgeries.



Figure 2. HMM that has been generated by model merging.

In-Vivo Measuring Surgical Workflow Activities in the OR

Loubna Bouarfa

Department of Biomechanical Engineering, Delft University of Technology, The Netherlands l.bouarfa@tudelft.nl Laurents P.S. Stassen Academic Hospital Maastricht, The Netherlands lps.stassen@mumc.nl Pieter P. Jonker Jenny Dankelman

Department of Biomechanical Engineering, Delft University of Technology, The Netherlands

ABSTRACT

Measuring surgical workflow activities in-vivo is a crucial task for medical computing to both track errors and reduce the costs in the OR. In this abstract we discuss our deployed framework for measuring surgical activities. Further we discuss the results of a previous pilot study using the proposed framework to measure surgeon activities in Laparoscopic Cholecystectomy procedures.

Author Keywords

Surgical workflow, environmental cues, HMM, signal simulation, sensing devices.

INTRODUCTION

Within the hospital, the Operating Room (OR) is one of the most resource intensive environments in healthcare, because it involves expensive labor resources (surgeon, anesthetists, nurses, etc.), expensive equipment and daily maintenance. In the Netherlands, the aggregated estimated cost of an OR is \notin 1100 per hour. It is therefore very important to maximize the OR efficiency.

In the Netherlands, each year more than 1700 patients die from preventable surgical errors, which are commonly known as Adverse Events (AEs) [1]. Surgical AEs account for one-half to three-quarters of all AEs in healthcare [2]. The annual cost of all AEs in the Netherlands is estimated to amount to a total of \notin 355 million, of which \notin 161 million is for preventable AEs in 2004 [3]; this is approximately 1% of the annual hospital budget of \notin 14.5 billion. Therefore, any improvement of patient safety by preventing AEs is cost saving.

Another aspect of OR cost management problems is the delay time between surgeries. A benchmark study showed that the average delay in surgery start times ranged from 25 to 103 minutes [4]. This is shown to be caused by the

failure of commonly used planning tools which do not account for the unpredictable time duration of operations. In the Netherlands those lost hours add up to 2150 hours, a loss of \notin 2.3 million per annum.

By having an adequate in-vivo workflow registration system for the OR deployed in the hospital, AEs can be consistently tracked allowing the hospital administrators to investigate the root cause of the problem and associated departments, staff or surgeons. Moreover, high quality data from the registration system can be used to reduce the start time delays in the OR and better estimate the length of a procedure for the planning system.

The goal behind this research is to measure surgical activities in-vivo in the OR in order to track the source of AEs and reduce the cost in healthcare, whilst keeping the quality and quantity of provided care equal. The general framework of such measuring system is discussed in section 2. Section 3 discusses a pilot study using the framework in section 2 for measuring Laparoscopic Cholecystectomy procedures. Our conclusions and future directions are discussed in section 3.

SYSTEM FRAMEWORK

Figure 1 illustrates the general framework of a surgical workflow registration system as used in our previous work [5, 6].

In step 1, we define the surgical workflow activities that need to be monitored in the OR. For this purpose we utilize existing surgical guidelines and protocols. At the end of this step we have a workflow model of the surgical activities

| Defining workflow tasks | (environment cues) | Training (workflow steps) | (workflow steps) |
|--|---|--|--|
| -Surgical guidelines -Surgical protocol | | • • • • • • • • • • • • • • • • • • • | |
| Observations OR -Brainstorming surgeons | -SimulationImplem entation: - Sensing devices | - (semi-) Supervised probabilistic models, e.g HMM | - Dynamic programming, e.g. Viterbi Algorithm |

Figure 1. General framework for a surgical workflow system.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

that needs to be in-vivo monitored in the OR.

In the second step we monitor environmental cues in the OR that may be significant for the activities derived in step 1. If the cues are easy to monitor with available easy-to-use sensors we acquire them automatically from the OR. In case adequate sensors are not available yet, we simulate the signal cues by means of video annotation and test their significance for the identification process (steps 3 and 4). In the third step, a probabilistic classification system is trained to infer the workflow steps from the environment cues. The system is trained with the cues (of step 2) as input features, and the workflow activities (of step 1) as different classes of surgical behavior.

After the system has been trained, we use dynamic programming to infer workflow activities from the input cues. The aim of the system is to record surgical activities in the OR in such a way they can be compared to the surgical protocol [5, 6].

PILOT [5]: LAPAROSCOPIC CHOLECYSTECTOMY

Laparoscopic Cholecystectomy (LC) is chosen because it is one of the most frequent and standardized procedures in minimally invasive surgery.

Step 1: Defining Workflow Steps

In the first step we defined the workflow activities of LC procedures by using LC protocol and discuss those activities with the cooperating surgeon. The workflow model of LC consists of 5 main surgical activities that need to be monitored. By further brainstorming with surgeons we assumed that instrument signals (i.e. instruments is usage) are significant cues in the OR to measure surgical activities.

Step 2: Monitoring Environment Cues

At the moment we do not have an adequate in-vivo sensing device to measure which instruments are being used in the LC procedure. In a previous study [5], we simulated the instrument signals by annotating laparoscopic video recorded during the operation. Figure 2 illustrates the simulated instrument cues as binary signals corresponding to instrument utilization; 1 if an instrument is in used, 0 if not in use; As ground truth data we measured also the surgical workflow activities (in step 1) as discrete time signals (Figure 3).



Figure 2. simulated instrument signals.



Figure 3. simulated instrument signals.

Step 3 & 4: training & Inference

In the training step [5], a Hidden Markov Model (HMM) is trained with 10 datasets from (step 2) using the binary instrument signals as observations and the ground truth surgical activity signals as hidden states.

After the training step, the system is ready for use. For a new set of instrument signals, derived form a new LC procedure the Viterbi algorithm is used to infer the surgical activities from instrument cues. The system shows a total accuracy of 90% of detected states as discussed in [5]. As expected, activities with higher accuracies were generally those with more data points (i.e. more instruments used).

CONCLUSIONS AND DISCUSSION

Recording surgical activities is crucial to save both lives and money in hospitals. After proving the significant cues for recording activities, we need to design appropriate sensing hardware or software solutions for in-vivo measurement. In the pilot study in section 3, we showed that instruments utilization is an important cue for the OR. The next step is to design a sensing device to allow tracking of instrument cues in-vivo during surgery.

- De Bruijne, M.C., Zegers, M., Hoonhout, L.H.F., Wagner, C, Onbedoelde schade in Nederlandse ziekenhuizen: Dossieronderzoek van ziekenhuizenopnames in 2004. Amsterdam: EMGO Instituut and NIVEL, The Netherlands, 2007.
- 2. Griffin, F. and D. Classen, Detection of adverse events in surgical patients using the Trigger Tool approach. *Quality and Safety in Health Care*, *17*, 4 (2008), 253.
- 3. Hoonhout, L., et al., Direct medical costs of adverse events in Dutch hospitals. *BMC Health Services Research*, 9, 1 (2009), 27.
- 4. Does, R., et al., Reducing start time delays in operating rooms. *Journal of Quality Technology*, *41*, 1 (2009), 95-109.
- 5. Bouarfa, L., P. Jonker, and J. Dankelman, Discovery of high-level tasks in the operating room. *Journal of Biomedical Informatics*, (2010).
- 6. Bouarfa, L., P. Jonker, and J. Dankelman. Surgical context discovery by monitoring low-level activities in the OR. (2009).

Learning Curve Assessment and Identification of Surgical Pitfalls of a New Hip Prosthesis Using Time-Action Analysis

Jakob van Oldenrijk jakobvanoldenrijk@gmail.com Elisa Rijk elisa_rijk@hotmail.com Matthias U. Schafroth M.U.Schafroth@amc.uva.nl

Department of Orthopaedic Surgery, Academic Medical Centre Amsterdam Meibergdreef , 1100 DD, Amsterdam, The Netherlands

Wouter C. Runne wouter@runne.nl **Rudolf W. Poolman** rwp@jointresearch.org

Department of Orthopaedic Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands

Cees C.P.M.Verheyen C.C.P.M.Verheyen@isala.nl **Cees van Egmond** c.van.egmond@isala.nl

Department of Orthopaedic Surgery, Isala klinieken Zwolle, The Netherlands

ABSTRACT

Outcome measures in hip joint replacement surgery, such as complication rates, are usually rare and require an extensive follow-up. These outcome measure are therefore often inadequate to monitor individual learning curves. We use time-action analysis to determine the level of efficiency of individual steps of a surgical procedure and thereby study the learning curve of surgeons and identify possible pitfalls. By analysing the unedited video recordings of the first 20 procedures of 4 surgeons using sports analysis software, the duration and the number of repetitions of each action during a surgical procedure is measured. This allows the construction of individual learning curves for each surgeon. Actions with a duration of more than the mean + 1sd are identified as inefficient. The video recordings of these actions are reviewed with the surgeon and the possible pitfall and solution is discussed after finalizing all procedures.

Author Keywords

Time action analysis, surgery, learning curve, pitfalls,

efficiency, hip prosthesis.

INTRODUCTION

In hip joint replacement surgery, outcome measures like complication rates are rare or require an extensive followup. They are therefore often inadequate to monitor individual learning curves. Time-action analysis (TAA) is a tool to objectively determine the level of efficiency of individual steps of a surgical procedure and identify possible pitfalls during a learning curve [1].

METHODS

By analyzing the unedited video recordings of the first 20 procedures of 4 surgeons, the number and duration of the actions needed for a surgeon to achieve his goal and the efficiency of these actions is measured.

We constructed a taxonomy or list of actions which together describe the complete surgical procedure. In the taxonomy we categorized the procedure in 5 different Goal Oriented Phases (GOP):

- 1. the incision phase.
- 2. the femoral osteotomy (bone cut) phase.
- 3. the acetabulum (hip joint socket) phase.
- 4. the stem phase.
- 5. the closure phase.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Setup in the operating theatre.

Since the incision phase and closure phase do not differ from the conventional surgical technique we primarily focus on the femoral osteotomy phase, the acetabulum phase and the stem phase.

Each GOP was subdivided in Goal Oriented Actions (GOA) and each GOA is subdivided in Separate Actions (SA) thereby defining all the necessary actions to complete the procedure. We grouped the SAs into GOAs since it would not be feasible to measure each SA.

Using Utilius vs. video sports analysis software (CCC software, Markleeberg, Germany), the duration of each GOA is recorded as well as the amount of delay. Actions



Figure 2. the complete surgical procedure is described by a list of actions, the taxonomy.

with a duration of more than the mean + 1sd are identified as inefficient. The video recordings of inefficient actions are reviewed and the possible problem and solution discussed with surgeons after finalizing all procedures, thereby constructing a list of potential pitfalls.

Ethical Statement

Since this trial is an observational study, a formal ethical approval was waived for this study by the OLVG Medical Ethics Committee. The OLVG Medical Ethics Committee declared to have no objections to this trial.

CONCLUSION

We describe time-action analysis as a method to identify potential problems during the learning curve of a new surgical procedure and to assess the level of efficiency of each surgeon individually.

REFERENCES

1. van Oldenrijk J, Schafroth MU, Bhandari M, Runne WC, Poolman RW. Time-action analysis (TAA) of the surgical technique implanting the collum femoris preserving (CFP) hip arthroplasty. TAASTIC trial identifying pitfalls during the learning curve of surgeons participating in a subsequent randomized controlled trial (an observational study). *BMC Musculoskelet Disord.*, 9 (2008), 93.

Symposium: Monitoring of Social Interactions and Initial Signs of Conflicts in Isolated Groups During Long-Term Missions

Roman Gorbunov

Faculty of Industrial Design, Technical University of Eindhoven, The Netherlands r.gorbunov@tue.nl Karl Tuyls Ddept. of Knowledge Engineering, Maastricht University, The Netherlands k.tuyls@maastrichtuniversity.nl

Emilia Barakova

Faculty of Industrial Design Technical University of Eindhoven, The Netherlands e.i.barakova@tue.nl

ABSTRACT

There are many missions which should be performed by small groups of people (crews) working in hard and extreme external conditions for a long period of time (space missions, submarines, polar stations and expeditions). Such missions are usually accompanied by a set of negative factors which have a strong influence on the physiological, psychological and sociological state of the crew, determining group performance and, as a consequence, the success of the whole mission. Interpersonal relations withing the group is one the main factors influencing performance of the crew. Inter-crew tension can cause formation of subgroups, disruptions of cohesion, scapegoating, refusal of communication and collaboration leading to a decrease of mission quality or even failure of the mission. Social health of the crew and psychological health of crew members are strongly related with each other being a complex psychosocial phenomenon.

In this respect it is important to have a possibility to monitor psychosocial state of the group during the mission. This approach requires a development of techniques for monitoring of social interactions, non-verbal behavior and initial signs of depression or aggression. Usage of such techniques can help to diagnose, predict and prevent conflicts on their earlier stages. There are different approaches to measure human behavior. Analysis of nonverbal communication is one of the powerful tool for monitoring mental state and relations among the crewmembers. Automatic recognition of face expression and voice intonation as well as analysis of body movements are promising techniques for monitoring psychological states of crew members as well as interpersonal relations in the group. Usage of virtual environments as a medium for

interactions between the crew members (for example online games) is another promising way to approach the problem of monitoring interpersonal relations in the group. In this approach interactions among the group members is restricted and formalized by the construction and, in this way, it is possible to simplify the data for the analysis as well as to focus on specific aspects of the relations between the group members. With the current fast development of multimedia technologies the gap between the real face-toface communication and virtual communication decreases. As a consequence the monitoring of physical parameters of real face-to-face communication and usage of virtual environments for interaction between peoples can be considered as techniques approaching the problem of monitoring human-human interaction from different directions with a tendency to merge into one broad filed of research in the future. In this respect a collaboration between researchers working in different filed is desirable.

Additionally to that virtual words were also shown to be effective tools in treatment of psychological disorder (different phobias). In this way, the role playing games, based on the usage of specially designed virtual words, can perform not only passive (monitoring) but also active (prevention and treatment) role with respect to psychological and interpersonal problems. Virtual environments can be enriched by intelligent virtual agents which can assist crew-members in performing complicated tasks which require extensive knowledge and intellectual resources. Maintenance of psychological and psychosocial health can be considered as one of the tasks performed by the virtual agents. Since virtual agents need to solve a variety of complex problems, state of the art techniques should be used for their development (evolutionary game theory, reinforcement learning).

The above mentioned challenges require a collaborative effort from leading researchers from different fields. This symposium is intended to bring together specialist from different disciplines to enable a productive and diverse discussing on the subject. The exchange of experience has a significant importance for further development of the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

computer systems for measuring human behavior and monitoring psychological and social health in small groups of people.

Author Keywords

Small groups, conflicts, isolation, interpersonal relations, nonverbal communications, emotional isolation, special confinements, long-term missions, conflicts predictions, space missions.

SYMPOSIUM CONTENTS

PSPA - Test for the Study of Individual Values and Mutual Perception in Small Groups

Vadim Gushin & Alla Vinokhodova (State Research Center, Institute for Biomedical Problems RAS, Russia).

Electronic Partners That Diagnose, Guide and Mediate Space Crew's Social, Cognitive and Affective Processes Jurriaan van Diggelen, Mark Neerincx, & Nanja Smets (TNO, The Netherlands) Mikael Wolff (ESA-ESTEC, The Netherlands) and Leo Breebaart (S&T, The Netherlands).

Automatic Mental Heath Assistant: Monitoring and Measuring Nonverbal Behavior of the Crew During Long-Term Missions

Natalia Voynarovskaya, Roman Gorbunov, Emilia Barakova, Matthias Rauterberg (Technical University of Eindhoven, The Netherlands).

PSPA - Test for the Study of Individual Values and Mutual Perception in Small Group

Vadim Gushin vgushin.57@mail.ru Alla Vinokhodova vinokhodova@imbp.ru

State Research Center, Institute for Biomedical Problems RAS, Russia

ABSTRACT

The new method for the small group under confinement observation is described. Results gained with PSPA testify that heterogeneity within small groups in extreme environments promotes the growth of conflict and tension, and it negatively affects the behavior and actions of the group members.

Author Keywords

Small group, hyperbaric chambers, repertory grid, individual values.

In the last ten years, several simulation studies took place, where multinational crews have been confined in hyperbaric chambers for prolonged periods to model aspects of space missions. The relations between crewmembers with the use of PSPA-test have been studied in international simulation experiments with participation of flight personnel provided by European Space Agency, Canadian Space Agency, NASDA, and IBMP (EXEMSI-92, HUBES-94, CAPSULS-94, ECOPSY-95, SFINCSS-99), in a French Antarctic study, sponsored by ESA (IAPP). Data from isolated groups in analogue situations (hibernation, expeditions) and in model experiments has supported the hypothesis that behavior inside a small group can be also defined by the specifics of crewmembers' perceptions of themselves and of each other, by their selfidentification with the group, and by changes under pressure of the stressors found in this extreme environment, such as isolation, monotony, and social deprivation. Experiments have shown that heterogeneity within small groups in extreme environments promotes the growth of conflict and tension, and it negatively affects the behavior and actions of the group members [1-2, 4-5].

Personal Self Perception and Attitudes (PSPA) is a system for analyzing participants' subjective attitudes. It is based on the semantic differential of Osgood and the repertory grid technique of F. Fransella and D. Bannister and allows to investigate the aspects related to crew cohesion, individual values and group identification [1]. Subject is choosing assessment criteria by himself, answering the question: What are the main features (traits) that allow you differentiate people from your close surrounding? Then he has to estimate the extent of psychological similarity between himself and his team, applying previously selected criteria.

We are planning to investigate group dynamics in space crews under the influence of psychologically stressful situation inherent in extended space flight. We expect that the level of group cohesion in space crews could be significantly modified, on the one hand, by the influence of the stress involved in space flight, and, on the other hand, by the heterogeneity of cultural values, attitudes and behavioral preferences in multinational International Space Station crews.

- 1. Fransella, F., Bannister, D. A Manual for Repertory Grid Technique, Academic Press, 1977.
- Gushin, V.I., Efimov, V.A., Smirnova, T.M., Kanas, N. et al Subject's Perception of the Crew Interaction under Prolonged Isolation. *Aviation, Space and Environmental Medicine*, 69, 6 (1998) 555-561.
- 3. Gushin, V.I., Pystinnikova, J., Smirnova, T. Interrelations between the small isolated groups with Homogeneous and Heterogeneous composition. *Journal of Human performance in Extreme environments*, *6*, 1 (2001), 25-33.
- Kozerenko, O.P., Gushin, V.I., Sled, A.D. et al Some Problems of Group Interaction in Prolonged Space Flights. *Human Performance in Extreme Environments*, 4, 1 (1999), 123-127.
- 5. Vinokhodova, A.G., Bystritskaya, A.F., Eskov, K.N. Inter-Group Relationship under Simulated Long-Term Isolation Conditions. *Space Technology*, *22*, 3-4 (2002), 103-112.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Electronic Partners That Diagnose, Guide and Mediate Space Crew's Social, Cognitive and Affective Processes

Jurriaan van Diggelen jurriaan.vandiggelen@tno.nl +31888662316 Mark Neerincx mark.neerincx@tno.nl +31 3463 56298

TNO Defense, Security, and Safety Kampweg 5, 3769 DE Soesterberg, The Netherlands

ABSTRACT

Long-duration missions (e.g. to the Moon, asteroids, or Mars) require astronauts to collaborate and interact with complex computerized equipment and facilities under dynamic and hazardous conditions. The Mission Execution Crew Assistant (MECA) comprises crew support that acts as an "electronic partner" (ePartner), helping the crew to take care of their mental and social conditions and to schedule crew activities. We have developed a simple prototype of MECA which we use to test MECA's functionalities in analogue environments where test participants are placed in isolation for longer periods of time.

INTRODUCTION

Manned long-duration missions beyond low earth orbit (LEO), e.g. to Mars, set high operational, human factors and technical demands for a distributed crew operations support system. The mission execution crew assistance (MECA) is such a system that shall enhance humanmachine teams' capabilities to cope autonomously with unexpected, complex and potentially hazardous situations. The MECA concept comprises distributed personal ePartners that help the team to assess the situation, to determine a suitable course of actions to solve a problem, to schedule on-the-job training, and to safeguard the astronauts/explorers from failures. One early prototype is being studied in the MARS500 mission, a unique opportunity to learn more about operations support required for long duration missions. Another experiment will be held in the Concordia base station, where twelve test participants will spend the winter season in Antarctica.

FRAMEWORK

Figure 1 illustrates the core high-level MECA functions that will be studied in the MARS500 and Concordia



Figure 1. High-level MECA functions.

experiments. MECA's activity monitoring and scheduling support proceeds in four stages, starting in the upper left corner:

- 1. Supporting crew activities: MECA is always present. In this sense MECA provides continuous support. However, for some activities, MECA is latently present and only becomes active in off-nominal situations. For other activities, MECA plays a more prominent role and provides the framework itself within which these activities take place.
- 2. Monitoring user state: MECA monitors the crew members' performed activities. This includes assessing what the crew members are doing, measuring their performance, and assessing their cognitive state (i.e. their emotional state and task load).
- **3. Providing feedback:** MECA provides feedback to their users by presenting the monitored data. This allows a user to better understand his current task behaviour in the context of his performance over time.
- **4. Scheduling activities:** MECA supports activity scheduling by offering an electronic timeline tool, and by automatically scheduling (or suggesting scheduling) tasks, depending on its interpretation of the data on previous activities.
- **5. Sampling user experiences:** MECA mediates the maintenance and appraisal of memorable experiences and events with a multimedia annotation tool. This function should support the astronauts to reflect on previous activities and happenings in a constructive way to improve resilience (cherishing of successes, coping with stressful events and learning).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Each of these high-level core functions can be implemented at different levels of sophistication. On the one hand, we can think of a full-fledged Artificial Intelligence (AI) implementation, where MECA automatically "reads" the user's mind to assess the user state, provides high level feedback by saying things like "take it easy", "come on", and "don't worry", and by acting as a smart secretary by appropriately (re-)scheduling the user's planning. On the other hand, we can think of a simple implementation, where the system monitors the user state by collecting questionnaire data, where the feedback consists of a statistical interpretation of this data, and where the system helps the user to reschedule the timeline by presenting relevant information.

Because the full-fledged AI implementation is currently still science fiction, MECA's current implementation is closer to the simple implementation than to the AI implementation. Nevertheless, we believe that also the principles behind more advanced support systems can be explored by evaluating more simple prototypes.

SEMI-AUTOMATIC MONITORING

A major focus of MECA is to support team resilience by monitoring team member's performance, and condition. Performance monitoring looks at effectiveness and efficiency of the member's operations and his/her related refresher trainings status, whereas condition monitoring assess the appropriateness of his / her social, cognitive and affective responses to a task or situation. In our approach, monitoring is a joint astronaut-ePartner activity. The ePartner anticipates on (near) future developments of robust sensing technology for monitoring the emotional states of the astronauts. In the current prototype versions, the majority of the "monitoring" data is provided by the human actors.

MECA combines a Cognitive Task Load (CTL) model with a model of the Emotional State (ES) to assess human responses in high-demand task domains. The CTL model distinguishes three load dimensions: the time occupied, the level of information processing (skill-, rule- or knowledgebased behaviour), and number of task-set switches. The ESmodel distinguishes two dimensions: the arousal level—low versus high—and the valence level—positive versus negative. ES and CTL are related: for specific load conditions a specific emotional state ("response") can be expected. For example, when task load increases, an adequate response is to invest extra effort (i.e., arousal increases) in order to maintain good performance.

In a first experiment, we used a Bayesian Network to train the Performance, ES and CTL models with their interrelationships. A "Naïve Bayesian Network" algorithm provided performance predictions with 86% and 74% accuracy for two data-sets, respectively: a data-set from a 'clean' lab-setting and in a dynamic field test environment. Both experiments used the SOWAT as a monitoring tool.



Figure 2. Example screen of the "Collaborative Training" environment for refreshment training of procedures.

MARS 500 EXPERIMENT

In the MARS500 experiment carried out in Russia, a 6person crew is isolated for 520 days to simulate a manned Mars mission (3 Russian, 2 European and 1 Chinese "astronaut"). In this setting, a prolonged and repeated interaction with ePartners' functions will be tested. First, via "Collaborative Training" with rotating trainer/trainee roles, the ePartner guides the joint training of procedures, assessing the interpersonal communication and mediating knowledge transfer from the "expert" to the "novice" crewmember (see Figure 2).

Second, via entertainment gaming the ePartner can be an "affecter" which improves the crewmember's mood. Figure 3 shows a simple entertainment that is integrated into the MECA system.

Third, via a timeline tool the ePartner motivates the crew members to perform certain activities and to maintain a personal (episodic) memory (see Figure 4).



Figure 3. Example screen of Lunar Lander.



Figure 4. The timeline and user experience tool.

The MECA experiment setting for MARS500 is as follows:

- two groups of three astronauts who train and game once a week for half an hour (incl. procedure training and an entertainment game)
- the astronauts communicate via text chat only
- MECA collects information on crew condition (social, cognitive, affective) and performance (effectiveness and efficiency of operations during the training and gaming
- MECA provides (simple) feedback on crew condition and performance
- MECA provides a task scheduling and multimedia annotation tool

A first experiment with the ePartners provided data to train Bayesian Network models on Cognitive Task Load (CTL) and Emotional State (ES), and their effects on performance with Bayesian Networks. CTL specifies the task demands in terms of time occupied, task complexity and task switches; ES specifies the emotional responses in terms of valence and arousal.

The screen in Figure 5 shows MECA's feedback mechanism for performance, emotional state (arousal, valence, and dominance), cognitive task load (time occupied, level of information processing, and task-set switches). On the upper part of the window, the user can



Figure 5. The tool for giving feedback on crew condition and previous performance.

see all values of previous sessions. The data is presented in such a way that correspondences between these three aspects can easily be recognized (e.g. the combination of high cognitive task load and poor performance). This could give the user a better insight in his (or her) own functioning.

On the lower left part of the window, the user can see the scores of the current session in comparison with the average score of the previous session. On the lower right part of the window, the user can add annotations to the current session, using text, audio, video, or photos. This could also help to improve the user's memory, and awareness of his (or her) functioning over time.

CONCORDIA EXPERIMENT

Concordia station is a permanent international research facility high on the Antarctic ice cap. The objective of Concordia station is to operate as an international research facility to conduct scientific programmes. Concordia provides a unique test platform for MECA, because of its setting in an isolated extreme environment for long duration experiments with high demands for crew autonomy. In this setting, more prolonged or repeated usage of MECA services can be tested. We are performing another design iteration of the MECA core support elements for nominal and off-nominal situations and corresponding conditions (i.e., the appropriateness of his or her cognitive and affective responses), which need this type of prolonged evaluation in extreme and isolated environments. Based on our findings in previous projects we have identified the following issues as promising to be included in this refinement:

- Goal-based timeline tool: We plan to enhance the MECA timeline functionality by using persuasive technology. This allows the timeline tool to motivate the crew members to perform certain activities. This improves mental health, and satisfaction. For example, physical exercise trainings are known to have a positive effect on mental health of astronauts. However, when the astronaut feels obliged to do these trainings, this positive effect disappears. We believe that, by allowing users to specify their goals when they schedule their activities, they become more motivated to perform these activities. In the scheduling phase, MECA and the users set up their goals collaboratively. In the feedback phase, MECA relates the user's performance to his or her goals.
- Social games: To maintain a healthy social atmosphere among a small group of isolated crewmembers, we plan to develop new tools for monitoring and (if needed) repairing social relations. We start from social interaction games, where multiple players have mixed objectives and must collaborate with each other to achieve their goals. We will add new gaming concepts, such as joint resource supervision, and tit-for-tat collaboration, to allow better monitoring of social relations, and to improve potential conflicts between crewmembers.

- **Personalization:** MECA will adapt itself to the personal characteristics of its users. During the activity monitoring phase, much information is gained about how the crewmembers behave under which circumstances. During the scheduling phase, we plan to use this information for offering personalized support. For example, MECA advises changes on the planned schedule, optimizing the users' performance and resources usage by to postponing, cancelling reassigning or scheduling new activities.
- First aid COLT: In previous MECA projects, we have gained positive experiences with procedure-training functionality. In Concordia, we plan to extend this functionality to the first aid domain. This allows the non-medical crew members to learn medical procedures from the medical crew members.

CONCLUSION

This paper presented ongoing research on electronic partners that diagnose, guide and mediate space crew's social, cognitive and affective processes. We follow a situated Cognitive Engineering methodology in which ePartner's functions are being developed incrementally. In collaboration with their human counterparts, these ePartners sense and affect crew's performance, task load and emotional responses. First prototypes are being tested at MARS500 and Concordia, evaluating the support functions and providing data to train our models. First results show that cognitive task load and emotion have additional value for performance assessment, and that Bayesian Networks can be used to train the models (e.g., it is possible to deal with missing data, and easy to extend and refine).

ACKNOWLEDGEMENTS

MECA is a development funded by the European Space Agency (contract numbers 19149/05/NL/JA and

21947/08/NL/ST). Project partners are TNO Human Factors (NL; coordinator), Science & Technology BV (NL), OK-Systems (E) and EADS-Astrium (D).

- 1. Fogg, B.J., (2002). *Persuasive Technology: Using Computers to Change What We Think and Do.* Morgan Kaufmann, SanFrancisco, CA.
- Grootjen, M., Neerincx, M.A., Stolk, K.D., Weert, J.C.M. van & Bierman, E.P.B. Design and user evaluation of an interface prototype that adapts to the operator's cognitive task load. In: D.D. Schmorrow, D.M. Nicholson, J.M. Drexler & L.M. Reeves (Eds), Proc. 4th Intern. Augmented Cognition. Arlington, Virginia: Strategic Analysis, Inc (2007), 97-106.
- 3. http://www.crewassistant.com
- 4. Neerincx, M.A., Kennedie, S., Grootjen, F., and Grootjen, M. Modelling Cognitive Task Load and Emotion for Adaptive Maritime Interfaces. In: Lecture Notes in Artificial Intelligence. Schmorrow, D.D.; Estabrooke, I.V.; Grootjen, M. (Eds.), Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience. Proceedings of the 5th International Conference of the Augmented Cognition, Berlin/Heidelberg: Springer (2009), 260 – 269.
- Neerincx, M.A. Bos, A., Olmedo-Soler, A. Brauer, U. Breebaart, L., Smets, N., Lindenberg, J., Grant, T., Wolff, M. The Mission Execution Crew Assistant: Improving Human-Machine Team Resilience for Long Duration Missions. Proceedings of the 59th International Astronautical Congress (IAC2008), Paris, France: IAF. DVD: ISSN 1995-6258.
- 6. Picard, R.W. *Affective Computing*. MIT Press, Cambridge, MA, 1997.

Automatic Mental Heath Assistant: Monitoring and Measuring Nonverbal Behavior of the Crew During Long-Term Missions

Natalia Voynarovskaya

Eindhoven University of Technology User System Interaction n.voynarovskaya@tue.nl

Emilia Barakova

Eindhoven University of Technology Designed Intelligence Group e.i.barakova@tue.nl

ABSTRACT

This paper presents a method for monitoring the mental state of small isolated crews during long-term missions (such as space mission, polar expeditions, submarine crews, meteorological stations, and etc.) The research is done as a part of Automatic Mental Health Assistant (AMHA) project which aims to develop set of techniques for automatic measuring of intra- and inter- personal states in working groups. The method is focused on those aspects of psychological and sociological states that are crucial for the performance of the crew. In particular, we focus on measuring of emotional stress, initial signs of conflicts, trust, and ability to collaborate. The present research is performed in collaboration with MARS-500 experiment in which a small group of people is isolated for a long period of time. The MARS-500 experiment, in this way, provides a unique platform for study of human-human interaction. The confinement study will imitate all key peculiarities expected to be present during future missions to Mars (i.e. ultra long duration flight, need for autonomy, complicated communication with a digital communication center due to signal delay, and limited stock of expendables). The developed method is also currently tested by usage of a web-based platform.

Author Keywords

Social network analysis, nonverbal communication, emotions, long-term missions, evolutionary game theory, colored trails.

Roman Gorbunov

Eindhoven University of Technology Designed Intelligence Group r.gorbunov@tue.nl

Matthias Rauterberg

Eindhoven University of Technology Designed Intelligence Group g.w.m.rauterberg@tue.nl

INTRODUCTION

Long-term missions usually are performed by small group of people, and characterized by extreme physical and psychological parameters. Specific conditions like special confinement, small group living, extreme social isolation, risk, working together in close proximity [2], could have a crucial negative influence on the mental state of crew members as well as the social atmosphere in the whole crew. The interpersonal issues, such as crew tension, cohesion, leadership, language, cultural factors [1], could cause personal conflicts which could affect the performance of the mission. According to Russian psychologists [3], the main objectives of psychological countermeasures during space flight, are (1) physiological reconstruction of the informational environment, (2) correction of psycho-emotional area of astronauts, and (3) prevention of insufficiency in social contacts. For measuring, prediction and prevention of the above mentioned problems it is crucial to have methods for automatic monitoring of psychological and social states in the group. The psychological states of the crew members as well as different aspects of interpersonal relations have been assessed using different approaches. In the current study we combine three approaches: strategic multiplayer game, [8, 11], techniques for automatic monitoring of nonverbal behavior [17,18,21,22,23], and self assessment based on questionnaires. In this way we are aiming to overcome limitation of every component and develop a self consistent and comprehensive technique for monitoring intra- and inter-personal state of the crew members.

RELATED WORK

The first concept of the AMHA project has been done within Mars-500 experiment for 105 days isolation [4].

In 2004 the Institute for Biomedical Problems (IBMP) in Moscow and the European Space Agency have started to plan a full-scale ground based simulation of a manned

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

mission to Mars. Such a full scale mission requires between 520 to 700 days of isolation. Referring to the lower end of this time frame the initiative was named Mars-500.

A crew of six candidates (four Russians and two from EU countries) are sealed insight the facilities of the Institute for Biomedical Problems in Moscow. An initial 105-day isolation period took a place in spring of 2009. The present concept is going to be tested in the full 520-day study, which is planned to launch in the first quarter of 2010.

Game Theory and Colored Trails

In this project, we explore the use of strategic multi-player games to alleviate stress, and more importantly as an unobtrusive tool to monitor the mental capacity of astronauts as well as the development of different social interaction patterns within the crew. We are primarily interested in games that feature the following properties:

- Simple enough for analysis
- Rich enough to reflect features of real life interactions
- Grounded in a situated task domain
- Strategic (i.e. partial information that promotes reasoning)
- Suited to measure social factors such as fairness

We are also interesting to produce data, which directly involve interpersonal relation and, as a consequence, can be interesting to monitoring social atmosphere in the crew.

As such a tool, we used a three-player negotiation variation [25] of the Colored Trails framework developed at Havard University [24]. The Colored Trails game is played on a board of colored squares. One square is designed as a "goal square". Each player has a piece on the board and possesses a set of chips in colors chosen from the same palette as the squares of the board. To move own piece into an adjacent square a player must turn in a chip of the same color as the square. Chips can be exchanged by the players if a mutual agreement is reached. The goal of the game is to move own piece as close to the goal square as possible using as less chips as possible. Distance to the goal is more important than number of chips left after the move.

METHOD

Design of the presented monitoring system is based on our previous analysis of the data obtained in the MARS-105 experiment. In particular we analyzed behavior of Responders. Situation, in which a responder needed to make a decision, was described by two numbers: how the Responder can improve his/her payoff by accepting the first offer and how he/she can improve his payoff by accepting the second offer. This description of the situation is a simplification because other factors can influence the decision of the Responder (for example what other player will get as a result of his/her decision or how good are the available offers in comparison with the offers which could be proposed). However, the two selected parameters are the



Figure 1. Colored Trails Game

only factors, which are determining behavior of the Responder and showing if the player's behavior is totally rational. So, the mentioned two parameters can be used to check how rational the behavior of Responder is. This check is important because any deviation from the rational behavior can potentially be an indication of psychological preferences or other interpersonal relations.

On the Figure 2 is shown the behavior of one of the Responders. The x- and y-axis corresponds to the improvement of the payoff which can be achieved by accepting the first and second offers, respectively. So, generally speaking, every point in the plot represents a particular situation in which the given Responder needed to make a choice. The color of the point indicates the choice which was made. The red/blue colors mean that the first/second offer was accepted. The green color means that the both offers were rejected.

The space on the graph is divided into three different regions. If a responder is totally rational, every region should only contain points of one corresponding color. For examples, if both offers worsen the payoff of the responder (the left bottom region) than both offers should be rejected



Figure 2. Dependence of the behavior of one Responder on the accepted offer.

(green points). The right-bottom region means that the first offer is better than the second one and it improves the payoff of the responder (so, in this case the first offer should be accepted (red points)). In the third region (lefttop) the second offer should be accepted (blue points).

Generalization of the Colored Trails

For 520 days isolations study, we propose a generalization of the Colored Trails game. In the new version of the game every user plays as Proposer. Moreover, in contrast to the previous version of the game, Proposers are free to choose a player (Responder) whom he/she wants to offer a chips exchange. As a consequence of these modifications, two or three players will play the Responder role in the second stage of the game. The introduced modifications provide several advantages.

First, we increased number of Proposer per game (from 2 to 3) and, as a consequence, we get more data about the behavior of Proposers.

Second, we increased the number of Responders (from 1 to 2 or 3). This way more information about behavior of Responders can be collected.

Third, we increased the variety of situations in which Responders can be. In particular, Responders have to choose from different number of offers coming from Proposers which are in different situations.

Forth, based on the analysis of the previous experiment, we added a phase at the game, which is aimed to assess irrational preferences of the players. At this stage, when each player is choosing a partner for a current game, player's behavior can not be based on rational thinking about state of the game, and provide explicit unconscientious choice of the partner.

In addition to the previous experiment we combine the usage of the negotiation game with direct monitoring of the nonverbal behavior of the subjects. In that way we want to study if our conclusions about user's persistence of the game and other player can be confirmed by the other kinds of measurements. In particular, the monitoring techniques can be used to detect if players are satisfied or not with offers from other players as well as decisions made by the Responders.

Design Concept

To build the AMHA concept model, firs of all, we defined the parameters, which are crucial for successful communication and collaboration, such as stress level, pleasure, trust, dominance, and etc. Furthermore, we defined a personal mental state and interpersonal relationships as two main components of the crew mental health. The combination of different techniques is required to assess different aspects of the psychosocial states as well as to perform cross validation and correct interpretation of the collected data (see Figure 3).



Figure 3. AMHA concept model.

To assess the emotional parameters during the game, players will be asked to fill in the questionnaires.

Usage of questionnaires for a self assessment is important component of our design, since crew members can provide explicit information about their psychological states and interpersonal relations in the crew. This information is important for interpretation of the game results in terms of perception of psychosocial state of the crew.

Self Assessment Techniques

Self assessment is an important component of mental health assistant since subjects can explicitly provide needed information about intra- and inter- personal states.

Big 5 Test

One of the intentions of our experiment is to find a relation between observed dynamics of the interpersonal relations with psychological parameters of the crew members. Such a relation could help us to generalize the behavior observed in particular crew and to predict, in this way, relations in groups consisting of members of similar psychological types.

In this sense, at the baseline of the experiment, we want to get data, which could give us the psychological characteristics of participants about personality and collaboration patterns at the same time. This data we can use as a reference point in the future data analysis. For our purpose we are using the Big 5 Test [10,12].

The Big Five model is considered to be one of the most comprehensive, empirical, data-driven research findings in the history of personality psychology. Over three or four decades of research, these five broad factors were gradually discovered and defined by several independent sets of researchers.



Figure 4. SAM test.

Self Assessment Manikin (SAM) Test

To assess emotional state of the participants, we are using the Self-Assessment Manikin test [13,14]. It offers the ability to avoid the verbal expression of the emotion assessment, so it establishes a quick and easy to use experimental procedure.

The test assesses the follow emotional states: "Pleasure", "Arousal" and "Dominance". In our study we used the initial verbal expressions of the three dimensions.

Usage of this technique has few advantages. The test is well established, it's aimed for the main and basic parameters, but the results are full enough for analysis. And it's simple, quick and intuitive to perform, which is crucial for our design.

Measuring Nonverbal Behavior

Automatic monitoring of face expression, voice intonation and body movements were proven to be useful techniques for monitoring of emotional states of subjects. These techniques are important since humans naturally express their emotions throw nonverbal behavior.

In our study we use records of facial expressions during the people are playing the Color Trails Game. Facial expressions give important clues about emotions. In contrast to self assessment techniques, the tracking of nonverbal behavior provides information about emotions and psychological state unconscious. It would allows us to get more objective data compare to results of questionnaires.

Using the software for analysis video (like Visual Recording by University of Amsterdam), we would be able to put in numbers 6 main emotions [9].

Social Network Analysis

To describe the relation between participants and analyze how the relations are developing through the time, we use Social Network Analysis (SNA). SNA defines networks as sets of ties linking several individuals. Ties or contacts may be of different kinds, formal or informal, frequent or infrequent, affect-laden or purely utilitarian. Network analysts use the terms "transactional content" or "tie type" to identify the type of exchange or relationship that takes place between actors in a network [15]. SNA approach allow not only to draw a 'map of ties' between structures/ institutions/ societies as macro-level of social reality, but, in addition, to track changes that occur with these 'actors' during social transformations in the contexts of interconnections at the micro-level.

DISCUSSION

In connection with the Mission Execution Crew Assistant (MECA) by TNO [19, 20], we have an unique possibility to test our design within MARS- 500 experiment. MARS-500 provides a unique test platform, because of its setting in which a small crew is isolated for a long duration to simulate a manned Mars mission. In this setting, more prolonged or repeated usage of AMHA can be tested. In the MARS-500 program, we will select a small set of core elements of AMHA that need this type of prolonged evaluation: a diagnostic method to measure psychosocial crew status and a feedback mechanism. The objective of this evaluation is to improve the requirements baseline and its design rationale for these elements, and to refine the corresponding models and methods.

To get quantitative data, we are running an additional webbased study. For this purpose the web-site with Colored Trails Game has been developed.

The research questions for this experiment, apart to get quantitative data, are:

- how the CT game can develop the patterns of existing relationships between people;
- how the CT game can create the relationships between people;
- how the CT game can monitor the behavior.

In these settings we are planning:

- 5-6 groups (by 3 person each) people who know each other
- 5-6 groups (by 3 person each) people who don't know each other
- About 15-20 groups of three people, which are know each other well and/or working together.
- Three sessions per week.
- The duration of the experiment is 15-18 sessions (5-6 weeks).

By the end of the experimental part of the project, we would be able to do data analysis. Two different experimental settings provide an opportunity to cross a validation and answer the question of transferability of results obtained with a single experiment with fixed constrains.

- Kanas, N., Salnitsky, V., Grund, E., Gushin, V. Social and cultural issues during Shuttle/Mir space missions. *Acta Astronautica*, 47, 2-9 (2000), 647-655.
- Manzey, D. Human missions to Mars: new psychological challenges and research issues. *Acta Astronautica*, 55 (2004), 781-890.
- 3. Gushin, V. Psychological countermeasures during space missions: russian experience. *Journal of Gravitational Physiology*, 9, 1 (2002), 311-312.
- Hennes, D., Tuyls, K., Neerincx, M.A., Rauterberg, G.W.M. Micro-Scale Social Network Analysis for Ultra-Long Space Flights. IJCAI-09/SMC-IT-09/IWPSS-09 Workshop on Artificial Intelligence in Space, 2009
- Hennes, D., Tuyls, K., Rauterberg, G.W.M. State-Coupled Replicator Dynamics. 8th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2009), 789-796.
- Hennes, D., Tuyls, K., Rauterberg, G.W.M. Formalizing Multi-state Learning Dynamics. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (IAT 2008) vol. 2, 266-272.
- Rauterberg, G.W.M., Neerincx, M.A., Tuyls, K., Loop, J. Entertainment computing in the Orbit. *New Frontiers for Entertainment Computing*, 279 (2008), 59-70.
- Gaal, Y., Grosz, B.J. Colored Trails: A Formalism for Investigating Decision-making in Strategic Environments, 2005.
- 9. Valenti, R., Sebe, N., Gever, T. Facial Expression Recognition: a Fully integrated approach, 2007.
- 10. John, O.P., Srivastava, S.: The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives, 2005.
- 11. Camerer, C.F. Behavioral Game Theory: Thinking, Learning, and Teaching, 2001.
- 12. Harvey, R.J., Murry, W.D., Markham S.E. A "Big Five" Scoring System for the Myers-Briggs Type Indicator, 1995.
- Bradley, M.M., Lang, P.J. Measuring Emotions: the Self-Assessment Manikin and the Semantic Differential. *J. Behav. Ther. and Exp.Psychiat.*, 25, I (1994), 49-59.

- Tsonos, D., Kouroupetroglou, G. A Methodology for the Extraction of Reader's Emotional State Triggered from Text Typography. *Tools in Artificial Intelligence* (2008), 139-154,
- 15. Nelson, E.. The Strength of Strong Ties: Social Networks and Intergroup Conflict in Organizations. *The Academy of Management Journal*, *32*, 2 (1989), 377-401.
- 16. Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan,S. Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information, 2004
- 17. Mase K. Recognition of facial expression from optical flow. *IEICE Transc.*, E. 74, 10 (1991), 3474–3483.
- Yacoob, Y., Davis, L. Computing spatio-temporal representations of human faces. Computer Vision and Pattern Recognition. *Proceedings CVPR '94, IEEE Computer Society Conference* (1994), 70–75.
- Neerincx, M.A., Lindenberg, J., Smets, N. The Mission Execution Crew Assistant: Improving Human-Machine Team Resilience for Long Duration Missions, 1998.
- 20. Breebaart, L., Bos, A., Grant, T., Soler, A.O., Neerincx, M.A., Lindenberg, J., Smets, N., Brauer, U., Wolff, M. The MECA Project — Ontology-based Data Portability for Space Mission Operations, 2008.
- Barakova, E., Lourens, T. Expressing and interpreting emotional movements in social games with robots. Personal and Ubiquitous Computing, 2010.
- 22. Lourens, T., Barakova, E. Retrieving Emotion from Motion Analysis: In a Real Time Parallel Framework for Robots. In C.S. Leung, M. Lee, and J.H. Chan (Eds.): *ICONIP 2009*, Part II, LNCS 5864, 430–438.
- 23. Lourens, T., Barakova, E. My Sparring Partner is a Humanoid Robot -A parallel framework for improving social skills by imitation. In J. R. Alvarez, editor, *IWINAC 2009*, number 5602 in Lecture Notes in Computer Science, 344-352.
- 24. Gal, Y., Grosz, B. J., Kraus, S., Pfeffer, A., Shieber, S. Colored trails: a formalism for investigating decisionmaking in strategic environments. In *Proc. of the IJCAI Workshop on Reasoning, Representation, and Learning in Computer Games*, 2005.
- 25. Ficici, S., Pfeffer, A. Modeling how humans reason about others with partial information, 2008.

Symposium: The Significance of Voluntary Exploration to Monitor Emotional Behavior of Rodents

Oliver Stiedl

Behavioral Neuroscience Group, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands oliver.stiedl@cncr.vu.nl

ABSTRACT

In this symposium we will provide examples from research approaches in rodents demonstrating the importance of voluntary choice and decision-making under baseline stress-free conditions for emotional behavior. These novel approaches allow for substantially improved interpretation of performance changes as opposed to currently used standardized behavior tests. Emotion cannot be measure directly but instead has to be inferred indirectly from behavioral and concomitant physiological (e.g. neural, autonomic) adjustments. Standard anxiety tests and fear learning tasks are characterized by human interference that affect emotional behavior and may bias the experimental outcome. This is particularly relevant for tests that depend on locomotor activity-derived measures as index of fear and anxiety. In response to this dilemma, we will discuss current methodological pitfalls and provide insights into our concepts and used methodological approaches including choice behavior with multiple measures extending to aspects of the organization of behavior (actualgenese: the moment-to-moment dynamics of behavior). Thus, human interference is largely excluded in these experiments while the duration of experiments is extended to several days. These novel approaches follow the main features of natural exploratory behavior starting out from a safe home base and driven by the motivation to investigate novel areas. The motivation for exploration is generally the drive to find resources or reproductive needs while facing the risk of the unknown e.g. potential predators. Exploration implies novelty detection and thus a discrimination from what is known and what is new. The distinction between exploration and cognition is not sharp. The same can be

said for the distinction between for instance anxiety and exploration. Thus, home cage observations that involve the activation of different motivational systems are an appropriate setting for assessing the integration of different motivational systems.

SYMPOSIUM CONTENTS

Fear Conditioning in an Automated Home Cage (DualCage) Environment

Oliver Stiedl Anton W. Pieneman & René F. Jansen (VU University, Amsterdam, The Netherlands), Christian Gutzen & Stephan Schwarzer (Biobserve, Germany).

Spontaneous Behavior in The Home Cage: A New Test Environment for Measuring Neuropathic Pain

Berry Spruijt, Raymond C. de Heer & Johanneke E. van der Harst (Delta Phenomics, The Netherlands)

An Automated Maze for Studying Working Memory and Decision-Making in Rodents

Jeansok Kim (University of Washington).

Genetic Dissection of Motor Activity Levels and Avoidance Behavior in The Home Cage; Translational Phenotypes for Mood Disorders

Martien Kas & Annetrude (J.G.) de Mooij-van Malsen (UMC Utrecht, The Netherlands), Berend Olivier (Utrecht University, The Netherlands).

Understanding Exploratory Behavior Step by Step

Ilan Golani, Ehud Fonio & Yoav Benjamini (Tel Aviv University, Israel).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Fear Conditioning in an Automated Home Cage (DualCage) Environment

Oliver Stiedl

oliver.stiedl@cncr.vu.nl

Anton W. Pieneman

Behavioral Neuroscience Group, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, The Netherlands

Christian Gutzen

info@biobserve.com

Stephan Schwarzer

Biobserve GmbH, Koenigswinterer Str. 418, 53227 Bonn, Germany

René F. Jansen

Behavioral Neuroscience Group, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, The Netherlands

Fear conditioning is an important test in behavioral neuroscience to investigate the neural systems and molecular basis of various aspects of emotional learning across a wide range of species [1]. Dysfunction of the fear circuits is assumed to underlie mechanisms of affective disorders and is frequently investigated in rodent models. Traditional fear conditioning methods require frequent handling of the animals which compromises the accuracy with which fearrelated parameters such as the behavioral response and autonomic responses such as heart rate of the animal can be measured. We therefore developed a novel fully automated DualCage environment to explore the dynamics of fear conditioning during spontaneous/voluntary behavior of male C57BL/6J mice.

The DualCage environment consists of a home cage attached to one or more additional cages separated from another by automated sliding doors. The sliding doors can be operated automatically, e.g. by the position of an animal, that is monitored by a camera tracking system (Viewer© software, Biobserve, Bonn, Germany). Each cage has an area of 24×30 cm thereby doubling the area offered in a type II cage. In the case of fear conditioning the attached

cage contains a shock grid and a loudspeaker. The animal is allowed to voluntarily enter the shock cage after the sliding door has been opened. As soon as the animal has fully entered the shock cage, the door is closed. Initially the mouse is allowed to explore the shock cage before a series of tone/shock stimuli are used to condition the mouse to the tone and the shock cage (context). After this training episode, the door is opened and the mouse can return to its home cage.

Retention tests are performed at different time points after the training session to explore the dynamics of memory formation and its expression. Tone retention tests can be performed in the home cage or in the shock cage. The latter obviously works only if and when the conditioned mouse voluntarily returns to the shock cage. Note that during the entire period of habituation, training and testing the mouse is not handled and allowed to show spontaneous exploratory behavior. The spontaneous behavior is characterized by risk assessment when access to the conditioning cage is granted by peeking through the door in stretch-attend posture before eventually revisiting the shock cage. Revisits of the shock cage eventually occur despite the lack of e.g. rewarding reinforcement indicating the naturalistic drive of mice to explore their environment if given a chance despite previous negative experiences. This behavior parallels the one observed in rats in a visible burrow system [2]. A separation between home cage and shock cage is necessary to exploit the novelty-seeking behavior of mice at distinct times after training, i.e. to investigate short- and long-term memory.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

This novel approach mimics naturalistic-like behavior that is characterized by trade-offs between avoidance and curiosity that can now be quantified on long time scales based on a voluntary choice of mice. Extracellular electrical activity in the brain as well as the heart rate of the animal are measured by wireless telemetry systems with signal receivers that are an integral part of the DualCage system. The DualCage system is designed to integrate any commercially available hardware into the system for multi-purpose use depending on the scientific aims. This approach with a separation of home cage and versatile add-on components will allow to perform a spectrum of diverse tests with full choice (voluntary decision) of the experimental animal and minimal unspecific interference by the experimenter. The importance of voluntary choice for the temporal organization of behavior has recently been demonstrated in open field behavior in mice [3]. Behavioral measures can now be complemented to autonomic [4] and neural measures of conditioned contextual fear without the adverse consequences of handling [5]. For the integration of electrophysiological measures see our presentation entitled "Extracellular Multi Unit Recording in Fear Conditioning in Mice Using a Telemetry Approach in an Automated Home Cage (DualCage) Environment" at this Measuring Behavior Meeting 2010.

Author Keywords

Behavior tracking, fear, anxiety, novelty, exploration, motivation.

- 1. LeDoux, J.E. Emotion circuits in the Brain. Annu. Rev Neurosci., 23 (2000), 155-184.
- Blanchard, R.J., Blanchard, D.C. Antipredator defensive behaviors in a visible burrow system. J. Comp. Psychol., 103 (1989), 70-82.
- Fonio, E., Benjamini, Y., Golani, I. Freedom of movement and the stability of its unfolding in free exploration of mice. *Proc. Natl. Acad. Sci. USA*, 106 (2009), 21335-21340.
- Stiedl, O., Jansen, R.F., Pieneman, A.W., Ögren, S.O., Meyer, M. Assessing emotional states through the heart in mice: implications for cardiovascular dysregulation in affective disorders. *Neurosci. Biobehav. Rev.*, 33 (2009), 181-190.
- 5. Stiedl, O., Tovote, P., Ögren, S.O., Meyer, M. Behavioral and autonomic dynamics during contextual fear conditioning in mice. *Auton Neurosci*, *115* (2004), 15-27.

Spontaneous Behavior in the Home Cage: a New Test Environment for Measuring Neuropathic Pain

Berry M. Spruijt

Raymond C. de Heer

Johanneke E. van der Harst

DeltaPhenomics BV, P.O. Box 80086, 3508 TB Utrecht, The Netherlands

INTRODUCTION

A sensitive animal model requires: 1) the distinction between novelty-induced and base line behavior, 2) the exclusion of confounding factors such as human handling, the continuous presence or preceding stressors; 3) the disentanglement of distinct motivational systems such as exploration, cognition involved in complex behavior. Thus, the activation of various motivational systems must be possible in a test situation allowing assessment of a distinct contribution of each of them to certain parameters.

A very good example of allowing the animal to display emotional behavioral expressions and to detect interactions between emotional states, for instance pain and anxiety,- is the monitoring of the development of neuropathic pain measured in the home cage.

Neuropathic pain is usually indirectly measured by applying temperature (hot, cold stimuli > allodynia) or mechanical stimuli and assessing the animal's hypersensitivity to those stimuli.

In those tests involving handling, novelty and exposure to a challenging stimulus, stress as a result of the mechanical or temperature stimulus may inhibit sensory input and, thus, the animal's perception of and/or expression of neuropathic pain.

We now monitor the spontaneous behaviors of animals developing neuropathic pain induced using a standard surgical procedure (constriction of the *N. Ischiadicus*).

The rationale is that animals suffering from pain will adapt their voluntarily movements to avoid pain and that pain influences behavioral patterns over time and develop behavioral changes as a result on the interaction of pain perception and other emotional expressions, i.e., anxiety. We expect that animals meet their primary demands for food, water and exploration, but that they will modify their spontaneous behaviors to avoid the use of the painful body part and due to the pain seek more shelter (safety, sleep, rest).

Thus, the home cage may allow the detection of a slowly developing chronic pain based on spontaneous reactions of the animal, most probably a direct consequence of the neuropathy and not an indirect reaction via hypersensitivity.

We aim to:

- Measure (the development of) symptoms of (chronic) pain by continuous automated monitoring of detailed potential changes in spontaneous behavior in the home cage after surgery that inflicts chronic neuropathic pain.
- Assess whether chronic pain affects other emotional states such as anxiety by measuring a possibly sensitized behavioral responses of animals having a constricted nerve in the hind paw to an aversive (light) stimulus in the home cage.
- Controlling the efficacy of the protocol used by assessing pain in a well known standard test to measure mechanical hypersensitivity, using the von Frey method.

We expect:

- to find changes in the distribution of time spent on various behavioral activities, such as time spent sleeping/sheltering, eating, exploring;
- that pain affects other emotional states such as anxiety, and thus, that animals subjected to the constriction injury will show altered behavioral responses to an anxiogenic stimulus.

METHODS

Subjects were twenty-four rats (Wistar, CharlesRiver Germany), that were housed socially in pairs in a standard macrolon cage containing bedding, nesting material and tubes. Water and food were available ad libitum. The animals were housed under a regular 12/12 day/night rhythm (light period: 6:00-18:00; dark period: 18:00-6:00) and were allowed to adjust to housing and management procedures for 2.5 weeks before the start of the experiment. The animals were randomly assigned to the CCI-group (n=12), or Control-group (n=12).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

| Day | Measurements | | | |
|-----|---|--|--|--|
| -4 | START PhenoTyper [24hr behavioral monitoring] | | | |
| -4 | continuous Behavioral monitoring | | | |
| 0 | | | | |
| 0 | SURGERY: Chronic constriction injury (CCI) at the sciatic nerve | | | |
| 0 | continuous Behavioral monitoring | | | |
| 7 | | | | |
| 7 | weighing, cage cleaning | | | |
| 7 | continuous Behavioral monitoring | | | |
| 14 | | | | |
| 14 | weighing, cage cleaning | | | |
| 14 | continuous Behavioral monitoring | | | |
| 21 | | | | |
| 22 | ANXIETY-test: Light spot on Food hopper during first 3hrs of dark period | | | |
| 24 | PAIN-test, mechanical hyper-sensitivity: Von Frey | | | |
| 25 | | | | |
| 26 | PAIN-test, mechanical nyper-sensitivity: Von Frey, 1 | | | |
| | PAIN-test, mechanical hyper-sensitivity: Von Frey, 2 | | | |

Table 1. Time schedule of the experimental protocol.

Phenotyper Anxiety Test: Light Spot At Day 22 on Feeding Area

After the start of the dark (i.e. active) period, during which rats normally spend a lot of time on feeding, a bright light spot (500 lux) was focused on the food trough for 3 hours. Parameters such as latency to enter the illuminated food zone, frequency of visits and time spent in this zone can be indicative of an aversive response and related to the level of anxiety. After the light spot has been switched off the latency to return and number of frequencies to this previously illuminated zone can be indicative of a longterm effect or sensitization.

Measurement of Mechanical Hypersensitivity

Mechanical hypersensitivity was measured using a set of graded von Frey Filaments (Somedic, Sweden). The animals are placed in a Plexiglas chamber (16x24x14 cm) with a grate bottom and left to adapt for at least 15 min; the animals have been habituated to the transport procedure and boxes (app. 20-30 min) the day before. The test is conducted during the light (inactive) phase between 9:30 and 13:00 hr. Mechanical stimuli are generated by touching the plantar region of the animals hind paw with von Frey filaments (glass fiber) of differential thickness, causing different levels of pressure.

The filament is pressed against the hind paw until it bends and is held steady for 5-6 seconds, scoring if the animal responds or not. The animals were measured randomly three times on the operated (left) paw. The measurements on the left paw of the non-operated control group serves as a control.

RESULTS

The results show acute consequences of the constriction injury of the sciatic nerve (CCI) on various parameters during the first week after surgery: e.g. Time spent on eating, distance moved, velocity and frequency of moving.

The following weeks more long-term effects of the constriction, and presumably indicating chronic pain, were noticed on how the animals spent their time and explored their home cage. In this abstract we limit ourselves to the effects of CCI on the time spent near the feeder and anxiety.

Time Spent Near Feeder: Increase

During the active period of their day CCI animals spent more time in the vicinity of the feeder (Figure 1). Body weight did not change over the weeks between the groups.



Figure 1. On the y-axis time spent in the feeding area is displayed per hour. The X-axis represents time over 7 days.

CCI animals showed a longer latency (Figure 2) before entering the food-area during the test and spent less time in that area after the test indicating a prolonged aversive effect or even sensitization to an aversive stimulus in CCIanimals.



Figure 2. Latency to enter the illuminated zone.

In addition, if the time spent near the feeder during the LightSpot is compared per animal to the time spent in that zone 24 hrs before (when light was off), a decrease in time spent near feeder is seen (data not shown).

At the end of the home cage-monitoring period, the von Frey test clearly verified the presence of neuropathic pain as assessed by this standard method.

DISCUSSION

It seems that the animals were capable of displaying all kinds of movements, and behaved differently in response to the feeder and an aversive light beam. Other parameters not shown here such as distance moved, velocity and time spent on shelter, immediately after surgery and after 3 weeks changed as well and are now statistically being analyzed and will be presented. Effects of pain on anxiety measured in the home cage have been demonstrated, which seems indicative of a change at the emotional level in the animals with neuropathic pain. The effects of the aversive stimulus on anxiety seems to be long lasting as these animals still showed a change in behavior (time spent in the previously lighted area) the next day. The effects of pain on a sensitized anxiety response underline the notion that the behavioral changes are due to the experience of pain and not a mere nociceptive response.

Notwithstanding the fact that the study was limited to 4 weeks, whereas it might be that a longer time-window and a more in-depth analysis is necessary to detect symptoms of (developing) chronic pain, the results are promising and opens avenues for assessing symptoms of neuropathic pain in the home cage.

An Automated Maze for Studying Working Memory and Decision-Making in Rodents

Jeansok J. Kim University of Washington, USA jeansokk@uw.edu

In rodents, the benchmark test for prefrontal cortex (PFC)mediated working memory is a delayed alternation task utilizing variations of T-maze or Figure-8 maze, which requires the animals to make specific arm entries to secure reward. Often, manual procedures are involved in shaping target behavior, imposing delays between trials, and delivering rewards, which can potentially influence the animal's behavior on the maze. Our laboratory developed an automated Figure-8 maze which does not require experimenter-animal interaction during shaping, training or testing. This system incorporates a computer vision system for tracking the animal on the maze, pneumatic air cylindercontrolled gates to impose delays, and automated reward delivery. The maze is controlled by custom software that records the animal's location and activates the gate according to the animal's behavior and a control algorithm. The program performs calculations of task accuracy, tracks movement sequence through the maze, and provides other measures such as running speed, time spent in different maze locations, and activity level during delay. This maze system can also be used to investigate decision-making in rodents by altering the reward volume, reward probability, reward delay, and reward effort.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Genetic Dissection of Motor Activity Levels and Avoidance Behavior in The Home Cage; Translational Phenotypes for Mood Disorders

Martien J.H. Kas

Rudolf Magnus Institute of Neuroscience, UMC Utrecht, The Netherlands m.j.h.kas@umcutrecht.nl **Berend Olivier**

Department of Psychopharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, The Netherlands B.Olivier@uu.nl

Annetrude (J.G.) de Mooij-van Malsen

Rudolf Magnus Institute of Neuroscience, UMC Utrecht, The Netherlands j.g.demooij@gmail.com

ABSTRACT

Identifying susceptibility genes for endophenotypes by studying analogous behaviors across species is an important strategy for understanding the pathophysiology underlying psychiatric disorders. This approach provides novel biological pathways plus validated animal models critical for selective drug development. One such endophenotype is avoidance behavior. In the present study, novel automated registration methods for longitudinal behavioral assessment in home cages are used to screen a panel of recently generated mouse chromosome substitution strains that are very powerful in quantitative trait loci (QTL) detection of complex traits. In this way, we identified chromosomes regulating avoidance behavior (increased sheltering preference) independent of motor activity levels (horizontal distance moved). Genetic information from the mouse QTL-interval was integrated with that from the

homologous human linkage region for a mood disorder.

We genetically mapped a QTL for avoidance behavior on mouse chromosome 15, homologous with a human genome region (8q24) linked to bipolar disorder. Integrating the syntenic mouse QTL-interval with genotypes of 1868 BPD cases versus 14,311 control subjects revealed two associated genes (ADCY8 and KCNQ3). Adenylyl cyclase 8 (Adcy8) was differentially expressed in specific brain regions of mouse strains that differ in avoidance behavior levels. Finally, we showed that chronic infusion of the human mood stabilizer carbamazepine (that acts via adenylyl cyclase activity) significantly reduced mouse avoidance behavior, providing a further link between human mood disorders and this mouse home cage behavior. Our data suggest that *Adcy8* might encode a translational behavioral endophenotype of bipolar disorder [1].

Author Keywords

Animal model, chromosome substitution strains, endophenotype, home cage environment, mood disorder, psychiatric disorders.

INTRODUCTION

Family and twin studies have revealed that genetic factors play a major role in psychiatric disorders, however, attempts to find susceptibility genes for these complex disorders have been largely unsuccessful. Therefore, new research strategies are required to tackle the complex interactions of genes, developmental, and environmental events. Recently, we have proposed a behavioral domain concept that focuses on the genetics of behavioral domains relevant to both animal behavior and across human psychiatric disorders [2]. We believe that interspecies trait genetics rather than complex syndrome genetics will optimize genotype-phenotype relationships for psychiatric disorders and facilitate the identification of biological substrates underlying these disorders. The development of automated paradigms that address these behavioral domains is a crucial step in this translational research field.

NOVEL VIEWS AND METHODOLOGIES

Identification of novel genetic loci in animal models for neurobehavioral traits relevant to psychiatric disorders relies on the fact that these traits are truly translatable across species. Once found, one can apply sensitive genetic strategies to these traits in order to unravel the underlying mechanisms. With the availability of a large variety of inbred mouse strains and their recently known genome sequences, mouse genetics offer a challenging way to study complex behavioural traits. For example, in contrast to patient populations, mouse strains can be used to control for phenotypic and genetic heterogeneity as well as for complex gene-environment interactions. Interestingly,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

recent studies have shown that genetic variation associated with psychiatric disorders affect analogous neural circuits and behavioural traits in mice and men, demonstrating that mouse models can contribute to systematic searches for genetic determinants of psychiatric disorders [3].

In general, rodent species have an innate preference for sheltered places that have lower light intensities than the outside-world and that provide a sense of safety via body contact with the shelter area surface (thigmotaxis). The assessment of this anxiety-related behavior is generally measured in relative short-lasting laboratory tasks and is highly dependent on strain differences in motor activity levels. In light of this, we have recently designed an automated home cage environment to assess separate behavioral domains over the 24-h day [4, 5, 6]. For instance, a hungry organism searching for food depends on an efficient exploration strategy in which finding the food resource in an appropriate period of time needs to be balanced against the risk of being exposed to potentially threats, such as predators. Thus, exploration for food relies on a balance between movement throughout the environment and avoidance behavior. To assess these behavioral domains as a function of time of day, a home cage environment for mice was designed with a sheltered and a non-sheltered feeding platform that would allow dissociation of the preference for shelter during feeding and for motor activity levels over several days and with minimal human disturbance [6].

For mouse genetics of these behavioral domains, there is an increasing appreciation of the properties of the set of mouse inbred strains which have been established over the last century of mouse genetics. Data are accumulating on each of this diverse collection of over 500 strains allowing strains to be chosen that cover a range of phenotypic variation in whatever phenotype is of interest. Traditionally, such strain combinations would be used to set up a cross or segregating population for genetic mapping purposes. More recently, Genetic Reference Populations (GRPs) with more optimal genetic properties are available or under construction. The prototype is the Recombinant Inbred (RI) panel which is generated from a cross between two inbred strains followed by an F1 intercross and 20 generations of inbreeding. The best characterized mouse RI panel, derived from C57BL/6J and DBA/2J (BXD) strains, has been a workhorse of behaviour genetics since the early 1990s. The BXD panel, of (until recently) 35 lines, gives only a coarse genetic resolution. However, several aspects of this picture have recently changed. One is the idea of treating transcript abundance and protein abundance or modification as phenotypes in their own right. This kind of geneticalgenomics is an extremely promising way of examining networks of function. Although it is possible with conventional genetic crosses or outbred populations, using (effectively immortal) inbred GRPs allows much more value to be extracted from each data set. This is also true of other phenotypic data, and there is a renewed interest in the

importance of accumulating data from many investigators. It is now possible to not only genetically map Quantitative Trait Loci (QTLs) on the BXD RI panel (recently expanded to 80 lines), but also to correlate new data with a large database of phenotypic data including gene expression data on several tissues. This is a major aid to positional cloning projects and multivariate analysis approaches currently being explored offer a way to assign some idea of function to many of the genes whose function is currently unknown. Other genetic reference populations have been developed. For example, chromosome substitution (consomic) panels have been generated for two strain combinations. This is done by repeated backcrossing to produce strains each with a single chromosome of one strain on the background of another. This is attractively simple to analyze and offers the simplification of multi-locus traits with only a single chromosome segregating [7].

RESULTS AND CONCLUSIONS

By testing a panel of 21 chromosome substitution strains in a wide variety of traditional and in an automated home cage environment, we have shown that behavioral components can be genetically dissociated. For example, we have shown that motor activity levels are under different genetic control than the preference to shelter by using a novel automated home cage task [6]. Further genetic mapping of these chromosomal regions revealed genetic loci that are syntenic with human linkage regions for mood disorders. Candidate gene selection within the QTL-intervals can nowadays be facilitated by combining quantitative phenotypic data from inbred mouse strains with their online available genome sequences. Single Nucleotide Polymorphism (SNP) databases provides gene-by-gene, SNP-by-SNP distribution patterns for various inbred lines and allow pinpointing SNP's in the QTL-regions that are associated with the behavioral trait of interest. Furthermore, biological pathway analysis can further be applied to provide additional candidate gene information. Subsequently, homologous candidate genes can then be tested in DNA samples from well-characterized psychiatric patient populations. In this way, interspecies genetics offers a great opportunity to translate essential behavioral traits in animals to human psychiatric disorders and to further understand the mechanisms underlying these traits. The automated home cage environments provides a modular system for the development of new longitudinal behavioral paradigms with translational value to disease.

- de Mooij-van Malsen AJ, van Lith HA, Oppelaar H, Hendriks J, de Wit M, Kostrzewa E, Breen G, Collier DA, Olivier B, Kas MJ. Interspecies trait genetics reveals association of Adcy8 with mouse avoidance behavior and a human mood disorder. *Biol Psychiatry*, 66, 12 (2009), 1123-30.
- 2. Kas M.J., Fernandes C., Schalkwyk L.C., and Collier D.A. Genetics of behavioural domains across the

neuropsychiatric spectrum; of mice and men. *Molecular Psychiatry*, *12*, 4 (2007), 324-30.

- Chen Z.Y., Jing D., Bath K.G., Ieraci A., Khan T., Siao C.J., Herrera D.G., Toth M., Yang C., McEwen B.S., Hempstead B.L., Lee F.S. Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behaviour. *Science*, *314*, 5796 (2006), 140-3.
- 4. Kas MJ, Van Ree JM. Dissecting complex behaviours in the post-genomic era. *Trends Neurosci.* 27, 7 (2004), 366-369.
- 5. De Visser L, van den Bos R, Kuurman WW, Kas MJ, Spruijt BM. Novel approach to the behavioural

characterization of inbred mice: automated home cage observations. *Genes Brain Behav.* 5, 6 (2006), 458-466.

- Kas M.J., de Mooij-van Malsen J.G., Olivier B., Spruijt B.M., Van Ree J.M. Differential genetic regulation of motor activity and anxiety-related behaviors in mice using an automated home cage task. *Behavioral Neuroscience*, 122, 4 (2008), 769-776.
- 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. Genetic dissection of complex traits with chromosome substitution strains of mice. *Science*. 304, 5669 (2004), 445-448.

Understanding Exploratory Behavior Step by Step

Ilan Golani ilan99@post.tau.ac.il Ehud Fonio ehudfo@gmail.com

Department of Zoology, Tel Aviv University, Tel Aviv 69978, Israel Yoav Benjamini Department of Staistics, Tel Aviv University, Tel Aviv 69978, Israel

ybenja@tau.ac.il

Exploration has been studied in rodents for almost a century. The measures used to characterize it are, however, limited in exposing the dynamics of the exploratory process, leaving the morphogenesis of its structure and meaning hidden. By freeing mice from the constraints imposed by forced introduction into the testing environment and from the confines of small cage and short session we reveal its meaning in the operational world of the mouse. In a new paradigm developed by us - The Dimensionality Emergence Assay (DIEM) - the setup consists of a large arena including a free passage between arena and home cage, and a long session. In this setup exploration consists of sequences of repeated motion. In a novel environment, these sequences show a gradual build up in extent and in dimensionality and complexity. Using advanced computational and exploratory data analysis tools we measure various aspects of the motions (e.g. distance traveled per motion, maximal speed per motion), plot them side-by-side, and compute the dynamics of their growth (actual genesis), as well as their interdependence within the overall developmental sequence (see Figure 1 on the next page).

The measurements provide a freedom of movement metric for the investigation of heretofore evasive issues like arousal and cognition, and for the assessment of drug action in animal models of human diseases of the CNS.

Our general approach to the quantification of behavior has been recently reviewed in a paper titled "Ten ways to improve the description of animal movement" [1]. In the present talk we will emphasize 2 out of the ten recommendations: i) the use of behavioral gradients to uncover animal-centered measures of behavior (the actual genesis of exploratory behavior provides an opportunity for a kind of "electrophoresis" of behavior whereby the various sequences of repeated motion "precipitate" in a prescribed, relatively stable order); ii) measuring kinematic variables (location, speed, path curvature, spatial spread and other variables reflecting the extent and complexity of motions), rather then scoring ad hoc classical behavior patterns or response categories.

The main message of our talk is that what you quantify is as important as how you quantify. A main help comes from the examination of overall statistical properties of the data of individual animals, of the strain, and of data collected across more than a single laboratory (the replicability issue, [2]).

Author Keywords

Mouse behavioral phenotyping, replicability, free exploration, Dimensionality Emergence Assay (DIEM), borderline roundtrips, home-related shuttles, wall-related shuttles, incursions, excursions, neophobia, freedom of movement.

ACKNOWLEDGEMENTS

Supported by a grant to IG and YB from the Israel Science Foundation (ISF). We thank Noldus Information Technology for the use of their EthoVision[®] system including the new EthoVision XT 7.0

- Benjamini, Y., Lipkind, D., Horev, G., Fonio, E., Kafkafi, N., Golani I. Ten ways to improve the quality of descriptions of whole-animal movement. *Neurosci. Biobehav. Rev.* (2010) (available online at doi:10.1016/j.neubiorev.2010.04.004).
- Kafkafi, N., Benjamini, Y., Sakov, A., Elmer, G.I., Golani, I. Genotype–environment interactions in mouse behavior: A way out of the problem. *Proc. Natl. Acad. Sci. USA*, 102 (2005), 4619-4624.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. The moment-to-moment developmental sequence of free exploration in a 3 h session of free open field behavior of a selected BALB/c mouse.

Symposium: Unveiling Affective Signals

Egon L. van den Broek

Human Media Interaction, University of Twente vandenbroek@acm.org +31 53 489 4650 Anton Nijholt Human Media Interaction, University of Twente anijholt@cs.utwente.nl +31 53 489 3686

Others

Joyce H.D.M. Westerink

Brain, Body, & Behaviour group, Philips Research joyce.westerink@philips.com +31 40 27 47793

ABSTRACT

The ability to process and, subsequently, understand affective signals is the core of emotional intelligence and empathy. However, more than a decade of research in affective computing has shown that it is hard to develop computational models of this process. We pose that the solution for this problem lays in a better understanding of how to process these affective signals. This article introduces a symposium that brought together various approaches towards unveiling affective signals. As such, it is envisioned to be a springboard for affective computing.

Author Keywords

emotion, affect, affective computing, methods, signal processing, pattern recognition.

ACM Classification Keywords

G.3. Probability and statistics; H.1.2. User/Machine systems, I.5.2. Pattern Recognition: Design methodology, I.5.4 Pattern Recognition: Applications.

SYMPOSIUM CONTENTS

Unveiling Affective Signals

Egon L. van den Broek & Anton Nijholt (University of Twente, The Netherlands), Joyce H.D.M. Westerink (Philips Research, The Netherlands).

Measuring Affective and Social Signals in Vocal Interaction

Khiet P. Truong (University of Twente, The Netherlands)

Facial EMG as a Tool for Inferring Affective States Anton van Boxtel. (Tilburg University, The Netherlands).

Motor, Emotional and Cognitive Empathic Abilities in

Children with Autism and Conduct Disorder Danielle M.A. Bons, Floor E. Scheepers, Nanda N.J. Rommelse, & Jan K. Buitelaar (University Centre

Nijmegen, The Netherlands). Mimicry as a Tool for Understanding the Emotions of

Marielle Stel & Kees van den Bos (University of Utrecht).

Social Signal Processing: Understanding Nonverbal Communication in Social Interactions

Alessandro Vinciarelli & Fabio Valente (Idiap Research Institute, Switzerland).

Automatic Measurement of Affect in Dimensional and Continuous Spaces: Why, What, and How?

Hatice Gunes & Maja Pantic (Imperial College London, U.K.).

Relative Affective Blindsight for Fearful Bodily Expressions

Bernard M.C. Stienen & Beatrice de Gelder (Tilburg University, The Netherlands).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Unveiling Affective Signals

Egon L. van den Broek

Human Media Interaction, University of Twente vandenbroek@acm.org +31 53 489 4650 Anton Nijholt Human Media Interaction, University of Twente anijholt@cs.utwente.nl +31 53 489 3686

Joyce H.D.M. Westerink

Brain, Body, & Behaviour group, Philips Research joyce.westerink@philips.com +31 40 27 47793

ABSTRACT

The ability to process and, subsequently, understand affective signals is the core of emotional intelligence and empathy. However, more than a decade of research in affective computing has shown that it is hard to develop computational models of this process. We pose that the solution for this problem lays in a better understanding of how to process these affective signals. This article introduces a symposium that brought together various approaches towards unveiling affective signals. As such, it is envisioned to be a springboard for affective computing.

Author Keywords

emotion, affect, affective computing, methods, signal processing, pattern recognition.

ACM Classification Keywords

G.3. Probability and statistics; H.1.2. User/Machine systems, I.5.2. Pattern Recognition: Design methodology, I.5.4 Pattern Recognition: Applications.

INTRODUCTION

The ability to process and, subsequently, understand affective signals of other people is the core of emotional intelligence and empathy. This capability continuously interacts with our behavior, in our everyday lives. Although a vast amount of work has been done on processing affective signals, progress is still limited. In particular, in ambulatory settings, automated processing of such signals is beyond science's current research [3]. We pose that substantial progress could be made when gaining on our understanding of affective signal processing.

PATTERN RECOGNITION BY MAN AND MACHINE

Recognition of affect, either by man or machine, is essentially a pattern recognition problem. The processing pipeline of pattern recognition (see also Figure 1) is as follows [1]: 1) a signal that is captured and, subsequently, 2) processed by i) a physical system (e.g., the eye or a CCD sensor). This system provides us with a ii) measurement space on which iii) feature selection and/or preprocessing is applied. This results in iv) a pattern space on which again v) feature selection is applied. This provides, vi) a reduced pattern space, which is used for 3) the pattern classification process. This classification process can either be the development of the classifying system or its execution on a new set of data. In the former case, the decision rule for the classifier is developed; in the latter case, the classification process provides a label for the signal that was captured. The classification process can be supervised or unsupervised. In the case a priori knowledge on the signal is available, 4) a classification error can be determined and 5) the classification process can be adapted. Without a priori knowledge these last two steps cannot be applied and unsupervised classification is applied. Please also see Figure 1, which provides a visualization of this pattern recognition processing pipeline.

Human's pattern recognition is only known in general lines. This makes it hard, not to say impossible, to define it as a computational model. Moreover, experimentation with parameters that are of possible importance in the pattern recognition process is hard with humans. In contrast, artificial pattern recognition systems can be defined up to the highest detail, manipulation of their parameters is easy, and obtaining results from them only requires some patience, as it can take some time.

Although the differences between human and artificial systems are overwhelming, they also have things in common. Both human and artificial pattern recognition systems often try to solve the same problems; e.g., playing chess, recognizing objects, or making decisions. If such a problem is solved, it is stated that the artificial pattern recognition system has been successful. Alternatively, human's pattern recognition system itself is sometimes taken as an example for artificial pattern recognition systems. Then, not only the results of the system are of interest but also to what extent the artificial system mimics its human counterpart. In the long run, the latter approach is also expected to bring significant progress in pattern recognition results by machines.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. The pattern recognition processing pipeline, inspired by the work of Meissel [1]. The gray boxes denote the stages that only apply to supervised learning strategies. With unsupervised learning strategies the decision algorithm is fixed, as no a priori knowledge is available on which the error detection can be based.

LEARNING FROM EXAMPLES

The ease with which humans learn is deceiving, as it is a refined process and evolutes over human one's life. Artificial pattern recognition aims to mimic human learning through applying adaptive algorithms, founded on decision rules. Throughout half a century, a broad range of adaptive algorithms have been proposed. In the continuous rat race to keep improving, these algorithms became more and more complex throughout the years.

To enable learning, the decision rule has to be able to adapt. For this, first, the error in the classification has to be identified; see also Figure 1. Second, a function has to be present that receives the error as input and enables the modification of the decision rule. Third, the decision rule has to be modified. Again, this is easier said than done. For example, in what stage of the pattern recognition pipeline, the modifying function has to hook into?

The type of examples on which the learning is based, is also of importance. For each class to identify, both positive and negative samples can be employed. Further, the level of deviation of the sample compared to the system's known set is of importance. With real world problems, the level of deviation can be expressed in both the number of dimensions in which the deviation is present and the distance in each of these dimensions.

Humans seem to have little problem with samples that urge them to adapt their pattern recognition process. However, this is possibly misleading. In either way, how humans adapt their pattern recognition processes is largely unknown. In general, machines try to adapt their pattern recognition pipeline through altering the following aspects: normalization, distance measures, dimensionality, and complexity of sample distributions.

STATE OF THE ART

Although significant differences exist between people's empathic abilities, most people sense affective signals automatically up to a level, machines are unable to reach. With the rise of the field affective computing, as coined by Picard [2], interest in machines that can sense people's emotions increased enormously. Now, more than a decade later, what is the status of affective computing and how was the development of this new subfield in science?

Affective computing has been mainly employed using three modalities: vision/image, speech, and biosignals. These three signals can be used to analyze facial expressions, speech utterances, movements and gestures, and physiological processes. However, it should be noted that the combination of biosignal processing with image and speech processing is rare. Most often, either image and/or speech processing of biosignal processing is applied.

Each of the three modalities applied in affective computing has its pros and cons. For example, computer vision/image processing techniques heavily rely on light sources, occlusion, and stereotype expressions. Speech processing is, in practice, heavily disturbed by environmental noise (e.g., from a radio) and is influenced by acoustic features of
environments. Nowadays, biosignals can be recorded with small sensors connected to tiny, light weight devices. However, biosignal recording is still experienced as rather obtrusive, when not integrated in other tools (e.g., a helmet or joystick) or clothes. Moreover, biosignals are sensitive to movement artifacts, signal loss (e.g., sensors that fall off), and humidity, to mention a few.

The mapping of human emotions on the three signal modalities is complex. Moreover, environmental influences can have a significant impact on both signal recording and the emotions people experience. Consequently, affective computing, although aimed to be used in our daily lives, is hardly applied outside well controlled lab environments.

IT STARTS WITH THE SIGNALS

Although the pattern recognition processes of man and machine are hard to compare, when going into more detail, they have things in common. One of these things is that both rely on the input for the pattern recognition process: the signal.

The quality of the signal is of the utmost importance for the pattern recognition process. Low signal quality can cause an ill defined measurement space. This is the foundation of the feature selection processes and everything follows from that. So, signal processing should be conducted with the utmost care.

Possibly even more important than the quality of the signal is our understanding of it. How is it originated and processed and how is it or should it be interpreted? A range of issues play a role in this; e.g., ethnic background, personality type, gender, and age. Nevertheless, some characteristics seem to be general; e.g., as has been shown with FACS for cultural diversity. However, also such generally accepted knowledge is a topic of debate.

SYMPOSIUM OVERVIEW

The symposium aims to initiate a multi-disciplinary knowledge exchange on all possible aspects that are of importance for affective signal processing. The session will discuss conceptual issues (e.g., ground truth) but also more applied issues of filtering and machine learning. Its rationale is to gather the available, but scattered, knowledge and to bring it from controlled lab settings to noisy real world applications.

Signals that will be discussed include social signals, biosignals, facial expression recognition, and nonverbal communication (e.g., movements). Differences between lab and real world studies will be discussed and, consequently, limitations of technology and challenges for science and technology will be identified [3]. Moreover, differences between special groups (e.g., as known in psychiatry) and generic applications will be assessed. Taken together, the session will bring together a group of distinguished scientists, from a range of disciplines. All will have their viewpoints and methods to unveil affective signals.

Bringing them together can possibly bring us a step further in the quest towards unveiling affective signals.

The symposium will be opened by the opening keynote speaker, Beatrice de Gelder [4] will discuss recent evidence on human's "relative affective blindsight for fearful bodily expressions". Nonconscious perception of emotions has repeatedly been shown for facial expressions. In contrast, this is not the case for bodily expressions, although being highly salient and known to influence our behavior towards others [4]. Using a parametric masking design, the unconscious perception of bodily expressions was measured with people. Participants had to detect in distinct experiments fearful, angry, and happy bodily expressions, among neutral bodily actions that served as distracters. Subsequently, the participants had to indicate their confidence. Results revealed a phenomenon that is coined relative affective blindsight, defined as two stimulus onset asynchronous conditions showing similar values, while the confidence ratings differed. In fact, this was only found for fearful bodily expressions, not for angry and happy bodily expressions.

In line with the presentation of De Gelder, Mariëlle Stel will explain how mimicry can be used "as a tool for understanding the emotions of others". She will address the question: How do people understand what others are emotionally experiencing? She argues that mimicking nonverbal expressions of other people (i.e., copying others' behavior) can facilitate the understanding of emotions they are experiencing. When people mimic nonverbally expressed emotions, this affects their own emotions, corresponding to an afferent feedback mechanism. As a consequence of this mechanism, the mimicker perceives the emotions of others more strongly, which facilitates emotion understanding. The first two presentations involve research on participants, who are considered to be a representative sample of the community of healthy adult people. However, as is known from various scientific disciplines (e.g., medicine and psychology), research on special cases and people suffering from disorders as well as on the development of people should be of the utmost interest.

Floor Scheepers and Jan Buitelaar will give an overview of the studies that investigated motor, emotional, and cognitive empathy in juveniles with autism or conduct disorder (CD). Studies that measured response to emotional faces with use of facial EMG, ECG, skin conductance, or eye-tracking are discussed [2]. In autism, facial mimicry, emotion recognition, and attention to the eyes seems to be reduced. In CD, facial mimicry and recognition of fear and sad facial expressions are impaired. Although further research is needed to investigate autonomic emotional empathic responses to emotional faces in both patient groups, major differences between autism and CD are hypothesized.

As indicated above, various biosignals are recorded and, subsequently, analyzed with the aim to unveil people's emotional state. One of these biosignals concerns the facial EMG. Facial EMG is a generally accepted tool for inferring affective states, as will be outlined by Anton van Boxtel. Van Boxtel will give a concise overview of methodological aspects of recording facial EMG signals as an index of affective states, which are known to be of the utmost importance. In addition, both strengths and weaknesses of the application of facial EMG in clinical and other applied settings will be emphasized; cf. [3].

Where Van Boxtel already outlines concerns when bringing affective signal processing to practice, Hatice Gunes and Maja Pantic will provide a brief overview of the current state-of-the-art in automatic measurement of affect signals. In classifying emotions, they distinguish dimensional and continuous spaces and, consequently, seek answers to the following questions: i) *why* has the field shifted towards dimensional and continuous interpretations of affective displays recorded in real-world settings? ii) *what* are the affect dimensions used, and the affect signals measured? and iii) *how* has the current automatic measurement technology been developed, and *how* can we advance the field?

As already indicated by Gunes and Pantic, not only biosignals are of interest also other (social) signals can show to be a rich source of information. Khiet Truong will discuss how and what type of measurements of vocal interactional behavior can be used to recognize both affective and social signals [2,5]. Three studies will be presented that deal with i) the collection and recognition of spontaneous vocal and facial expressions, ii) the detection of laughter, and iii) the meaning of overlapping speech (i.e., interruptions) in conversations. Following the results of these studies, both the pros and cons of affective speech processing will be evaluated. In addition, fundamental issues such as 'ground truth labeling' and collection of spontaneous data will be discussed.

The second key-note speaker, Alessandro Vinciarelli, will close the symposium. In line with the presentation of Truong, he will introduce the recently emerged field of social signal processing (SSP); i.e., understanding nonverbal communication in social interactions [5]. In a clinical context, this issue was already briefly touched by Scheepers and Buitelaar. SSP is founded on the idea that nonverbal communication is physical and, thus, provides machine detectable evidence of mutual relational attitudes (social signals). Subsequently, SSP aims to automatically analyze, model, and synthesize nonverbal behavior in human-human (and human-machine) interactions.

Thus throughout the symposium, a plethora of affective signals are brought to attention. Both methodological and signal processing issues have been discussed. Various settings in which affective signals are recorded were touched upon; e.g., from psychiatry to gaming. Moreover, not only the technical aspects but also social aspects (e.g., people's mimicry) have been brought to attention.

CONCLUSION

A wider scope is needed for affective computing as progress is limited and the techniques employed are still too fragile to bring from lab to life. We pose that not the pattern analysis and machine learning component of affective computing constitute is the bottle neck but instead this is formed by the affective signals that are the input for the classification processes.

Understanding affect and its signals requires an interdisciplinary approach, as is adopted for this symposium. Engineering, neuroscience, psychological, biological, and clinical approaches are explored. Through their integration, we envision a significant leap in unveiling affective signals.

The range of disciplines the speakers originate from enables a true interdisciplinary discussion and knowledge exchange. We hope that, in time, this symposium will show to be a small but significant step forward in unveiling affective signals. In this way, step by step, affective computing will find its scientific foundation and the path to true progress can be finally paved.

ACKNOWLEDGMENTS

This work has been supported in part by the European Community's Seventh Framework (FP7/2007-2013) under grant agreement no. 231287 (SSPNet) and grant agreement no. 211486 (SEMAINE). The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. Further, would the organizers of the symposium like to thank the members of their program committee, Gert-Jan de Vries and Henriette C. van Vugt (Philips Research) and Frans van der Sluis (University of Twente), for making this session possible.

REFERENCES

- 1. Meisel, W.S. *Computer-oriented approaches to pattern recognition*. Series of Mathematics in Science and Engineering, vol. 83. New York, NY, USA: Academic Press, Inc., 1972.
- 2. Picard, R.W. *Affective Computing*. Boston, MA, USA: MIT Press, 1997.
- 3. Broek, E.L. van den & Westerink, J.H.D.M. Considerations for emotion-aware consumer products. *Applied Ergonomics* 40, 5 (2009), 1055-1064.
- de Gelder, B. Why bodies? Twelve reasons for including bodily expressions in affective neurosciences. *Philosophical Transactions of the Royal Society B: Biological Sciences 364*, 1535 (2009), 3475-3484.
- Vinciarelli, A., Pantic, M. and Bourlard, H. Social signal processing: Survey of an emerging domain. *Image and Vision Computing* 27, 12 (2009), 1743-1759.

Measuring Affective and Social Signals in Vocal Interaction

Khiet P. Truong

Human Media Interaction, University of Twente P.O. Box 217, 7500 AE Enschede, The Netherlands k.p.truong@ewi.utwente.nl

ABSTRACT

In this paper, I will discuss how and what type of measurements of vocal interactional behavior can be used to recognize affective and social signals. Three studies will be presented that deal with 1) the collection and recognition of spontaneous vocal and facial expressions in a gaming context, 2) the detection of laughter in meetings, and 3) the relation between dominance and overlapping speech in multiparty conversations. On the basis of these studies, (dis-)advantages and issues in speech processing for affective and social computing will be evaluated. Acoustic features, but also simple speech or no-speech information were employed in these studies. In addition, fundamental issues such as 'ground truth labeling' and collection of spontaneous data are also discussed.

Author Keywords

speech, social signals, affective computing.

ACM Classification Keywords

H5.1. Multimedia information systems: Audio input/output H5.2. User Interfaces: Natural language

INTRODUCTION

In speech analysis research, there is a long history of analysing speech signals with respect to human affective and social signals. One of the first studies carried out on real-life, natural emotion speech material [1], described an acoustic analysis of a well-known radio broadcast of the explosion of the Hindenburg Zeppelin in 1937. The audio sample contains the speech of a radio reporter who witnessed the explosion live during broadcasting and who continues his report in an emotional tone of voice. Speech analyses were performed by hand on a relative small amount of speech data. In the 70s, automatic speech feature extraction methods and statistical learning algorithms made it possible to develop simple automatic speech recognition systems (ASR). The tasks of these systems was to recognize *what* is said. Automatically recognizing *how* something is said gained more interest in the past 10-15 years, when the number of serious efforts into the automatic recognition and understanding of human affective and social signals was (and still is) growing steadily. I will briefly present three studies that we have carried out and discuss how the speech modality was used for our research. Not only speech-specific issues will be discussed but also some general issues in affective computing such as emotion labeling and the collection of spontaneous data.

SPEECH-BASED RECOGNITION OF AROUSAL AND VALENCE

An increasing number of researchers adopt the arousalvalence dimensional model of affect for automatic affect recognition. However, only a few databases containing spontaneous vocal expressions have been developed with continuous arousal-valence annotations. In addition, we are also interested in investigating differences between 'felt' affect annotations performed by people who have undergone the emotion themselves, and 'perceived' annotations performed by naive observers. Hence, we decided to record our own audiovisual spontaneous corpus of affect. We describe how we developed speech-based affect recognizers that are trained to predict a location in the arousal-valence space.

Collecting and Labeling Data: the TNO-Gaming Corpus

An audiovisual corpus containing expressive vocal and facial behavior was collected by inviting people to play a multiplayer videogame. The gaming sessions took place at TNO in Soesterberg, the Netherlands. In order to obtain affect annotations, each participant labeled his/her own affect after each gaming session. Seventeen males and eleven females with an average age of 22.1 years (2.8 standard deviation) participated in the gaming experiment. Vocal behavior was especially stimulated and was recorded by microphones that were attached near the mouths of the participants to reduce the effects of crosstalk (facial expressions were also recorded by webcams). The video content of the game itself was also stored. A more detailed description of this corpus can be found in [2].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Affect Rating by the Gamers Themselves ('Felt') and Naive Observers ('Perceived')

After each gaming session, the participants annotated their own emotions based on the video recordings and the videostream of the game content itself. We asked the participants to recall what they were *feeling* during playing. The participants rated the running video and could not pause or rewind the video. Two scales were used for rating, namely the arousal (active-passive) and valence (positivenegative) scale. Each 10 seconds, an arrow appeared on the screen to signal the participants to give an arousal and valence rating separately on a scale from 0 to 100. We will call these ratings the *self*-ratings. In this way, a total of 7473 affect-rated speech segments were obtained (after speech segmentation). Out of this total, 2400 segments were selected (sampling the whole arousal-valence space of the self-annotations evenly) for re-rating performed by 6 naive observers. Similar to the rating procedure of the gamers, the naive observers were asked to rate what they perceived on the arousal and valence scale. Each observer rated different parts of the dataset that overlapped with parts that were rated by other observers, such that each segment was rated by 3 different observers. In order to obtain unique ratings for each segment, the 3 different ratings were averaged. We refer to these ratings as other.avg-ratings.

Agreement Between Gamers and Naive Observers

To what extent do the ratings given by the gamers themselves differ from the ones given by the observers? To answer this question, the continuous ratings were discretized into 5 classes and Krippendorff's α (ordinal, [3]) was applied to assess the level of agreement between the *self*-ratings and the *other.avg*-ratings. Relatively low agreement scores were obtained: 0.27 and 0.36 for arousal and valence respectively (the agreement *among* the external observers was 0.28 and 0.57 for arousal and valence respectively). How does this discrepancy influence the performance of affect recognizers developed with these two types of ratings?

Recognition Task

Two types of speech-based arousal and valence recognizers were developed in parallel: one based on the self-ratings and the other based on the other.avg-ratings. The task of the affect recognizers was to predict scalar values on continuous scales of arousal and valence, rather than to recognize categories of emotions.

Method and Features

A Support Vector Regression approach was used to predict arousal and valence values (see [2]). As speech features, we extracted a selection of features that were commonly used in emotional speech research as described in the literature. First, a voiced-unvoiced detection algorithm (available in Praat [4]) was applied to find the voiced units. The features were extracted over each voiced unit of a segment. In addition, global information calculated over the whole segment (instead of per voiced unit) about the speech rate and the intensity and pitch contour was included. The following features were extracted: 4 pitch-related (mean, standard deviation, range (max-min), mean absolute pitch), 4 intensity-related (Root-Mean-Square, mean, range (maxmin), standard deviation), 5 energy-distribution-inspectrum-related (slope Long-Term Averaged Spectrum, Hammarberg index, standard deviation, centre of gravity, skewness), 24 Mel Frequency Cepstrum Coefficents (12 coefficients, 12 first order derivatives), and 6 other features (speech rate1, speech rate2, mean positive/negative slope pitch, mean positive/negative slope intensity). Subsequently, the features extracted on voiced-unit-level were aggregated to segment-level by taking the mean, minimum, and maximum of the features over the voiced units. Hence, we obtained per segment a feature vector with 117 dimensions. These features were normalized for speaker variation by transforming the features to z-scores: $z=(x-\mu)/\sigma$, with μ and σ calculated over a development set.

Results

Similarly to the human-human agreement assessment, the results are expressed in terms of Krippendorff's α (to allow for comparison): predicted values and reference ratings were first discretized into 5 classes, and Krippendorff's α was computed. We can observe in Table 1 that the recognizer based on *other.avg*-ratings performs much better than the *self*-based recognizer; the *other.avg*-ratings appear to describe the affect perceived more consistently.

| | self | other.avg |
|---------|------|-----------|
| arousal | 0.22 | 0.42 |
| valence | 0.10 | 0.28 |

Table 1. Results recognition task, expressed in Krippendorff's α [3] (i.e., agreement between man-machine).

Discussion

What we consider 'ground truth' is of great influence on the performance of the recognizer. In this study, the *self*-ratings do not appear to be consistent enough for the learning algorithm, whereas the perceived affect ratings give better performance. It remains a challenge to acquire reliable affect ratings as the perception of affect is influenced by many factors such as context, culture and personality. With respect to automated recognition of affect, the acousticsbased recognizers have difficulty recognizing valence; a fusion with other modalities, such as facial expressions, could boost performance considerably. Finally, there appears to be an increasing interest for the recognition of gradations of affect. Adopting the dimensional approach and applying regression techniques to recognize these gradations is a relatively new approach, and hence, more investigation is needed into e.g., adequate evaluation procedures for these models.

AUTOMATIC DETECTION OF LAUGHTER IN MEETINGS

In this study, our goal was to detect paralinguistic events in speech (see also [5]). One the most recognizable events for humans, and (luckily) often annotated event is laughter. The first step towards this goal was to discriminate between laughter and speech. One could ask the question why developing a separate laughter detector is necessary as most ASR systems already have laughter models included. ASR systems are typically not tuned to detect paralinguistic events, in fact, these events are usually seen as 'garbage'. In addition, the computation cost and labour needed to train such a system would not make ASR a good candidate for the detection of paralinguistic events.

Data - ICSI Meeting Corpus

As speech data, the ICSI Meeting Corpus was used [6]. For training and testing, respectively 30 and 3 meetings were used, totaling an amount of approximately 90 minutes (>2500 segments) for each class of laughter and speech separately.

Method and Features

We extracted 12 Perceptual Linear Coding Coefficients (PLP, which model spectral properties of the speech according to a model adapted to the properties of the human ear) plus 1 energy component and their 1st order derivatives. After normalization to z-scores, these features were used in Gaussian Mixture Models (GMMs) to train a laughter and speech model. In classification, a log-likelihood ratio is used to decide the class.

Results

With relatively straightforward methods and features, an Equal Error Rate (EER, the point where the false alarm rate is equal to the miss rate) of approximately 6% was achieved. Note that only audible laughter segments were included in the speech material; e.g., unvoiced laughter was not part of this evaluation task.

Discussion

The next step would be to perform (real-time) laughter detection (e.g., [7]) and to use additional modalities (e.g., [8]). Furthermore, the detector does not give an interpretation of the laughter yet, and does not make distinctions between types of laughter (e.g., voiced vs unvoiced). Rather, we view the output of this detector as a useful feature for a higher-level affect recognition system.

ANALYSIS OF OVERLAPPING SPEECH AND DOMINANCE IN MULTIPARTY CONVERSATION

In social sciences and psychology, interuptions have frequently been studied with respect to cultural, gender, and status aspects. Traditionally, interruptions are treated as indicators of power, control, and dominance [12]. However, while some interruptions may indeed be seen as power displays, some of these are actually rapport displays. But in general, given the assumptions that turn-taking is regulated by the notion of 'one speaker speaks at a time', the social convention that it is impolite to speak at the same time when someone else is speaking, and the fact that it is difficult to decode the message when two speakers are speaking at the same time, interruptions are generally perceived as rude and competitive. In this study, we explored how interruptions can tell us something about social group dynamics in multiparty conversation, dominance and speaker role in particular.

Data - AMI Meeting Corpus

As multiparty conversation data, the AMI Corpus was used, see [9] for a detailed description. In short, the AMI corpus is comprised of recorded meetings in which 4 participants are brainstorming about the development of a tv remote control. Each participant has a role: there is a project manager, a user interface expert, a marketing expert, and an industrial designer. The corpus has multimodal and multi-layered annotations of e.g., gaze, head and hand gestures, and dominance. For our analyses, we used the manual word transcriptions and the dominance annotations as described in [10]. Each meeting (we used IS1000a, IS1001b, IS1003b, and IS1006b) was divided into 5-minute segments in which each meeting participant was ranked by dominance by 3 different raters. Only those segments were used where there was a majority agreement on ranked dominance.

Overlapping Speech Analysis

Overlapping speech parts were automatically found based on the manual speech transcriptions and the following definition of overlapping speech: we speak of overlap when there is more than 1 person talking at the same time. The person who performs the overlapping is the active overlapping speaker, while the overlappee is the passive overlapped speaker. We adopted the following measurements of vocal behavior as proposed by [10] who have analyzed overlapping speech in the context of interviews: an attack/resist ratio R and the overall frequency of active overlaps D. R is calculated as (A-P)/(A+P) where A is the number of words spoken by the active overlapper and P is the number of words spoken by the passively overlapped. A negative R indicates that the speaker is more a floor keeper than an overlapper. D is calculated as 100*A/(M+P+A) where M is the number of words spoken in a mono-speaker condition. A low D indicates low density of active overlaps. In [10], these measures were succesfully applied to journalists and interviewees: journalists had higher R and D than interviewees.

Results

The R and D measurements were compared to the dominance annotations which gave mixed results. In 2 of the 4 meetings analyzed, the measures appeared to show relation to dominance, see Figure 1. However, counter-intuitively, low R and D appeared to be associated with high dominance. In addition, correlating R and D to role (i.e., project manager etc.) did not reveal any patterns.



Figure 1. '+' is the most dominant person, '.' the least dominant, R on the x-axis, D on the y-axis.

Discussion

We have to keep in mind that the R and D measures are designed for a journalist-interviewee setting. However, the thought behind these measurements may be the same: namely, an overlapper is usually perceived as dominant. The fact that we did not obtain the expected values for R and D may have to do with the observation that in multiparty conversational speech there are a lot more cooperative interruptions, e.g., backchannels ('mm-hmm'), than in journalists-interviewees settings. Calculating these measures for competitive overlaps *only* may show improvement.

CONCLUSIONS AND DISCUSSION

We have given a brief overview of some analyses of affective and social signals in speech to show how different measures of vocal behavior can inform us about human affective and social behavior. Specifically, we looked at arousal/valence prediction, laughter detection, and a classic example of a social signal namely dominance. First, the study in arousal/valence prediction showed how differences in annotation affect the recognizer's performance, and the study showed that the collection of spontaneous affect data remains a challenge: for speech analysis, close-talk microphones that pick up few surrounding noise or talk are preferred (in the first study, the microphones were attached with tape to the face near the mouth), although in realworld situations, this is not always possible. We also looked at a more low-level feature of affect, namely laughter, for which relatively reliable annotations are available. In this case, relatively high detection performances were obtained with the disadvantage that the detector trained only detects laughter without giving an interpretation. Modeling the interpretation is not only a matter of collecting a sufficient amount of data for the training of algorithms, it is also about modeling the *context*. For example, depending on the situation, laughter can signal humor, joy, politeness or shyness. The first two studies covered the modeling of affect with acoustic features whereas in the third study, we looked at simple speech and no-speech information to model social signals such as dominance. With the on-andoff pattern of speech, overlaps in multiparty speech were analyzed to see whether certain overlap behavior could be related to social group behavior. In this case, information about the competitive or cooperative nature of the overlap in speech would help to reveal this relation. In general, speech as a modality offers an inobtrusive way of measuring affective and social signals. The use of speech with other measurements such as physiological measures is an interesting combination. Humans have high control over the vocal apparatus so that emotions in the voice can be suppressed, while it is much harder to suppress certain physiological measures. We also have to keep in mind that social conventions may pose limitations to what is expressed. For future research, we suggest to investigate how context and personality can be modelled for the interpretation of affect and social signals.

ACKNOWLEDGEMENTS

This research has been supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet).

REFERENCES

- 1. Williams, C.E. and Stevens, K.N. Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*, 52 (1972), 1238–1250.
- Truong, K.P., van Leeuwen, D.A., Neerincx, M.A., and de Jong, F.M. Arousal and Valence Prediction in Spontaneous Emotional Speech: Felt Versus Perceived Emotion. In *Proceedings of Interspeech* (2009), 2027-2030.
- 3. Krippendorff, K. Computing Krippendorff's Alpha-Reliability (2007). Retrieved from http://www.asc.upenn .edu/usr/krippendorff/webreliability.doc.
- 4. Boersma, P. Praat, a system for doing phonetics by computer. *Glot International*, *5*, 9-10 (2001), 341-345.
- 5. Truong, K.P., and van Leeuwen, D.A. Automatic discrimination between laughter and speech. *Speech Communication*, 49 (2007), 144-158.
- Janin, A., Ang, J., Bhagat, S., Dhillon, R., Edwards, J., Macias-Guarasa, J., Morgan, N., Peskin, B., Shriberg, E., Stolcke, A., Wooters, C., and Wrede, B. The ICSI meeting project: resources and research. In: NIST ICASSP2004 Meeting Recognition Workshop, Montreal, Canada, 2004.
- Melder, W.A., Truong, K.P., den Uyl, M., van Leeuwen, D.A., Neerincx, M.A., Loos, L.R., and Plum, B.S. Affective multimodal mirror: sensing and eliciting laughter. In Proceedings of the International workshop on Human-Centered Multimedia (HCM) (2007), 31-40.
- Reuderink, B., Poel, M., Truong, K. P., Poppe, R., and Pantic, M. Decision-Level Fusion for Audio-Visual Laughter Detection. In *Proceedings of MLMI*, (2008), 137-148.
- 9. Carletta, J. Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus.

Language Resources and Evaluation Journal, 41, (2007), 181–190.

- Hung, H., and Gatica-Perez, D. Identifying Dominant People in Meetings from Audio-Visual Sensors. In *Proceedings of FG 2008*, 1-6.
- 11. Adda-Decker, M., Barras, C., Adda, G., Paroubek, P., de Mareuil, P. B., Habert, B., Annotation and analysis of

overlapping speech in political interviews. In *Proceedings of LREC* (2008), 3105-3111

12. West, C., Against our will: Male interruptions of females in cross-sex conversations. *Annals of the New York Academy of Sciences*, 327 (1979), 81-97.

Facial EMG as a Tool for Inferring Affective States

Anton van Boxtel Tilburg University

P.O. Box 90153, 5000 LE Tilburg, The Netherlands

a.vanboxtel@uvt.nl

ABSTRACT

In this presentation, I will give a concise overview of several important methodological aspects of recording facial EMG signals as an index of affective states. In addition, several strengths and weaknesses of this technique during practical applications will be emphasized.

Author Keywords

Facial EMG, EMG recording, EMG signal processing, emotion.

INTRODUCTION

The human face may be considered the richest source of information for revealing someone's affective state. Healthy persons during daily life automatically recognize affective facial expressions quite well. For scientific purposes, affective facial expressions can be quantitatively analyzed by trained experts coding elementary facial actions, or by automated systems recognizing facial expressions through visual analysis of facial movements. Another method is recording electromyographic (EMG) signals of specific facial muscles. Both visual and EMG methods have their strengths and weaknesses. I will present a concise overview of several advantages and disadvantages of EMG signals as a tool for inferring affective states.

RECORDING AND ANALYSIS OF FACIAL EMG

Electrodes

Facial EMG is generally recorded bipolarly with small surface electrodes (contact area diameter ≤ 4 mm) located close to each other. EMG activity is frequently recorded from specific muscles playing a prominent role in the expression of elementary emotions, like happiness, surprise, anger, sadness, fear, and disgust (Figure 1). Although affective facial EMG responses may show bilateral differences in individual subjects, group results generally do not show systematic differences between both sides of the face during spontaneous emotional expressions [3].



Figure 1. Electrode locations for measuring facial EMG activity.

Conditioning of EMG Signals

Following amplification, the EMG signal must be bandpass filtered within the frequency range 20-500 Hz, being the predominant frequency range of facial EMG signals. Effective high-pass filtering at 20 Hz is essential because of the strong influence of low-frequency artifacts such as motion potentials, eye movements, eyeblinks, activity of neighboring muscles, respiration, swallowing, etc. [12]. If not removed, low-frequency artifacts may dominate the real facial EMG potentials (which under natural circumstances are often small; see Figure 2) and may thus strongly affect the estimation of real EMG activity. In most practical applications occurring outside an electrically-shielded laboratory, it may also be necessary to remove 50-Hz power line interference by applying 50-Hz notch filtering.

Quantification of EMG Amplitude

The EMG is a signal with random properties. Its amplitude can be quantified by calculating the mean rectified EMG amplitude during a fixed time interval on the basis of the rectified, or rectified and smoothed (low-pass filtered), EMG signal (Figure 3). The duration of the optimal analysis epoch depends on the purpose of the study. A longer interval may be necessary if one is interested in relatively

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 2. Empirical EMG power spectrum (above) and theoretical power spectrum (below). The low frequency range (< 20 Hz) is dominated by large frequency components caused by artifacts.

steady emotional states, such as a subject's mood state [9]. A shorter interval will be required if one is interested in dynamic changes in emotional responses, for example short-lived facial mimicry responses [1]. Using a shorter interval, temporal resolution becomes better but random error also increases, with negative consequences for reliability. In my experience, the optimal analysis interval for tracking fast dynamic changes in facial expression, while avoiding large effects of random error, has a duration in the order of magnitude of 100 ms (Figure 4).

Standardization of EMG Responses

Baseline EMG amplitudes and affective EMG response magnitudes strongly vary between individuals, not only because of differences in affective processes but also due to anatomical and biophysical differences. This implies that, when determining group means, individual contributions will strongly differ in weight. An adequate method to standardize individual results, and making them comparable between individuals, is expressing EMG response magnitudes as a proportion of an adequate baseline value. As EMG amplitudes are measured on a ratio scale, expressing them as a proportion of baseline level is preferred rather than expressing them as difference scores between baseline and response levels. This standardization also enables a direct comparison of affective responses in different muscles within the same person. It also provides a solution for the problem that EMG amplitudes of a certain person may considerably vary over repeated measurement sessions, even when precautions have been taken to place



Figure 3. Raw EMG signal (upper frame), rectified raw EMG signal (middle frame), and smoothed rectified EMG signal (lower frame).

electrodes on exactly the same locations. Relatively minor changes in location may have strong effects on absolute signal amplitude [7].

An even better standardization method would be expressing EMG activity of a specific muscle as a proportion of the EMG level during maximal voluntary contraction of that muscle [4,14]. Theoretically, this leads to a better compatibility between different subjects or different measurement sessions from the same subject. However, in practice this procedure is somewhat complicated since it requires training the subject to perform selective maximal contractions of specific muscles.



Figure 4. Dynamic corrugator EMG response with a time resolution of 100 ms to a brief film clip showing a dynamic facial expression of anger [1]. The arrow indicates the apex of the dynamic angry expression.

ADVANTAGES AND DISADVANTAGES OF FACIAL EMG AS AN INDEX OF AFFECTIVE STATES

Sensitivity

A basic problem associated with systems relying on the analysis of observable facial motions is that weak or moderate affective responses may be accompanied by visually undetectable facial actions [11]. Using EMG techniques, even the weakest responses, remaining under the visual detection threshold, can be detected, especially since most facial muscles are located at close distance from the surface electrodes. Besides their great sensitivity in the amplitude domain, EMG signals also have a good time resolution so that rapid changes in activity can be reliably measured. Using techniques relying on observable facial movements, small dynamic transitions in activity may be less well observed since they may be masked by the stiffness of overlying cutaneous and subcutaneous tissues. Good dynamic response properties are a prerequisite, among others, for accurate measurement of response latencies to affective stimuli or rapid changes in emotional state during social interactions (e.g., emotional mimicry).

Selectivity

An important application of facial EMG activity may be discriminating between different elementary emotions.

| Elementary emotions | Muscles involved | Produced actions |
|---------------------|--|---|
| Happiness | Orbicularis oculiZygomaticus major | Closing eyelids Pulling mouth corners upward and laterally |
| Surprise | FrontalisLevator palpebrae superioris | Raising eyebrows Raising upper eyelid |
| Fear | Frontalis Corrugator supercilii Levator palpebrae superioris | Raising eyebrows Lowering eyebrows Raising upper eyelid |
| Anger | Corrugator supercilii Levator palpebrae superioris Orbicularis oculi | Lowering eyebrows Raising upper eyelid Closing eyelids |
| Sadness | FrontalisCorrugator superciliiDepressor anguli oris | Raising eyebrows Lowering eyebrows Depressing lip corners |
| Disgust | Levator labii superioris Levator labii superioris alaeque nasi | Raising upper lip Raising upper lip and wrinkling nasal skin |

 Table 1. Predominant facial actions during the expression of elementary emotions according to Ekman and Friesen [2].



Figure 5. Time course of corrugator and zygomaticus EMG activity while watching a humorous TV-commercial with a climax toward the end of the film clip. The attention of the participants was experimentally manipulated (high vs. low involvement).

These emotions are characterized by specific configurations of facial actions [2] (see Table 1). There is some interindividual variability in these patterns which, at least partly, may be due to interindividual differences in the morphology of the facial musculature. A robust, and often replicated, finding is that positive and negative affective states can be reliably distinguished on the basis of corrugator and zygomaticus responses, corrugator responses showing a negative linear relationship with emotional valence and zygomaticus responses showing a positive curvilinear relationship [8]. Corrugator activity is not only facilitated during negative emotional states but is also inhibited during positive emotions [1,8] (see also Figure 5). A challenging application would be discriminating between elementary emotions on the basis of the pattern of EMG responses of different muscles. Multivariate analysis of EMG response patterns may be a useful technique for this purpose. Subjecting facial EMG response patterns to discriminant analysis, a reasonably accurate identification could be made of (a) elementary mood states induced by mental imagery, and (b) posed expressions of elementary emotions (happiness, fear, anger, sadness) [4].

However, reliable discrimination between specific positive or negative emotions on the basis of facial EMG response patterns remains complex as yet. There are two important factors contributing to this problem. First, recordings of EMG activity from different muscles are less selective than desirable. The major reason for this disadvantage is crosstalk, that is, the phenomenon that electrical activity generated by a specific muscle spreads to adjacent areas through volume conduction. This activity, at least partly, will be detected by electrodes located on muscles in the vicinity of the target muscle. However, using a high-density grid of surface electrodes, estimations of real facial EMG activity could be considerably improved [6]. Another impeding factor is that emotional experiences under natural circumstances often consist of a mixture of elementary emotions which, in addition, may rapidly change so that EMG response patterns may thus be a function of such undetermined or dynamic emotional states.

Confounding Factors

A limitation of EMG recordings (but also of all other techniques analyzing facial movements) as an index of emotional processes is that the human face does not only display affective responses but also produces a large variety of activities unrelated to emotional processes like speech, mental effort or mental fatigue, task involvement or performance motivation, anticipation of sensory stimuli, preparation of motor responses, orienting responses, and startle reflexes [10,13,15,16,17]. In experimental studies, we can try to control for such disturbing influences but during practical applications outside the laboratory this will be much more difficult. Anyhow, the influence of such ubiquitous factors should be carefully evaluated to avoid invalid conclusions regarding a person's affective state.

Obtrusiveness

A basic limitation of using facial EMG recording in practical applications is the obtrusiveness of this technique. The extensive preparation being required and the connection with the recording equipment through electrode leads may interfere with spontaneous, natural behavior. Although the tactile sensations associated with the presence of skin electrodes generally quickly habituate during the absence of facial movements, strong dynamic facial movements may generate tactile sensations due to the high density of sensitive mechanoreceptors in the facial skin [5]. This may particularly occur when a large array of closely spaced electrodes is used.

Conclusion

EMG recording may be considered a sensitive technique for inferring subjective mood states or affective responses. However, it has limitations for many applications under natural life circumstances due to its obtrusiveness and the fact that facial activity is influenced by many other, nonaffective, behavioral factors. Also, methodological improvements are necessary to enhance its effectiveness as a tool for reliable differentiation between specific emotions.

REFERENCES

- 1. de Wied, M., van Boxtel, A., Zaalberg, R., Goudena, P.P., and Matthys, W. Facial EMG responses to dynamic emotional facial expressions in boys with disruptive behavior disorders. *Journal of Psychiatric Research*, 40 (2006), 112-121.
- 2. Ekman, P., and Friesen, W.V. Facial Action Coding System (FACS): A technique for the measurement of facial action. Consulting Psychologists Press, Palo Alto, CA, 1978.
- 3. Ekman, P., Hager, J.C., and Friesen, W.V. The symmetry of emotional and deliberate facial actions. *Psychophysiology*, *18* (1981), 101-106.
- 4. Fridlund, A.J., Schwartz, G.E., and Fowler, S.C. Pattern recognition of self-reported emotional state from multiplesite facial EMG activity during affective imagery. *Psychophysiology*, *21* (1984), 622-637.
- Johansson, R. S., Trulsson, M., Olsson, K. Å., and Westberg, K.-G. Mechanoreceptor activity from the human face and oral mucosa. *Experimental Brain Research*, 72 (1988), 204-208.
- 6. Lapatki, B.G. The facial musculature. Characterization at a motor unit level. Unpublished doctoral dissertation, Radboud University, Nijmegen, 2010.
- Lapatki, B.G., Oostenveld, R., van Dijk, J.P., Jonas, I.E., Zwarts, M.J., and Stegeman, D.F. Optimal placement of bipolar surface EMG electrodes in the face based on single motor unit analysis. *Psychophysiology*, 47 (2010), 299-314.
- Larsen, J.T., Norris, C.J., and Cacioppo, J.T. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40 (2003), 776-785.
- 9. Sirota, A.D., Schwartz, G.E., and Kristeller, J.L. Facial muscle activity during induced mood states: differential growth and carry-over of elated versus depressed patterns. *Psychophysiology*, 24 (1987), 691-699.
- 10. Stekelenburg, J.J. and van Boxtel, A. Pericranial muscular, respiratory, and heart rate components of the orienting response. *Psychophysiology*, *39* (2002), 707-722.
- Tassinary, L.G., and Cacioppo, J.T. Unobservable facial actions and emotion. *Psychological Science*, 3 (1992), 28-33.
- 12. van Boxtel, A. Optimal signal bandwidth for the recording of surface EMG of facial, jaw, oral, and neck muscles. *Psychophysiology*, *38* (2001), 22-34.
- 13. van Boxtel, A., Damen, E.J.P., and Brunia, C.H.M. Anticipatory EMG responses of pericranial muscles in relation to heart rate during a warned simple reaction time task. *Psychophysiology*, *33* (1996), 576-583.
- 14. van Boxtel, A., Goudswaard, P., van der Molen, G.M., and van den Bosch, W.E.J. Changes in EMG power spectra of facial and jaw-elevator muscles during fatigue. *Journal of Applied Physiology*, 54 (1983), 51-58.

- 15. van Boxtel, A., and Jessurun, M. Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a two-choice serial reaction task. *Psychophysiology*, *30* (1993), 589-604.
- 16. Veldhuizen, I.J.T., van Boxtel, A., and Waterink, W. Tonic facial EMG activity during sustained information processing: Effects of task load and achievement

motivation. Journal of Psychophysiology, 12 (1998), 188-189.

17. Waterink, W., and van Boxtel, A. Facial and jaw-elevator EMG activity in relation to changes in performance level during a sustained information processing task. *Biological Psychology*, *37* (1994), 183-198.

Motor, Emotional and Cognitive Empathic Abilities in Children with Autism and Conduct Disorder

Danielle M.A. Bons^{1,2} d.bons@psy.umcn.nl +31 (0)488 – 469 611

Nanda N.J. Rommelse^{1,2} n.lambregts-rommelse@psy.umcn.nl +31 (0)24 351 2222

¹Karakter child- and adolescent psychiatry University Centre Nijmegen, Zetten-Tiel P.O. Box 104, 6670AC Zetten, The Netherlands

ABSTRACT

This paper gives an overview of the studies that investigated motor, emotional and cognitive empathy in juveniles with autism or conduct disorder. Studies that measured response to emotional faces with use of facial EMG, ECG, skin conductance, eye-tracking or emotion recognition are discussed. In autism facial mimicry and emotion recognition, as well as attention to the eyes, seem to be reduced. In conduct disorder facial mi micry seems to be impaired as well as recognition of fear and sad facial expressions, and possibly associated with lack of attention to the eyes. Further research is needed to investigate autonomic emotional empathic response to emotional faces in both patient groups. Major differences between ASD and CD are hypothesized.

Author Keywords

Empathy, Autism, Conduct Disorder, facial mimicry, EMG, heart rate, skin conductance, eye-tracking, emotion recognition

ACM Classification Keywords

J.3 Life and medical sciences: Health, Medical information systems, J.4 Social and behavioral sciences: Psychology

INTRODUCTION

Autism spectrum disorders (ASD; including autism and Asperger syndrome) are characterized by atypical communication, impaired social interaction and restricted **Floor E. Scheepers**¹ f.scheepers@karakter.com

Jan K. Buitelaar^{1,2} j.buitelaar@psy.umcn.nl

²Department of Psychiatry UMC St. Radboud P.O. Box 9101, 6500HB Nijmegen, The Netherlands

repetitive patterns of behavior, interests and activities. Children with conduct disorder (CD) show a pattern of behavior violating the basic rights of others and ageappropriate norms and rules, which may develop in antisocial behavior in adulthood. At first sight these disorders appear to have little in common. However, lack of empathy is a core symptom in both ASD and CD.

Empathy is assumed to consist out of three components: motor, emotional and cognitive empathy [5]. Motor empathy refers to unconsciously mirroring the facial expressions of another, which is suggested to induce shared representations of perception and emotional contagion. Emotional empathy refers to the experience of emotions consistent with and in response to those of others. Cognitive empathy is the ability to rationally understand the emotional state of others. These three aspects of empathy have been frequently studied, mostly using facial expressions as stimuli, since facial expressions are essential in social communication and empathy. Our aim was to review the studies that investigated the three components of empathy in juveniles with ASD or CD and normal intelligence, in order to determine the overlap and specificity of empathic abilities in these disorders. Studies were included that used the 6 universal basic emotions; happy, angry, sad, fear, disgust and surprise [11]. In addition, eye-tracking studies are included, since attention to the eyes may play an essential role in facial emotion recognition as well as motor and emotional empathy, and possibly explaining impairments in these young patients.

MOTOR EMPATHY

Motor empathy is usually measured as facial mimicry using electromyography (EMG) electrodes to record muscle activity, on the cheek at the zygomator major muscle (smiling muscle) and corrugator supercilii muscle above at the inside of the eyes (frowning muscle). EMG amplitude in response to emotional faces is compared to a pre-stimulus

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

baseline level. Normal facial reaction patterns are increased zygomaticus activity during happy faces or positive pictures and increased corrugator activity during angry faces or negative images [10].

Motor Empathy in Autism

Recently, two studies measuring motor empathy in autism were conducted with children. Results were inconsistent, as were findings in adults. One study investigated spontaneous facial mimicry to emotional faces (happy, angry and fear) without using an emotion recognition task [3]. Children with ASD (n=11, age 7-13) did not show congruent rapid facial EMG responses to happy, angry or fearful faces. The second study [18], focused on timing of facial mimicry. No significant differences were found in emotion recognition or facial mimicry. However, facial mimicry appeared to be delayed in ASD (n=13, age 8-12) across all emotional expressions. This delay of facial mimicry in ASD might explain the inconsistent findings. When EMG response was measured within the first second after stimulus onset reduced facial mimicry in ASD was found [3]. Whereas measuring facial mimicry up to 2 seconds after stimulus onset showed no significant differences [18]. In addition, inconsistent results between studies may also be explained by using an emotion recognition task or measuring spontaneous responses. Indeed, it has to be noted that the study without a task and automatic facial EMG response within 1 second did find reduced facial mimicry in ASD [3]. In contrast, results of the other study using an emotion recognition task and EMG response measured up to 2 seconds after stimulus onset showed no significant difference between ASD and control subjects [18]. Therefore, voluntary imitation associated with active cognitive emotion processing may explain different results. Based on these results it seems to be that juveniles with ASD are impaired in motor empathy, at the level of spontaneous automatic facial mimicry (<1second). If they are cognitively processing and naming the emotion in an emotion recognition task they may use voluntary imitation in order to do so, which is delayed (1<second<2) as compared to automatic responses. The impairment in spontaneous automatic facial mimicry may be related to the lack of empathy in ASD, since empathy has been shown to be associated with facial mimicry in healthy individuals [20].

Motor Empathy in Conduct Disorder

Up to date there are only two reports of facial mimicry in 22 boys (age 8 - 12) with disruptive behavior disorder (DBD, including CD and oppositional defiant disorder, ODD). In the first report [8], film clips were shown with 2600ms dynamic happy or angry expression. Boys with DBD showed significantly reduced mean EMG amplitude corrugator response during the dynamic phase of angry faces, as compared to the control group. In addition, they scored lower on an emotional empathy self-report. In the second report [9], again reduced corrugator EMG reactivity was found in the DBD boys in response to both

sadness- and anger inducing film fragments from documentaries with a length of 58 to 158 seconds. No differences were found for happy facial expression or happiness inducing film clips. Motor empathy seems to be impaired in DBD for negative emotions, although only a few children with CD were included in these studies, therefore further research is needed in CD specifically.

EMOTIONAL EMPATHY

Emotional empathy is usually measured as heart rate (HR) or skin conductance response (SCR) to distressing or threatening pictures as compared to pleasant pictures. Though faces may be included, the stimuli are usually not emotional facial expressions only. Thus, these stimuli might provoke an emotional response and induce increased arousal, but not necessarily shared basic emotional experiences with others. A better indicator for emotional empathy would be the autonomic response to emotional facial expressions, which is discussed below.

Emotional Empathy in Autism

Autonomic response to facial emotional expressions specifically, has not been studied yet in ASD. No difference was found in SCR to distressing stimuli between juveniles with ASD and controls [4]. Further research is needed concerning emotional empathy in ASD, investigating autonomic response to emotional facial expressions of other people and sharing these emotions, rather than showing situations where just *one's own* emotion is triggered.

Emotional Empathy in Conduct Disorder

A reduced basal heart rate, increased heart rate reactivity and reduced basal SC was found associated with conduct problems and antisocial behavior [14,19]. Four studies investigated autonomic response in juveniles with CD or DBD to movie scenes. The presence of callous unemotional (CU) traits (e.g., lack of guilt and empathy, callous use of others) may be determining findings. CD children (n=33, age 7-11) with CU traits showed reduced basal HR and HR response to a scared boy, while no differences were found for CD children without CU traits (n=29) [1]. In two other studies HR response to film clips with sad people in DBD was investigated. Reduced HR response in DBD (n=22, age 8-12) was shown in one study [9], but no difference was found in the other study [16]. Using dynamic emotional faces, rather than movie scenes, no difference in HR response for the DBD (n=22, age 8-12) and control group was found [8]. Effects may be extinguished due to heterogeneity in the patient groups. DBD might include children with CU traits being autonomic hyporesponsive, hence showing reduced HR/SC response and lacking normal emotionality, as well as autonomic hyperresponsive children being highly emotional reactive.

COGNITIVE EMPATHY

Cognitive empathy can be measured with an emotion recognition task, using static or dynamic stimuli of emotional faces, or pictures of the eyes.

Cognitive Empathy in Autism

Emotion recognition abilities have been studied extensively in autism. Nevertheless, findings of studies on basic emotion recognition in juveniles with high functioning autism or Asperger are inconsistent. This may be the consequence of differences in methodology. Some of the studies analyzed recognition accuracy of all emotions separately [13], while other studies measured total accuracy of all basic emotions accumulated [7]. Although several databases for facial emotional expression were used, number of trials per emotion varied and accuracy analyses differ among studies; these factors did not explain the inconsistent findings. Interestingly, the majority of the reviewed studies did not find differences in basic emotion recognition between ASD and control groups, when verbal mental age and IQ were taken into account [21, 18]. In contrast, those studies with differences in verbal mental age or IQ, did report reduced emotion recognition in ASD [7, 13]. It seems that reduced recognition of the basic emotions in juveniles with ASD is probably explained by differences in verbal mental age or IQ, rather than diagnosis [23]. However, emotion recognition deficits may exist in ASD with normal IQ, for the more complex social emotions.

Cognitive Empathy in Conduct Disorder

Recently, a meta-analysis was conducted [15] considering 20 emotion recognition studies in antisocial individuals, with and without CU traits, including children and adults. Most studies used Ekman's [11] static facial stimuli and a multiple-choice response format. Results showed significant deficits in recognizing fear and sadness. Results were not related to CU traits, age or gender. Only a few studies have focused on conduct disorder or disruptive behavior disorder specifically, rather than antisocial behavior in general. Two studies were published after 2006 and therefore not included in this meta-analysis. These studies [22, 12] are in line with the results of the metaanalysis in the way that juveniles with CD have deficits in recognition of sad and/or fearful facial expressions, while no consistent relation with CU traits was found.

ATTENTION TO THE EYES

Eye gaze can be followed with an infrared eye-tracker device. Usually relative fixation time or fixation frequency is calculated for the areas of interest; the eyes and the mouth of the emotional faces.

Attention to the Eyes in Autism

In juveniles with ASD reduced attention to the eyes was found consistently for static and dynamic emotional faces [2, 7, 17]. Actually, only one study did not find reduced attention to the eyes in ASD [21]. The inconsistent finding in this particular study [21] could be explained by the fixation points counted from 100ms, while most other studies counted 20ms to 50ms as being fixation points or even just accumulated all time spent looking at the eyes. (However, see also [17].) The other explanation could be the fact that in this study [21] both ASD and controls groups had maximum scores on emotion recognition, which is probably related to eye fixation [2]. In addition, reduced emotion recognition was reported in studies with reduced eye fixation in ASD [7]. In boys with ASD eye fixation seems also to be associated with amygdala activity [7], which might indicate increased emotional responsiveness to the eyes. Thus, emotional *hyper*responsiveness might explain eye avoidance in individuals with ASD.

Attention to the Eyes in Conduct Disorder

No studies with CD juveniles and eye-tracking have been conducted. Only one eye-tracking study was done with juveniles concerning CU traits. It was reported that poor fear recognition in a community sample of 100 boys (age 8-15) was associated with higher CU traits, though not when explicitly instructed to look at the eyes [6]. Moreover, they found reduced time and frequency looking to the eyes in the group scoring high on CU traits. Further research is needed to investigate whether this applies to juveniles with conduct disorder as well, as they are particularly impaired in fear and sad expression recognition.

CONCLUSION

In juveniles with autism spectrum disorder (ASD) motor and cognitive empathy seem to be impaired: reduced mirroring of emotional faces was reported in ASD as well as impaired emotion recognition. The latter finding appears to be related to verbal developmental level. Emotional empathy, i.e. the autonomic response to emotional faces is not yet studied in ASD. Reduced attention to the eyes has consistently been found in ASD and could be related to emotional autonomic *hyper*responsivity resulting in eye contact avoidance.

In juveniles with conduct disorder (CD) cognitive empathy was found to be impaired, specifically sad and fear recognition being reduced. Motor and emotional empathy have not been studied in CD specifically. Although studies in a broader category of behavior disorders suggest reduced mirroring of emotional faces (motor empathy), further research is needed to investigate whether this is the same in CD specifically. Reduced autonomic response to emotional eliciting stimuli (emotional empathy) has been reported in CD, which may be related to callous unemotional traits, and impaired emotion recognition in these juveniles. No eyetracking studies have been conducted in CD yet, though reduced sad and fear recognition may be associated with lack of attention to the eyes.

In both ASD and CD emotional empathy needs to be further investigated and attention to the eyes may play an essential role. Importantly, major differences between the patient groups can be hypothesized. Whereas in autism eye contact might be avoided because of emotional *hyper*responsiveness, in CD a lack of attention to the eyes might be associated with emotional *hypo*responsiveness. In both disorders the outcome would be impaired emotional and cognitive empathy.While in some studies electrophysiological autonomic measurements and eyetracking were combined, it has not been measured *simultaneously* in response to emotional faces in ASD or CD. This would be the best suited method for investigating the association between attention to the eyes and emotional response to facial expressions. In addition, wireless electrophysiological equipment gives the possibility of obtaining measurements in real life or virtual reality situations. This could be a great step forward, since people with autism might be functioning much better in a structured test-setting than in chaotic unpredictable real life. Juveniles with CD may show social appropriate behavior in a test-setting, in contrast to real life following their personal benefits despite the harm of others.

Finally, psychiatric patients are of particular interest in order to investigate the underlying neurophysiological processes of empathy, because they are differentially impaired in the various subcomponents of empathy, allowing us to disentangle the motor, emotional and cognitive aspects of empathy. Besides a better understanding of empathy and emotional processing, this might provide opportunities for developing diagnostic or socio-emotional training tools for patients.

REFERENCES

- 1. Anastassiou-Hadjicharalambous X, Warden D. Physiologically-Indexed and Self-Perceived Affective Empathy in Conduct-Disordered Children High and Low on Callous-Unemotional Traits. *Child Psychiatry* & Human Development 2008;39(4):503-17.
- 2. Bal E, Harden E, Lamb D, Van Hecke AV, Denver JW, Porges SW. Emotion recognition in children with autism spectrum disorders: relations to eye gaze and autonomic state. *J Autism Dev Disord* 2010 March;40(3):358-70.
- 3. Beall PM, Moody EJ, McIntosh DN, Hepburn SL, Reed CL. Rapid facial reactions to emotional facial expressions in typically developing children and children with autism spectrum disorder. *J Exp Child Psychol* 2008 November;101(3):206-23.
- 4. Ben SD, Mostofsky SH, Hazlett RL et al. Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism. *J Autism Dev Disord* 2006 April;36(3):395-400.
- 5. Blair RJR. Responding to the emotions of others: dissociating forms of empathy through the study of typical and psychiatric populations. *Conscious Cogn* 2005 December;14(4):698-718.
- Dadds MR, El MY, Wimalaweera S, Guastella AJ. Reduced eye gaze explains "fear blindness" in childhood psychopathic traits. J Am Acad Child Adolesc Psychiatry 2008 April;47(4):455-63.
- Dalton KM, Nacewicz BM, Johnstone T et al. Gaze fixation and the neural circuitry of face processing in autism. *Nat Neurosci* 2005 April;8(4):519-26.

- de Wied M., van BA, Zaalberg R, Goudena PP, Matthys W. Facial EMG responses to dynamic emotional facial expressions in boys with disruptive behavior disorders. *J Psychiatr Res* 2006 March;40(2):112-21.
- de Wied M., Boxtel AV, Posthumus JA, Goudena PP, Matthys W. Facial EMG and heart rate responses to emotion-inducing film clips in boys with disruptive behavior disorders. *Psychophysiology* 2009 September;46(5):996-1004.
- Dimberg U, Thunberg M, Elmehed K. Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* 2000 Jan;11(1):86-9.
- 11. Ekman P, Friesen WV. *Pictures of facial effect.* Palo Alto, CA: Consulting Psychologists Press; 1976.
- 12. Fairchild G, Van Goozen SH, Calder AJ, Stollery SJ, Goodyer IM. Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. J Child Psychol Psychiatry 2009 May;50(5):627-36.
- 13. Kuusikko S, Haapsamo H, Jansson-Verkasalo E et al. Emotion recognition in children and adolescents with autism spectrum disorders. *J Autism Dev Disord* 2009 June;39(6):938-45.
- 14. Lorber MF. Psychophysiology of aggression, psychopathy, and conduct problems: a meta-analysis. *Psychol Bull* 2004 July;130(4):531-52.
- 15. Marsh AA, Blair RJ. Deficits in facial affect recognition among antisocial populations: a metaanalysis. *Neurosci Biobehav Rev* 2008;32(3):454-65.
- 16. Marsh P, Beauchaine TP, Williams B. Dissociation of sad facial expressions and autonomic nervous system responding in boys with disruptive behavior disorders. *Psychophysiology* 2008 January;45(1):100-10.
- 17. Norbury CF, Brock J, Cragg L, Einav S, Griffiths H, Nation K. Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. *J Child Psychol Psychiatry* 2009 July;50(7):834-42.
- 18. Oberman LM, Winkielman P, Ramachandran VS. Slow echo: facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. *Dev Sci* 2009 July;12(4):510-20.
- 19. Ortiz J, Raine A. Heart rate level and antisocial behavior in children and adolescents: a meta-analysis. *J Am Acad Child Adolesc Psychiatry* 2004 February;43(2):154-62.
- 20. Sonnby-Borgstrom M. Automatic mimicry reactions as related to differences in emotional empathy. *Scand J Psychol* 2002 December;43(5):433-43.
- 21. van der Geest JN, Kemner C, Verbaten MN, van EH. Gaze behavior of children with pervasive developmental disorder toward human faces: a fixation time study. *J Child Psychol Psychiatry* 2002 July;43(5):669-78.

- 22. Woodworth M, Waschbusch D. Emotional processing in children with conduct problems and callous/unemotional traits. *Child Care Health Dev* 2008 March;34(2):234-44.
- 23. Wright B, Clarke N, Jordan J et al. Emotion recognition in faces and the use of visual context in young people with high-functioning autism spectrum disorders. *Autism* 2008 November;12(6):607-26.

Mimicry as a Tool for Understanding the Emotions of Others

Marielle Stel m.stel@uu.nl +31 30 253 4705 Kees van den Bos k.vandenBos@uu.nl +31 30 253 3460

Department of Social and Organizational Psychology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

ABSTRACT

How do people understand what others are emotionally experiencing? We argue that mimicking the nonverbal expressions of other people (i.e., copying the behaviors of others) can be a tool for facilitating the understanding of the emotions that they are experiencing. People express their emotions nonverbally and when mimicking these nonverbal expressions, this affects the mimickers' emotions correspondingly due to an afferent feedback mechanism. As a result of this mechanism, the mimicker catches the emotions of others more strongly, which facilitates emotion understanding. Implications are discussed for affective computing.

Author Keywords

Mimicry, nonverbal expressions, emotional contagion, empathy, emotional understanding.

ACM Classification Keywords

H.5.m Information interfaces and presentation: Miscellaneous

INTRODUCTION

How do people understand what others are emotionally experiencing? We argue that one way to know which emotions other people experience is by mimicking their nonverbal expressions (i.e., copying the behaviors of others). As people often nonverbally express their emotions and as these nonverbal expressions are linked to emotions, mimicking these expressions leads to experiencing the same kind of emotions as the expressor of the emotions is experiencing. This sharing of emotions due to mimicry enhances the understanding for what other people are emotionally experiencing.

MIMICKING EMOTIONAL EXPRESSIONS

Emotions are expressed verbally as well as nonverbally. Emotions are nonverbally expressed, for instance, in the face, shoulders, movements, gestures, postures, and tone of voice. For example, when people feel angry, they often lower and bring together their eyebrows, clench a fist, speak louder, and have an active posture. These nonverbal expressions of emotions are often spontaneously mimicked by others. For example, it has been shown that newborns already have the tendency to spontaneously mimic happy and angry facial expressions of other people [e.g. 10]. Furthermore, it has been shown that people spontaneously mimic the facial expressions of persons shown on photographs: they activate the same facial muscles as the person on the picture [4].

Mimicry, often, is unintentional and occurs outside our awareness. It also occurs among complete strangers [e.g. 2]. One of the proposed functions of this mimicry is that it bonds people together: When mimicking or being mimicked, people feel more liking for the other person, feel more close to the other person, and feel that the interaction with the other person is more positive [e.g. 1, 2, 14]. In addition to this bonding function, mimicry serves as a means to create emotional understanding between people [e.g. 17].

FEEDBACK MECHANISM

An important mechanism in how mimicry can be a tool for emotional understanding is the feedback mechanism. The feedback mechanism entails that the muscles that are activated elicit emotions and cognitions that are associated with these muscles [3, 7, 19]. When moving your eyebrows inward and downward, activating the corrugator supercilii muscle which is associated with feeling angry, you will due to the feedback of this activated muscle — feel anger. The mechanism has been demonstrated often for facial expressions, vocal expressions, and postures [e.g. 6, 8]. So when people mimic the emotional expressions of others, they will experience the same kind of emotions due to the feedback of the activated muscles.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

EMPIRICAL EVIDENCE

This idea that mimicry leads to sharing each others emotions is also confirmed by research. For instance, it has been demonstrated that participants who observe a person showing facial expressions displayed on a video caught the emotions that were displayed by this person more strongly when mimicking these facial expressions [17]. This emotional contagion occurred for positive, as well as for negative emotional expressions. Furthermore, in a study in which two participants interacted face-to-face, it was demonstrated that mimicry led the two participants to experience more similar emotions [14]. More specifically, the results of this study showed that when mimicry occurred the emotions between the two interactants did not significantly differ, whereas when mimicry did not occur the two interactants experienced significantly different levels of happiness and anger. When the experienced emotions are used to infer what others are feeling, this enhances the understanding of others' emotions. A study confirming this idea showed that, as a result of mimicry. people who mimicked reported to understand the emotions and cognitions experienced by the other person better compared to people who did not mimic [15]. Thus, when one mimics the emotional expressions of another person, one becomes more emotionally attuned to this person, which leads to more emotion understanding for the person being mimicked.

In addition to mimicry facilitating catching and understanding the emotions of others, mimicry has been shown to facilitate the interpretation of affective signals [e.g. 13]. The emotions that are caught due to mimicking other people's facial expressions can serve as a propioceptive cue in the recognition of emotions. Indeed, women who were not constrained from mimicking positive or negative facial expressions from pictures were faster in recognizing the affective valence of these facial expressions than women who were constrained from mimicking [13]. This advantage of mimicry was not present for male participants. The gender difference is explained in that women are more emotionally expressive than men, and facial feedback may be more important in fast emotionrelated processing. The finding that gender differences exist in the facilitation of mimicry on emotion recognition contrasts findings of studies on emotional contagion that did not show gender differences in the effect of mimicry. This can be explained by a difference in the presentation of the stimuli. In the emotion recognition study [13] the facial expressions were very briefly presented. When having sufficient time to process emotions as in the emotional contagion studies [14, 15, 17], both men and women have an advantage of mimicry in processing emotions.

In sum, mimicry has been shown to facilitate feeling what the other person is feeling, leading to more emotion understanding. Although mimicry often occurs unintentional and spontaneous, the effects of intentional (instructed) mimicry do not differ from the effects of spontaneous (non-instructed) mimicry [e.g. 16]. Both types of mimicry (spontaneous or intentional) lead one to adopt the same expressions as their interaction partner, which – via the feedback of activated muscles– lead to experiencing the corresponding emotions. Thus, mimicry can be intentionally used as a tool for facilitating the understanding of the emotions experienced by others.

LIMITATIONS

At this point we would also like to discuss the limitations of using mimicry as a tool for emotion understanding. First of all, as mimicry enhances emotion understanding via the feedback mechanism, this means that mimicry does not serve as a tool for emotion understanding when having a different working feedback mechanism, or possibly, when having no feedback mechanism at all. This has been demonstrated by a study showing that participants with Autistic Spectrum Disorders (ASD) have a differently working facial feedback mechanism: Participants with ASD did not rate pictures of objects as more likeable when muscles were activated that correspond with positive emotions than when no emotional muscles were activated, whereas participants without ASD did rate the pictures as more likeable due to feedback of the muscles that correspond with positive emotions [16]. This deficit in the feedback mechanism for ASDs can explain why they experience a deficit in understanding the emotions of others. In part, ASDs' trouble with emotional understanding is due to a deficit in spontaneous mimicry: Participants with ASD were impaired in spontaneous mimicry [9]. However, simply instructing ASDs to intentionally mimic the expressions of others (of which they were capable of) did not lead to more emotion understanding because they have a different working feedback mechanism [16]. So mimicry can serve as a tool for emotional understanding, only when there is no impairment in the feedback mechanism.

Another limitation is that mimicry leads to more understanding of the emotions that are expressed by the other person. This means that when persons are less expressive of their emotions or try to mislead how they are truly feeling by expressing different emotions, mimicking does not help in understanding what the person may be emotionally experiencing. In the case of deception when a person intentionally tries to mislead another person by displaying emotions that are not truly felt, mimicry even leads to a worse understanding of how the person may be truly feeling. This is because via mimicry, you become to feel the emotions that are expressed by the other person. When you use these emotions to infer how another person might be feeling, you are less accurate than when you did not mimic these false expressions. This is demonstrated by a study showing that not mimicking (compared to spontaneous and instructed mimicry) the facial expressions of a liar led to more accurate estimations of how the person truly felt [18].

Furthermore, it is important to note that mimicry may not lead to an exact match in the extent to which mimickers and mimickees feel a certain mix of emotions. The mimicked expressions sent afferent feedback to the brain, which reenacts associated information, that is, information is reenacted that is associated with the muscles within the mimicker. This means that mimickers will feel the emotion that for them is associated with the activated muscles. Although a smile is associated with feeling positive for almost everyone, differences may exist in the strength of the association and the intensity of the emotion associated with the expression. For emotional expressions that differ between cultures, this implies that mimicking a person from another culture who shows culture-specific expressions does not lead to the re-enactment of associated emotions.

Finally, we do not, of course, argue that mimicry is the only way to understand the emotions of others. A lot of processes play a role in emotional understanding. We do argue, however, that mimicry is an important process that influences emotional understanding and, therefore, can be used as a tool to facilitate the understanding of the emotions experienced by others.

AFFECTIVE COMPUTING

Current knowledge on mimicry facilitating emotional understanding via facial feedback could be used in affective computing. The findings that mimicking the expressions of others lead to better emotional understanding imply that there are observable signals in the person's face that can facilitate emotion understanding when mimicking these emotions. In the field of affective computing, this means that computers could be more accurate in learning about the emotions of others by registering and translating the nonverbal expressions of a person. To be able to do so, the computer should first read the coordinates of the facial expressions and register any movements in them. This can be done, for instance, using a face tracker system [12]. The advantage of a tracker system over human perception is that computers are more accurate in 'observing' the expressions as humans do not attend to all facial cues and are less capable in reading expressions of mixed emotions. Secondly, a feedback mechanism should be built --similar to the mechanism of humans- that attach meaning to the movements. Many classification systems exist [e.g., 11], but the system is expected to be most accurate when based on a database of non-posed human expressions that are linked to what these people report to be truly experiencing at the moment of the expression.

Difficulties that could arise in accurately reading the emotions of others using a computer are similar to the difficulties that arise when using mimicry. First, the feedback mechanism that needs to be built linking expressions with emotions should be accurate. Although at this point expressions of the more basic emotions can be accurately classified, it is still difficult to read emotions that have an overlap in the facial features with other emotions [e.g., 11]. Finally, a computer should take into account the possibility that people may not be expressing what they are truly experiencing due to their motives in certain situations. This is a problem because motives are hard to be derived from expressions and context. Truly felt and false emotional displays, however, do differ in the muscles that are activated, as well as in the intensity, duration, laterality, and timing of this activation [e.g., 5]. These should be included in the classification system as well.

CONCLUSION

In conclusion, mimicking the expressions of others causes one to feel the emotions that are associated with these expressions, leading to more emotional understanding. Thus, when people are expressing what they are truly experiencing, mimicry can be used as a tool for a better understanding of the emotions of other people. This knowledge can be applied in the field of affective computing in which the challenge is to build an accurate feedback system that can link expressions with emotions and that can take into account context and individual motives.

REFERENCES

- Ashton-James, C.E., Van Baaren, R.B., Chartrand, T.L., Decety, J., & Karremans, J. Mimicry and me: The impact of mimicry on self-construal. *Social Cognition*, 25 (2007), 518-535.
- 2. Chartrand, T.L., & Bargh, J.A. The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76 (1999), 893-910.
- 3. Darwin, C. *The expressions of the emotions in man and animals.* Chigago: University of Chigago Press, 1965. (Original work was published in 1872).
- 4. Dimberg, U. Facial electromyography and emotional reactions. *Psychophysiology*, 27 (1990), 481-494.
- Ekman, P., Friesen, W.V., & O'Sullivan, M. Smiles when lying. *Journal of Personality and Social Psychology*, 54 (1988), 414–420.
- 6. Hess, U., Kappas, A., McHugo, G. J., Lanzetta, J. T., & Kleck, R. E. The facilitative effect of facial expression on the self-generation of emotion. *International Journal of Psychophysiology*, *12* (1992), 251-265.
- 7. Izard, C. E. Human emotions. New York: Plenum, 1977.
- McIntosh, D. N. Facial feedback hypotheses: Evidence, implications, and directions. *Motivation and Emotion*, 20 (1996), 121-147.
- McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. L. When the social mirror breaks: Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Developmental Science*, 9 (2006), 295–302.

- Reissland, N. Neonatal imitation in the first hour of life: Observations in rural Nepal. *Developmental Psychology*, 24 (1988), 464-469.
- Sebe, N. Cohen, I. Huang, T.S. Handbook of Pattern Recognition and Computer Vision. World Scientific, Singapore, 2005.
- Stel, M., Van Baaren, R., Blascovich, J., McCall, C., Pollmann, M.H., van Leeuwen, Mastop, J., M., & Vonk, R. Effects of a priori liking on the elicitation of mimicry. *Experimental Psychology* (in press).
- Stel, M. & Van Knippenberg, A. The role of facial mimicry in the recognition of affect. *Psychological Science*, 19 (2008), 984-985.
- 14. Stel, M. & Vonk, R. Mimicry in social interaction: Benefits for mimickers, mimickees and their interaction. *British Journal of Psychology*, 101 (2010), 311-323.

- 15. Stel, M., Vonk, R., van Baaren, R.B., & Smeets, R.C. *The social consequences of mimicry: Effects on empathy and bonding.* Manuscript submitted for publication.
- 16.Stel, M., Heuvel, van den, C., & Smeets, R.C. Facial feedback mechanisms in Autistic Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 38 (2008), 1250–1258.
- 17.Stel, M., Van Baaren, R.B., & Vonk, R. Effects of mimicking: acting prosocially by being emotionally moved. *European Journal of Social Psychology*, 38 (2008), 965-976.
- 18. Stel, M., Van Dijk, E., & Olivier, E. You want to know the truth? Then don't mimic! *Psychological Science*, 20 (2009), 693-699.
- Tomkins, S.S. Affect theory. In P. Ekman (Ed.), *Emotion in the human face* Cambridge: Cambridge University Press (2nd ed., 1982), 353-395.

Social Signal Processing: Understanding Nonverbal Communication in Social Interactions

Alessandro Vinciarelli

University of Glasgow / Idiap Research Institute Sir A.Williams Bldg – G12 8QQ Glasgow (UK) CP592 – 1920 Martigny (Switzerland) vincia@dcs.gla.ac.uk Fabio Valente

Idiap Research Institute CP592 – 1920 Martigny (Switzerland) fvalente@idiap.ch

ABSTRACT

This paper provides a short overview of Social Signal Processing, the domain aimed at bridging the social intelligence gap between people and machines. The focus of Social Signal Processing is on nonverbal behavioral cues that human sciences (psychology, anthropology, sociology, etc.) have identified as conveying social signals, i.e. relational attitudes towards others and social situations. The rationale is that such cues are the physical, machine detectable and synthesizable evidence of phenomena nonotherwise accessible to computers such as empathy, roles, dominance, personality, (dis-)agreement, interest, etc. After providing a brief state-of-the-art of the domain, the paper outlines its future perspectives and some of its most promising applications.

Author Keywords

Social Signal Processing, human-human communication, nonverbal behavior, social interactions.

ACM Classification Keywords

A.1 Introductory and survey, H.1.2 User/ Machine Systems, I.2 Artificial intelligence, I.2.10 Vision and scene understanding, J.4 Social and behavioral sciences.

INTRODUCTION

Several decades of research in human sciences have shown that nonverbal communication is the main channel through which we express *social signals* [3], i.e. our relational attitudes (e.g., sympathy, interest, hostility, agreement, etc.) towards others and social situations. Nonverbal communication is the wide spectrum of nonverbal behavioral cues (e.g., facial expressions, vocalizations, postures, gestures, appearance, etc.) that we display when we interact with others [10], with machines [8] and with media [11]. From a computing point of view, this is important for two reasons: The first is that nonverbal behavioral cues play the role of a physical, hence machine detectable evidence of social signals. The second is that nonverbal cues synthesized through some form of embodiment (conversational agents, robots, etc.) express the same relational attitudes as when they are displayed by humans [4], thus are likely to synthesize social signals [8].

Social Signal Processing (SSP) relies on the above to bridge the social intelligence gap between humans and machines [13,15]. Social intelligence [13] is the facet of our cognitive abilities that aims at dealing effectively with social interactions and, at its core, it includes two main aspects: The first is the correct interpretation, in terms of social signals, of nonverbal behavioral cues displayed by others. The second is the generation of nonverbal cues expressing social signals appropriate in a given situation. In other words, SSP brings social intelligence in machines via *modeling, analysis* and *synthesis* of nonverbal behavior in social interactions [13,15].

Correspondingly, SSP addresses three main research questions:

- Is it possible to detect automatically nonverbal behavioral cues in data captured with suitable sensors (e.g., microphones and cameras)?
- Is it possible to automatically infer social signals from nonverbal behavioral cues detected through sensors?
- Is it possible to synthesize social signals for embodiment of social behaviors in artificial agents, robots or other devices?

As people tend to display the same behavioral cues when they interact with others and with machines [8], the above questions are relevant to HCI scenarios as well. Furthermore, behavioral cues often convey affective information, thus progress in SSP can contribute to Affective Computing as well (and viceversa).

While still in its early and pioneering stages, SSP has attracted significant attention in both scientific and business communities. The SSP state-of-the-art is rich and constantly

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

expands towards new research directions, but the domain is still characterized by high entry barriers, in particular the need of large annotated corpora and software tools covering a wide spectrum of functionalities (e.g., facial expression analysis, prosody extraction, data annotation, etc.). In this respect, a European collaboration called SSPNet (Social Signal Processing Network) is building an extensive online repository (www.sspnet.eu) of articles, data and software tools. The goal is to smooth the entry barriers and allow any potentially interested researcher to start working on SSP [12].

The rest of this paper proposes a short survey of works where turn-taking analysis is used to infer social phenomena (a good example of how SSP works), outlines some future perspectives from both scientific and application points of view, and draws some conclusions.

STATE-OF-THE-ART: THE TURN-TAKING EXAMPLE

Extensive surveys of SSP, at least for the analysis component, are available in [13,15]. Due to space limitations, it is not possible to consider all possible problems and modalities. Thus, this section focuses on turn-taking, the cue that so far has given the most satisfactory results in the themes most commonly addressed in SSP.

Current efforts aim at a systematic and rigorous definition of social signals as well as at the identification of behavioral variables to be taken into account in automatic analysis and synthesis of social signals [3]. Furthermore, several works explore the possibility of using computational approaches to validate psychological findings like, e.g., the impact of facial features on the perception of personality traits, or the effect of depression on nonverbal cues. This kind of works is particularly interesting because it closes the loop between human and computing sciences: on one hand, computational approaches integrate human sciences findings to automatically analyze nonverbal behavior, on the other hand, human sciences apply computational approaches to confirm and assess their findings.

On the analysis side, the state-of-the-art concentrates on interactions in small groups, the most common and primordial forms of social exchange [7]. The most extensively addressed problem is the recognition of roles people play in different situations, including radio and television programs, where the setting is highly formal and roles correspond to specific tasks (e.g., *anchorman* or *guest*) [14], and spontaneous meetings, where roles correspond to social functions (e.g., *attacker* or *supporter*) [16].

Automatic role recognition is mostly based on the analysis of turn-taking patterns, i.e. *on who talks when, to whom and how much*. Thus, the first step of the process is typically the application of a speaker clustering approach that segments the audio channel of the interaction recordings into time

intervals expected to correspond to an individual voice. In other words, speaker clustering techniques identify *turns*, i.e. time segments during which one person talks and the others listen to her. In the meantime, each turn is automatically assigned a label corresponding to a speaker so that the process not only identifies the points where the speaker changes, but also what are the turns during which each speaker talks.

Speaker clustering techniques are typically based on agglomerative clustering approaches that group vectors of acoustic observations, extracted at regular time steps (typically once every 10 milliseconds), based on their similarity, i.e. on how likely they are to belong to the same voice, and on their temporal proximity, i.e. on how likely they are to belong to the same turn. Agglomerative clustering techniques are iterative approaches where, at each step, the two most similar clusters are merged. The process is continued until a model selection criterion (typically the Bayesian Information Criterion) is met. Two clusters are considered similar when, after having been merged, the fitness of the clustering to the data (typically measured in terms of log likelihood) improves.

Once the speaker clustering has been performed, the actual role recognition step can take place using two main approaches. The first is to represent the turn-taking pattern of each person with a feature vector and then to map this last into one of the roles using a classifier (Support Vector Machines, Neural Networks, Bayesian Networks, etc.). The second is to extract a feature vector from each turn and then to align the resulting sequence of observations with a sequence of roles using probabilistic sequential models (Hidden Markov Models, Dynamic Bayesian Networks, Conditional Random Fields, etc.).

In the first case, the most common features are the number of times a given speaker talks, the fraction of total conversation time a speaker talks for, how many adjacent turns each speaker has with all of the others, what is the centrality of each speaker (i.e. how many times the speaker talks between each pair of two other speakers), etc. In the second case, the most common features aim at capturing sequential aspects such as the number of times a given sequence of speakers is observed, how many turns there are between two consecutive interventions of the same speaker, etc.

There is no evidence about what approach is better, but sequential approaches seem to be more promising because they allow one to assign different roles to the same person in the course of the same interaction. This is an important requirement when considering roles inspired by social theories (like those proposed by Bales in his works on small groups), more general than those related to specific scenarios considered so far like broadcast data and meetings. Furthermore, some variants of the most common probabilistic sequential approaches, e.g. Factorial Hidden Markov Models, Influence Models, Layered Hidden Markov Models, Latent Conditional Random Fields, Dynamic Bayesian Networks, etc., allow one to model multiple streams of observations which might correspond to several persons (particularly suitable to investigate how people react to one another) as well as to behavioral cues extracted from multiple modalities (particularly suitable to study how multiple cues concur to convey the same social signal).

Besides roles, other phenomena often investigated include conflict and disagreement [2], given the disruptive impact they can have on the life of a group, and recognition of dominant individuals, given the impact these have on the group outcome [6]. In these cases as well, the turn-taking component plays a major role (see above), but more behavioral cues are taken into account such as facial expressions, gaze behavior and movement. This requires to jointly analyze multiple modalities, a major problem when the behavioral cues expressed in each of the modalities take place at different time-scales like, for example, facial expressions (half a second to a second) and turns (few seconds to some minutes).

Two main approaches have been followed in these cases. The first is called early fusion and simply consists of using a single probabilistic sequential model (see above) fed with observations extracted from multiple modalities. This is based on the assumption, often unrealistic, that the processes taking places in different modalities are lockstep, i.e. that observations resulting from different modalities are always determined by the same hidden, underlying state. The second approach, called late fusion, analyzes processes taking place in different modalities separately and then fuses the output of the models with classifier combination approaches. This approach is based on the assumption, once again unrealistic, that behaviors captured through different modalities are independent.

While being based on patently wrong assumptions, both early and late fusion approaches lead to satisfactory results even if, in general, one modality alone is responsible for most of the performance, the others simply bringing small relative improvements.

FUTURE PERSPECTIVES

The state-of-the-art addresses a long list of social phenomena, but new directions are still emerging in SSP research.

On the short term, research efforts explore aspects of nonverbal behavior that, while having been extensively investigated in human sciences, have been neglected in the computing community. One example is the use of space and environment to express social relational messages [5]. This includes analysis and synthesis of socially relevant information from the spatial configurations (called Fformations) people assume during interactions, the inference of social distance from physical distance, analysis of territoriality, social behavior in surveillance and monitoring scenarios, etc. Other examples are the multimodal generation of spontaneous behavior, or the simulation of subtle behavioral phenomena like mirroring and phonetic convergence.

Signal processing and machine learning approaches applied so far are not necessarily suitable and major challenges must be faced. In particular, the state-of-the-art has concentrated on small groups (four to six participants), but many important scenarios involve larger numbers of individuals. This is likely to require approaches focusing less on the detailed behavior of each individual and more on collective phenomena. This might lead towards Social Network Analysis like techniques where no prediction is made at the individual level, and analysis is possible only in terms of presence of social groups, detection of prominent individuals (in terms of the number of network paths passing through them), overall connectedness of the network, etc.

The perspectives are rich in terms of potential applications as well. Multimedia indexing is likely to profit from SSP because social interactions are one of the main channels through we access reality and to index the data in terms of social interaction means to make retrieval approaches closer to our perception of data content. Healthcare applications, especially when it comes to mental problems or cognitive deterioration due to ageing and related problems, can apply SSP to identify subtle symptoms in the first stages of illnesses. Human Computer Interaction can be improved thanks to the introduction of socially adept technologies capable of dealing with users like humans deal with other people. Computer mediated communication can profit from SSP by allowing the detection and transmission of those nonverbal cues like gaze that most contribute to the naturalness of face-to-face interactions.

CONCLUSIONS

This paper has provided a short introduction to Social Signal Processing, including an overview of its main principles and goals, a description of some approaches based on turn-taking (one of the nonverbal cues most commonly applied in SSP), and some of the most promising research perspectives currently emerging in the community.

Social aspects of human behavior attract attention in many different areas because they seem to provide an explanation for many experimental observations not otherwise understandable. For example, neurosciences have identified social interaction and learning through imitation as the main goal of mirror neurons, physiology has shown that our ears are tuned to human voices more than to any other sound to maximize the chance of social contacts, some psychology theories explain the existence of stable personality traits as a means to ensure predictability in social exchanges, ethology recognizes social interaction as one of the main reasons behind observable behaviors, and the list could continue.

Computing sciences could not be immune to such a wave of interest. Nowadays, computers are much more than an improved version of old tools (like word processors used to be with respect to typewriters), they are the platform through which we communicate, we entertain ourselves, we shop, we join and form large communities of interest, etc. Furthermore, computers are at the core of technologies expected to seamlessly integrate our everyday life such as smart ambients, robots, smart interfaces and, more in general, human centered technologies [9], i.e. technologies dealing with their users following the natural interacting modes of humans.

As social interactions, and their non-verbal component in particular, are such a natural aspect of our behavior, SSP technologies, centered around non-verbal communication, are likely to bring a major improvement in all of the scenarios outlined above. Last, but not least, computational approaches for behavior understanding appear to be an instrument helping human sciences to better understand human behavior. This in turn might further improve SSP and open a cycle where human and computing sciences are not only integrated, but also mutually supporting.

ACKNOWLEDGMENTS

This work has been funded by the EC (FP7/2007-2013) under grant agreement No. 231287 (SSPNet), and by the Swiss National Science Foundation through the National Center of Competence in Research IM2.

REFERENCES

- 1. Albrecht, K. Social Intelligence: The new science of success. John Wiley & Sons Ltd, 2005.
- 2. Bousmalis, K., Mehu, M. and Pantic, M. Spotting Agreement and Disagreement: A Survey of Nonverbal Audiovisual Cues and Tools. Proceedings of IEEE Workshop on Social Signal Processing, 2009.
- 3. Brunet, P.M., Donnan, H., McKeown, G., Douglas-Cowie, E. and Cowie, R. Social signal processing: What are the relevant variables? And in what ways do they relate? Proceedings of the *IEEE International Workshop* on Social Signal Processing (2009).
- 4. Cassell, J. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine*, 22, 4 (2001), 67-83.
- 5. Cristani, M., Murino, V. and Vinciarelli, A. Socially Intelligent Surveillance and Monitoring: Analysing

Social Dimensions of Physical Space. In *Proceedings of International Workshop on Socially Intelligent Surveillance and Monitoring* (2010), to appear.

- 6. Hung, H. and Jayagopi, D. and Yeo, C. and Friedland, G. and Ba, S. and Odobez, J.M. and Ramchandran, K. and Mirghafori, N. and Gatica-Perez, D. Using audio and video features to classify the most dominant person in a group meeting. In *Proceedings of the ACM International Conference on Multimedia* (2007), 838-841.
- Levine, J. and Moreland, R. Small groups. In *The* handbook of social psychology, D. Gilbert and G. Lindzey (eds.). Oxford University Press (2008), 415-469.
- 8. Nass, C. and Brave, S. Wired for Speech. How Voice Activates and Advances the Human-Computer Relationship. MIT Press, USA, 2005.
- 9. Pantic, M. and Nijholt, A. and Pentland, A. and Huang, T.S. Human-Centred Intelligent Human Computer Interaction (HCI2): how far are we from attaining it? *International Journal of Autonomous and Adaptive Communications Systems*, 1, 2 (2008), 168-207.
- 10. Poggi, I. Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication. Weidler Buchverlag Berlin, 2007.
- 11. Reeves, B. and Nass, C. The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press New York, NY, USA, 1996.
- 12. Vinciarelli, A., Pantic, M. *sspnet.eu, a web portal for Social Signal Processing.* IEEE Signal Processing Magazine, to appear, (2010).
- Vinciarelli, A., Pantic, M. and Bourlard, H. Social Signal Processing: Survey of an Emerging Domain. *Image and Vision Computing Journal* 27, 12 (2009), 1743-1759.
- Vinciarelli, A. Capturing Order in Social Interactions. IEEE Signal Processing Magazine 26, 12 (2009), 133-137.
- 15. Vinciarelli, A., Pantic, M., Bourlard, H. and Pentland, A. Social Signal Processing: State-of-the-art and future perspectives of an emerging domain. In *Proceedings of the ACM International Conference on Multimedia* (2008), 1061-1070
- 16. Zancanaro, M., Lepri, B. and Pianesi, F. Automatic detection of group functional roles in face-to-face interactions. *In Proceedings of the International Conference on Multimodal Interfaces*, (2006), 28–31.

Automatic Measurement of Affect in Dimensional and Continuous Spaces: Why, What, and How?^{*}

Hatice Gunes

Imperial College London, U.K. University of Technology Sydney, Australia h.gunes@imperial.ac.uk

ABSTRACT

This paper aims to give a brief overview of the current state-of-the-art in automatic measurement of affect signals in dimensional and continuous spaces (a continuous scale from -1 to +1) by seeking answers to the following questions: i) *why* has the field shifted towards dimensional and continuous interpretations of affective displays recorded in real-world settings? ii) *what* are the affect dimensions used, and the affect signals measured? and iii) *how* has the current automatic measurement technology been developed, and *how* can we advance the field?

Author Keywords

Automatic measurement of human affect, dimensional and continuous affect recognition, multicue and multimodal affect recognition.

ACM Classification Keywords

A.1 Introduction and survey, H.1.2 User/machine systems: Human information processing, I.5.4 Pattern recognition applications

WHY MEASURE AFFECT IN DIMENSIONAL SPACES?

According to research in psychology, three major approaches to affect modelling can be distinguished [10]: categorical, dimensional, and appraisal-based approach. The categorical approach claims that there exist a small number of emotions that are basic, hard-wired in our brain, and recognized universally (e.g. [5]). This theory on universality and interpretation of affective nonverbal expressions in terms of basic emotion categories has been the most commonly adopted approach in research on automatic measurement of human affect.

Maja Pantic

Imperial College London, U.K. University of Twente, The Netherlands m.pantic@imperial.ac.uk

However, a number of researchers have shown that in everyday interactions people exhibit non-basic, subtle and affective states like rather complex thinking, embarrassment or depression. Such subtle and complex affective states can be expressed via dozens of anatomically possible facial and bodily expressions, audio or physiological signals. Therefore, a single label (or any small number of discrete classes) may not reflect the complexity of the affective state conveyed by such rich sources of information [23]. Hence, a number of researchers advocate the use of dimensional description of human affect, where affective states are not independent from one another; rather, they are related to one another in a systematic manner (e.g., [10, 23, 24]).

It is not surprising, therefore, that automatic affect sensing and recognition researchers have recently started exploring how to model, analyse and interpret the subtlety, complexity and continuity (represented along a continuum from -1 to +1, without discretisation) of affective behaviour in terms of latent dimensions, rather than in terms of a small number of discrete emotion categories.

The most widely used dimensional model is Russell's twodimension 'circumplex model of affect', where emotions are seen as combinations of arousal and valence [23].

Scherer and colleagues introduced another set of psychological models, referred to as componential models of emotion, which are based on appraisal theory [7, 10, 24]. In the appraisal-based approach emotions are generated through continuous, recursive subjective evaluation of both our own internal state and the state of the outside world (relevant concerns/needs) [7, 8, 10, 24]. How to use the appraisal-based approach for automatic measurement of affect is an open research question as this approach requires multicomponential complex, and sophisticated measurements of change. One possibility is to reduce the appraisal models to dimensional models (e.g., 2D space of arousal-valence).

Another model, known as OCC [19] is also established as a standard cognitive appraisal model for emotions, and has mostly been used in affect synthesis (in embodied conversational agent design).

^{*}An extended version of this paper has been published in the International Journal of Synthetic Emotions, in Jan. 2010.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

An individual's inner emotional state may become apparent by subjective experiences (how the person feels), external/outward expressions (audio/visual signals), and internal/inward expressions (bio signals). However, these may be incongruent, depending on the context (e.g., feeling angry and not expressing it outwardly). This poses a true challenge to automatic sensing and analysis.

Currently, a number of affect recognisers attempt to label both the felt (e.g., [1, 17]) and the internally/externally expressed (e.g., [14, 15]) emotions.

Affect Dimensions

Despite the existence of the abovementioned emotion models, in automatic measurement of dimensional and continuous affect, *valence* (how positive or negative the affect is), *activation* (how excited or apathetic the affect is), *power* (the sense of control over the affect), and *expectation* (the degree of anticipating or being taken unaware) appear to make up the four most important affect dimensions [7]. Although ideally the intensity dimension could be derived from the other dimensions, to guarantee a complete description of affective colouring, some researchers include *intensity* (how far a person is away from a state of pure, cool rationality) as the fifth dimension. However, search for optimal low-dimensional representation of affect remains open [7].

Visual Signals

Facial actions (e.g., pulling eyebrows up) and facial expressions (e.g., producing a smile), and to a much lesser extent bodily postures (e.g., backwards head bend and arms raised forwards and upwards) and expressions (e.g., head nod), form the widely known and used visual signals for automatic affect measurement. Dimensional models are considered important in this task as a single label may not reflect the complexity of the affective state conveyed by a facial expression, body posture or gesture. Ekman & Friesen [6] considered expressing discrete emotion categories via face, and communicating dimensions of affect via body as more plausible. A number of researchers have investigated how to map various visual signals onto emotion dimensions. For instance, [23] mapped the facial expressions to various positions on the two-dimensional plane of arousal-valence, while [4] investigated the emotional and communicative significance of head nods and shakes in terms of arousal and valence dimensions, together with dimensional representation of solidarity, antagonism and agreement.

Audio Signals

Audio signals convey affective information through explicit (linguistic) messages, and implicit (acoustic and prosodic) messages that reflect the way the words are spoken. There exist a number of works focusing on how to map audio expression to dimensional models. Cowie et al. used valence-activation space (similar to valence-arousal) to model and assess affect from speech [2, 3]. Scherer and colleagues have also proposed how to judge emotion effects on vocal expression, using the appraisal-based theory [10].

Bio Signals

The bio-signals used for automatic measurement of affect are galvanic skin response that increases linearly with a person's level of arousal [1], electromyography (frequency of muscle tension) that is correlated with negatively valenced emotions [13], heart rate that increases with negatively valenced emotions such as fear, heart rate variability that indicates a state of relaxation or mental stress, and respiration rate (how deep and fast the breath is) that becomes irregular with more aroused emotions like anger or fear [1, 13]. Measurements recorded over various parts of the brain including the amygdala also enable observation of the emotions felt [22]. For instance, approach or withdrawal response to a stimulus is known to be linked to the activation of the left or right frontal cortex, respectively. A number of studies also suggest that there exists a correlation between increased blood perfusion in the orbital muscles and stress levels for human beings. This periorbital perfusion can be quantified through the processing of thermal video (e.g., [26]).

HOW IS THE CURRENT TECHNOLOGY DEVELOPED?

Data Acquisition and Annotation

Cameras are used for acquisition of face and bodily expressions, microphones are used for recording audio signals, motion capture systems are utilized for recording 3D affective postures and gestures, and thermal (infrared) cameras are used for recording blood flow and changes in skin temperature. In the bio-signal research context, the subject being recorded usually wears a headband or a cap on which electrodes are mounted, a clip sensor, or touch type electrodes. The subject is then stimulated with emotionally-evocative images or sounds. Acquiring affect data without subjects' knowledge is strongly discouraged and the current trend is to record spontaneous data in more constrained conditions such as an interview setting, where subjects are still aware of placement of the sensors and their locations.

Annotation of the affect data is usually done separately for each modality assuming independency between the modalities. A major challenge is the fact that there is no coding scheme that is agreed upon and used by all researchers in the field that can accommodate all possible communicative cues and modalities. In general, the annotation tool Feeltrace is used for annotating the external expressions (audio and visual signals) with *continuous* traces (impressions) in dimensional spaces. Feeltrace allows observers to watch an audio-visual recording and move their cursor within the affect space to rate their impression about the affective state of the subject [2]. For annotating the internal expressions (bio signals), the level of valence and arousal is usually extracted from subjective experiences (subjects' own responses) (e.g., [17, 22]) due to the fact that feelings induced by an image can be very different from subject to subject. When discretised dimensional annotation is adopted (as opposed to continuous one), researchers seem to use different intensity levels: either a ten-point Likert scale (e.g., 0-low arousal, 9-high arousal) or a range between -1.0 and 1.0 (divided into a number of levels) [11]. The final annotation is usually calculated as the mean of the observers' ratings. Development of an easy to use, unambiguous and intuitive annotation scheme that is able to incorporate inter-observer agreement levels remains an important challenge in the field.

Obtaining high inter-observer agreement is another challenge in affect data annotation, especially when (continuous) dimensional approach is adopted. To date, researchers have mostly chosen to use self-assessments (subjective experiences, e.g. [13]) or the mean (within a predefined range of values) of the observers' ratings (e.g. [16]). Modelling inter-observer agreement levels within automatic affect analyzers, and finding which signals better correlate with self assessments and which ones better correlate with independent observer assessments remains as a challenging issues in the field.

Automatic Measurement of Affect in Continuous Spaces

After affect data has been acquired and annotated, representative and relevant features need to be extracted prior to the automatic measurement of affect in dimensional and continuous spaces. The feature extraction techniques used for each communicative source are similar to the previous works (reviewed in [12]) adopting a categorical approach to affect recognition.

There are a number of additional issues which need to be taken into account when applying a dimensional approach to affect recognition.

The interpretation accuracy of expressions and physiological responses in terms of continuous emotions is very challenging. While visual signals appear to be better for interpreting valence, audio signals seem to be better for interpreting arousal [11]. A thorough comparison between all modalities would indeed provide a better understanding of which emotion dimensions are better recognised from which modalities (or cues).

The window size to be used to achieve optimal affect recognition is another issue that the existing literature does not provide a unique answer to. Current affect recognizers employ various window sizes depending on the modality, e.g., 2-6 seconds for speech, 3-15 seconds for bio-signal [15]. There is no consensus on how the efficiency of such a choice should be evaluated.

Measuring the intensity of expressed emotion appears to be modality dependent. The way the intensity of an emotion is apparent from physiological data may be different than the way it is apparent from visual data. Moreover, little attention has been paid so far to whether there are definite boundaries along the affect continuum to distinguish between various levels or intensities. Currently intensity is measured by quantizing the affect dimensions into arbitrary number of levels such as neutral, low and high (e.g., [16, 17, 27]). Separate models are then built to discriminate between pairs of affective dimension levels, for instance, low vs. high, low vs. neutral, etc. Generalizing intensity analysis across different subjects is a challenge yet to be researched as different subjects express different levels of emotions in the same situation.

The Baseline problem is another major challenge in the field. For tactile modality (bio signals) this refers to the problem of finding a condition against which changes in measured physiological signals can be compared (a state of calmness). For audio modality this is usually achieved by segmenting the recordings into turns using energy based voice activity detection and processing each turn separately. For visual modality the aim is to find a frame in which the subject is expressionless and against which changes in subject's motion, pose, and appearance can be compared. This is achieved by manually segmenting the recordings, or by constraining the recordings to have the first frame containing a neutral expression. However, expecting expressionless state in each recording or manually segmenting recordings so that each segment contains a baseline expression are strong, unrealistic constrains for analysis and processing of affective information.

Feature space with high dimensionality hinders automatic affect measurement. For instance, various works in the field have reported that they extract 2,520 features for each frame of an input facial video, 4,843 features for each utterance, 16,704 EEG features for each stream etc. (see [11] for details). Having fewer training samples than features per sample impedes the learning of the target classification. Various dimensionality reduction or feature selection techniques have been applied (e.g., Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA), kernel PCA (KPCA) Sequential Backward Selection) to mitigate this problem. Creating dimensionality reduction techniques with specific applications to automatic measurement of affect in dimensional and continuous spaces remains as an issue to be explored.

Generalisation capability of automatic affect analysers across subjects is still a challenge in the field. Kulic & Croft [17] reported that for bio-signal based affect measurement subjects seem to vary not only in terms of response amplitude and duration, but for some modalities, a number of subjects show no response at all. This makes generalisation over unseen subjects a very difficult problem. When it comes to other modalities, most of the works in the field report only on subject dependent dimensional affect measurement and recognition due to limited number of subjects and data (e.g., [27]).

Modality fusion refers to combining and integrating all incoming unimodal events into a single representation of the affect expressed by the user. When it comes to

integrating multiple modalities, the major issues are: i) when to integrate the modalities (at what abstraction level to do the fusion), ii) how to integrate the modalities (which criteria to use), iii) how to deal with the increased number of features due to fusion, iv) how to deal with the asynchrony between the modalities (e.g., video is recorded at 25 Hz, audio is recorded at 48 kHz while EEG is recorded at 256-512 Hz), and v) how to proceed with fusion when there is conflicting information conveyed by the modalities. Despite a number of efforts in the discrete affect recognition field (reviewed in [12]), these issues remain yet to be explored for dimensional and continuous affect recognition.

Classification methods used for dimensional and continuous affect measurement should be able to produce continuous values for the target dimensions. Some of the classification schemes that have been explored for this task are, namely, Support Vector Regression (SVR), Conditional Random Fields (CRF), and Long Short-Term Memory Recurrent Networks (LSTM-RNN). Overall, there is no agreement on how to model dimensional affect space (continuous vs. quantised) and which classifier is better suited for automatic, multimodal, continuous affect analysis using a dimensional representation. The design of emotion-specific classification schemes that can handle multimodal and spontaneous data is one of the most important issues in the field.

Evaluation measures applicable to categorical affect recognition are not directly applicable to dimensional approaches. Using the Mean Squared Error (MSE) between the predicted and the actual value of arousal and valence, instead of the recognition rate (i.e., percentage of correctly classified instances) is the most commonly used measure by related work in the literature (e.g., [14, 27]). However, using MSE might not be the best way to evaluate the performance of dimensional approaches to automatic affect measurement and recognition. Therefore, the correlation coefficient, that evaluates whether the model has managed to capture patterns inhibited in the data at hand, is also employed by several studies (e.g., [14, 18]) together with MSE. Overall, however, how to obtain optimal evaluation metrics for continuous and dimensional emotion recognition remains an open research issue [11].

HOW CAN WE ADVANCE THE FIELD?

The analysis provided in this paper indicates that the automatic affect sensing field has slowly started shifting from categorical (and discrete) affect recognition to dimensional (and continuous) affect recognition to be able to capture the complexity of affect expressed in *real* world settings, by the *real* people. Despite the existence of a number of dimensional emotion models, the two-dimensional model of arousal and valence appears to be the most widely used model in automatic measurement from audio, visual and bio signals. The current automatic measurement technology has already started dealing with

spontaneous data obtained in less-controlled environments using various sensing devices, and exploring a number of machine learning techniques and evaluation measures. However, real-world settings pose many challenges to continuous affect sensing and recognition (e.g., when subjects are not restricted in terms of mobility, the level of noise in all recorded signals tends to increase).

To date, only a few systems have actually achieved dimensional affect recognition from multiple modalities. These are reviewed in [11]. Overall, existing systems use different training/testing datasets (which differ in the way affect is elicited and annotated), they differ in the underlying affect model (i.e., target affect categories) as well as in the employed modality or combination of modalities, and the applied evaluation method. As a consequence, it remains unclear which classification method is suitable for dimensional affect recognition from which modalities and cues. These challenges should be addressed in order to advance the field while identifying the importance and feasibility of the following issues. 1) Among the available remotely observable and remotely unobservable modalities, which ones should be used for automatic dimensional affect recognition? Should we investigate the innate priority among the modalities to be preferred for each affect dimension? Does this depend on the context (who the subject is, where she is, what her current task is, and when the observed behaviour has been shown)? 2) When labelling emotions, which signals better correlate with self assessment and which ones correlate with independent observer assessment? 3) How does the baseline problem affect recognition? Is an objective basis (e.g., a frame with an expressionless display) strictly needed prior to computing the dimensional affect values? If so, how can this be obtained in a fully automatic manner from spontaneous data? 4) How should intensity be modelled for dimensional and continuous affect recognition? Should the aim be personalizing systems for each subject, or creating systems that are expected to generalize across subjects? 5) In a continuous affect space, how should duration of affect be defined? How can this be incorporated in automated systems? Will focusing on shorter or longer observations affect the accuracy of the measurement process?

Finding straightforward answers to these questions is beyond the scope of this paper. Although research fields such as engineering, computer science, psychology, neuroscience, and cognitive sciences seem to be somewhat detached and have their own research community and audience, emotion research is inherently multi-disciplinary. Great advances in emotion research are possible, however, depend on all the aforementioned fields stepping out of their labs, working side-by-side together in real-life applications, and sharing the experience and the insight acquired on the way, to make emotion research *tangible* for the *real world* and the *real people* [20]. Pioneering projects representing such inter-disciplinary effort have already started emerging, ranging, for instance, from publishing compilation books of related work papers (e.g., [9]) to projects as varied as affective human-embodied conversational agent interaction (e.g., European Union FP 7 SEMAINE [25]), and affect sensing for autism (e.g., [21]).

ACKNOWLEDGMENTS

The research of Hatice Gunes is funded by [FP7/2007-2013], grant agreement no 211486 (SEMAINE). The research of Maja Pantic is funded in part by the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

REFERENCES

- 1. Chanel, G., Ansari-Asl, K. & Pun, T. Valence-arousal evaluation using physiological signals in an emotion recall paradigm, *Proc. of IEEE SMC* (2007), 2662-2667.
- Cowie, R., Douglas-Cowie, E., Savvidou, S., Mcmahon, E., Sawey, M. & Schröder, M. 'FEELTRACE': An instrument for recording perceived emotion in real time, *Proc. ISCA W. on Speech and Emotion* (2000), 19–24.
- 3. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. & Taylor, J. G. Emotion recognition in human-computer interaction, *IEEE Signal Processing Magazine*, 18, 1(2001), 32-80.
- 4. Cowie, R., Gunes, H., McKeown, G., Vaclau-Schneider, L., Armstrong, J. & Douglas-Cowie, E. The emotional and communicative significance of head nods and shakes in a naturalistic database, *Proc. of LREC Int. Workshop on Emotion* (2010), 42-46.
- 5. Ekman, P. & Friesen, W.V. Unmasking the face: a guide to recognizing emotions from facial clues (2003) Cambridge, MA.
- 6. Ekman, P. & Friesen, W. V., Head and body cues in the judgment of emotion: A reformulation, *Perceptual and Motor Skills*, 24 (1967), 711-724
- Fontaine, J. R., Scherer, K. R., Roesch, E. B. & Ellsworth, P. The world of emotion is not twodimensional, *Psychological Science*, 18 (2007), 1050-1057.
- 8. Frijda, N. H. (1986) *The emotions*, Cambridge University Press.
- 9. Gokcay, D. & Yildirim, G. (Eds.) (2010) Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives, IGI Global.
- Grandjean, D., Sander, D. & Scherer, K. R. Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization, *Consciousness and Cognition*, 17, 2 (2008), 484-495.
- 11. Gunes, H. & Pantic, M. Automatic, dimensional and continuous emotion recognition, *Int. Journal of Synthetic Emotions*, 1, (1) 2010, 68-99.
- 12. Gunes, H. Piccardi, M. & Pantic, M. From the lab to the real world: Affect recognition using multiple cues and modalities, *In Affective Computing, Focus on Emotion Expression, Synthesis and Recognition* (2008), 185-218.

- Haag, A., Goronzy, S., Schaich, P. & Williams, J. Emotion recognition using bio-sensors: First steps towards an automatic system, *LNCS 3068* (2004), 36-48.
- 14. Kanluan, I., Grimm, M. & Kroschel, K. Audio-visual emotion recognition using an emotion recognition space concept, *Proc. of the 16th European Signal Processing Conference* (2008).
- 15. Kim, J., Bimodal emotion recognition using speech and physiological changes, *In Robust Speech Recognition and Understanding*, (2007) 265-280.
- Kleinsmith, A. & Bianchi-Berthouze, N., Recognizing affective dimensions from body posture, *LNCS* 4738 (2007), 48-58.
- Kulic, D. & Croft, E. A. Affective state estimation for human-robot Interaction, *IEEE Tran. on Robotics*, 23, 5(2007), 991-1000.
- 18. Nicolaou, M. A., Gunes H. & Pantic, M. Automatic segmentation of spontaneous data using dimensional labels from multiple coders, *Proc. of LREC Int. Workshop on Multimodal Corpora* (2010), 43-48.
- 19. Ortony, A., Clore, G. L. & Collins, A. (1988) *The* cognitive structure of emotions, Cambridge Univ. Press.
- 20. Picard, R.W., Emotion research by the people, for the people, *Emotion Review* (2010), in press.
- 21. Picard, R.W., Future affective technology for autism and emotion communication, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1535 (2009), 3575-3584
- 22. Pun, T., Alecu, T.I., Chanel, G., Kronegg, J. & Voloshynovskiy, S. Brain-computer interaction research at the Computer Vision and Multimedia Laboratory, University of Geneva, *IEEE Tran. on Neural Systems* and Rehabilitation Engineering, 14, 2 (2006), 210-213.
- 23. Russell, J. A. A circumplex model of affect, J. of *Personality and Social Psychology*, 39 (1980)., 1161-1178.
- 24. Scherer, K.R., Schorr, A., Johnstone, T. (Eds.). (2001) Appraisal Processes in Emotion: Theory, Methods, Research, Oxford Univ. Press.
- 25. Schröder, M., Bevacqua, E., Eyben, F., Gunes, H., Heylen, D., Maat, M., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., Sevin, E., Valstar, M. & Wöllmer, M. A demonstration of audiovisual sensitive artificial listeners, *Proc. of Int. Conf. on Affective Computing and Intelligent Interaction* (2009), 1, 263-264.
- 26. Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I. & Frank, M.G. Imaging facial physiology for the detection of deceit, *Int. J. of Computer Vision*, 71, 2 (2007), 197-214.
- 27. Wöllmer, M., Eyben, F., Reiter, S., Schuller, B., Cox, C., Douglas-Cowie, E. & Cowie, R. Abandoning emotion classes - Towards continuous emotion recognition with modelling of long-range dependencies, *Proc. of Interspeech* (2008), 597-600.

Relative Affective Blindsight for Fearful Bodily Expressions

Bernard M.C. Stienen

B.M.C.Stienen@uvt.nl

Beatrice de Gelder B.deGelder@uvt.nl

Faculty of Social and Behavioural Sciences, Tilburg University PO Box 90153, 5000 LE Tilburg, The Netherlands

ABSTRACT

Nonconscious affective perception has repeatedly been shown for facial expressions but not for bodily expressions while being highly salient and known to influence our behavior towards others. Lau & Passingham [7] found a case of relative blindsight using a parametric masking design. We used a comparable approach to find this relative case of blindsight for affective information in which the affective information can be processed independently of visual awareness. Participants had to detect masked fearful bodily expressions among masked neutral bodily actions as distractors and subsequently the participants had to indicate their confidence. The onset between target and mask (Stimulus Onset Asynchrony, SOA) varied from -50 to +133 milliseconds. D-prime as well as the confidence ratings showed that the bodies could be detected reliably in all SOA conditions. Importantly, a phenomenon which we coined relative affective blindsight was found, defined as two SOA conditions showing same d-prime values, while the confidence ratings differed.

Author Keywords

Emotion, masking, blindsight.

ACM Classification Keywords

J.4 Social and behavioral sciences

INTRODUCTION

Backward masking is one of the most widely used techniques for exploring unconscious processing of visual emotional information in neurologically intact observers. Esteves and Öhman [4] found that short (e.g. 33 ms) presentation of a facial expression (happy and angry) replaced immediately by a neutral face (mask) with a longer duration (e.g. 50 ms) is below the participants' identification threshold. Esteves, Parra, Dimberg and Öhman [5] reported that participants prevented from conscious recognition of conditioned angry faces by

backward masking still showed elevated skin conductance response to these faces, while Esteves, Dimberg and Öhman [3] found that this response could not be conditioned when happy faces were used. Dimberg, Thunberg and Elmehed [2] used EMG to show that participants respond to happy and angry faces with corresponding specific muscles in the face while not being conscious of the presentation of the faces.

A critical issue in most of the backward masking experiments concerns the measure adopted for visibility or visual awareness of the target. Most often this is assessed in a separate post test session. This clearly complicates the interpretation of studies using masking procedures because visibility of the target co-varies with the performance for each target presentation. To counter these methodological problems Lau and Passingham [7] performed an elegant parametric masking study. They presented their participants with masked diamonds and squares and asked them on each trial to identify the target and, next, to indicate whether they had seen the target. The onset between target and mask (Stimulus Onset Asynchrony, SOA) varied from minus 50 to 133 milliseconds. This method provided information about whether the participant was aware of the presence of a stimulus on a trial by trial basis and controls for the possibility that participants are likely to be more aware when the SOA value is larger and thus better visible. In particular, Lau and Passingham [7] coined the term "relative blindsight" to refer to two SOA conditions where participants were performing equally in the identification task but differed in reporting whether they had seen the target or not.

We adapted a similar approach to investigate whether bodily expressions can be detected independently of subjective visual awareness. Participants had to detect masked and unmasked fearful bodily expressions among masked and unmasked distractors (a neutral action; combing). The pictures and the mask were controlled for several factors such as lighting, size of the postures, contrast, and importantly the actors were uniformly dressed in black clothing. A mask was presented at 12 different SOAs varying from minus 50 to 133 milliseconds. The participants were instructed to detect the emotional

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

expression and subsequently to indicate whether they were sure or whether they were guessing.

In line with the methodology of Lau and Passingham [7] we expected to find this relative case of blindsight for affective information in which the affective information can be processed independently of visual awareness. Because we used a pattern mask it is expected that the lowest detection performance and confidence ratings will be around the SOA of 0 milliseconds and will be U-shaped [1]. Following Lau and Passingham [7] we conjectured that this U-shape implies that we can find SOA conditions where the objective performance is the same, but the confidence ratings are different.

METHOD

Participants

Twenty-three undergraduate students of The University of Tilburg participated in exchange of course credits or a monetary reward (16 women, 7 men, M = 19.8 years, SD = 2,3). All participants had normal to corrected-to-normal vision and gave informed consent according to the declaration of Helsinki.

Stimuli and Materials

Photos of actors expressing fear and combing their hairs were selected from a photoset. During the photo shoot pictures were projected on the wall facing the actor meant to trigger the fear response as natural as possible. Moreover, a short emotion inducing story related to the image projected was told by the experimenter. For the combing pictures the actors were asked to pretend that they had a comb in their hands and that they were straightening their hairs.

The faces were covered with an opaque oval patch to prevent that the facial expression would influence the identification of the emotional body expression. The color of the patch was the average grey value of the neutral and emotional face within the same actor. In addition, the colors were saturated to white and black with the color of the mask as anchor point. In this way, the participants were forced to base their judgments on the contours of the body because by isolating only two colors the color differences within the clothing disappeared. A total of 16 pictures (2 Fear/Combing x 2 gender x 4 actors) were selected for use in the present study. Average height of the bodies was 7.78 degrees, the average maximum width (distance between the hands) was 2.83 degrees and the average waist was 1.39 degrees of visual angle.

Using Adobe Photoshop 7.0 © a pattern mask was constructed by cutting the target bodies in asymmetric forms which were scrambled and replaced in the area occupied by the bodies. The parts were grouped with the restriction that parts containing white had to be grouped within the area occupied by the hands (which were saturated to white) and parts containing black had to be grouped within the area occupied by the bodies (which were black). Finally, the resulting picture was duplicated,

rotated 180 degrees and pasted at the background to induce symmetry and extra noise to avoid the percept of a body. The result is the mask in Figure 1. The height of the mask was 9.85 degrees and the maximum width was 6.48 degrees of visual angle. The mask covered the area of the stimuli completely.

The stimuli were presented on a 17" PC screen with the refresh rate set to 60 Hz. We used Presentation 11.0 to run the experiment. A white cross of 1.22×1.22 degrees was used as a fixation mark in the center of the screen. Finally, all stimuli were pasted on a gray background.

The SOA values were -50, -33, -17, 0, 17, 33, 50, 67, 83, 100, 117 and 133 milliseconds. The actual presentation time was calibrated with the use of a photodiode and an oscilloscope measuring the latency between onset of the target and the mask. Negative SOA values represent forward masking and positive SOA values backward masking. Moreover a target-only condition and a mask-only condition were included. One complete run summed up to a total of 224 trials (8 identities x 2 actions (Fear/Combing) x 14 timing conditions (including target-only and mask-only) which were randomly presented.

Procedure

Participants were comfortably seated in a chair in a soundproof experimental chamber approximately 90 cm from the screen. A trial started with a white fixation cross on a gray background. The disappearance of this cross signaled the beginning of a trial. After 500 milliseconds the target stimulus appeared for 33 milliseconds. After a variable interval the mask was presented for 50 milliseconds (sometimes the mask was presented first). This means that when the SOA value was -33, -17, 0 and 17 milliseconds the body and the mask overlapped. The body was always presented at the foreground. The participants were instructed to push a predefined button using the index finger of their left hand as soon as they thought a fearful bodily expression was presented (GO) and to withhold their response when they thought the neutral action was presented (NO-GO). Two thousand milliseconds after the target a screen was presented with the text "Sure or Guessed?". They had to respond with the other hand with



Figure 1. Visual representation of a trial (on the left), example target and distracter stimuli per detection task (middle right) and the mask (bottom right).



Figure 2. Detecting fearful bodily expressions as a function of SOA values (a) Confidence ratings as a function of SOA values (b) Error bars indicate standard error mean. * = p < .05.

two different buttons on the same response box labeled with "Sure" and "Guessed". These buttons were counterbalanced across participants. It was stressed that they had to respond as accurate and fast as possible and that they could use their "gut feeling" if they did not have seen the body. Finally a gray screen was presented with a random duration between 17 milliseconds and 767 milliseconds. This jitter was added to prevent that the participants would be caught in a mechanical rhythm. In total the trials were on average 4025 milliseconds. See Figure 1 for a visual representation of a trial.

Previous to the experimental sessions the participants had to perform two practice sessions consisting of 33 trials each (16 target-trials, 16 distractor- trials, and 1 mask-only trial). Other identities than the ones used in the main experiment served as targets. When the participants had more than 12 hits and gave notice of a full understanding of the procedures the main experiment was started. A total of four runs were presented adding up to a total of 896 trials. Before each run the trials were shuffled. Every 112 trials there was a 3 minute break.

The sensitivity to the signal (detection of expressions) was estimated by calculating the d-prime (d'). The d' is a measure for the distance between the signal and noise distribution means in standard deviation units [6]. A d' of 0 means that the participants are not able to discriminate the fearful bodily expressions from the neutral bodily actions. D-prime was calculated as:

$$d' = \Phi^{-1} (H') - \Phi^{-1} (FA')$$

Where H' is the corrected hit rate and FA' is the corrected false alarm rate. The function Φ^{-1} converts the rates into z-scores. The correction of the hit- and false alarm rates was performed to protect against ceiling effects as proposed by Snodgrass and Corwin [8]:

H' = (h + 0.5) / (h + m + 1)FA' = (f + 0.5) / (f + cr + 1).

Where h is the number of hits, m is the number of misses, f is the number of false alarms and cr is the number of

correct rejections. See also Tamietto, Geminiani, Genero, and de Gelder [9].

Confidence ratings were calculated in analogue with the dprime in the sense that information from all four cells (hits, misses, false alarms and correct rejections) were used. Per SOA condition the number of sure responses when the detection of the emotional expression or the rejection of the neutral action was incorrect was subtracted from the number of sure responses when the response was correct. This was divided by the total number of correct and incorrect answers. A resulting value of zero would mean that the participants indicate subjectively that they are not more confident of their correct answers then their incorrect answers which is taken as a measure of subjective visual awareness. This measurement will be called confidence ratings in the remaining paper.

RESULTS AND DISCUSSION

As shown in Figure 2a, the d-prime pattern shows a classical pattern masking curve with the lowest point of the curve when the SOA was 0 milliseconds [1]. There was a main effect of SOA as indicated by a MANOVA with SOA as a factor (F(12,11) = 29.73, p < .001). Interestingly, the d-prime was above zero when the SOA was 0 milliseconds (t(22) = 9.64, p < .01), indicating that the participants were capable of detecting the fearful bodily expressions. The confidence ratings are plotted in Figure 2b. Also, a main effect of SOA is found here (F(12, 11) = 86.37, p < .001). Strikingly, participants were still more confidence ratings being still significantly above zero when the SOA was 0 milliseconds (t(22) = 5.76, p < .01).

When exploring Figure 2a detection performance seems equal between SOA values of -50 & +33, -50 & +50, -33 & +33, and -33 & +50 milliseconds, this was confirmed with paired t-tests showing no significant differences. However, when performing statistical comparisons between the same SOA conditions on the confidence ratings it appeared that the confidence ratings differed significantly for the SOA

value pair -50 & +33 milliseconds (t(22) = 2.23, p < .05). This was also the case for the comparison of the SOA conditions of -33 & +33 milliseconds; the d-prime did not differ, while the confidence ratings did (t(22) = 2.25, p < .05).

While participants are in some conditions equally capable in detecting the fearful body posture, the subjective confidence ratings differed. The dissociation between the detecting performance and the confidence ratings seems to indicate that the fearful bodily expressions are being processed independently of subjective visual awareness. This process could be playing an important role in everyday vision supporting us unconsciously with important information of our surroundings. However, while in creating our stimuli we took care of triggering a fear response in the actors as natural as possible a drawback of the laboratory setting is that the spontaneous observation of fear is difficult to simulate.

Further research using neurological measures could give us insight which pathways are indeed mediating the independency of detecting fearful signals from visual awareness and how it interacts with other sources of information like auditory or other visual information like the face. Finally, it should also be examined if this phenomenon is general to the perception of all emotions or whether it is specific to the perception of fearful signals.

REFERENCES

1. Breitmeyer, B.G. and Ganz, L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review*, 83, 1 (1976), 1-36.

- 2. Dimberg, U., Thunberg, M., and Elmehed, K. Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*, 1 (2000), 86-89.
- 3. Esteves, F., Dimberg, U., and Öhman, A. Automatically elicited fear: conditioned skin conductance responses to masked facial expressions. *Cognition and Emotion*, *9*, 5 (1994), 99-108.
- 4. Esteves, F. and Ohman, A. Masking the face: recognition of emotional facial expressions as a function of the parameters of backward masking. *Scandinavian Journal of Psychology*, *34*, 1 (1993), 1-18.
- Esteves, F., et al. Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology*, *31*, 4 (1994), 375-385.
- 6. Green, D.M. and Swets, J.A., Signal detection theory and psychophysics. Wiley, New York, 1966.
- Lau, H.C. and Passingham, R.E. Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 49 (2006), 18763-18768.
- 8. Snodgrass, J.G. and Corwin, J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 1 (1988), 34-50.
- 9. Tamietto, M., et al. Seeing fearful body language overcomes attentional deficits in patients with neglect. *Journal of Cognitive Neuroscience*, *19*, 3 (2007), 445-454.

Full Papers

Proceedings of Measuring Behavior 2010 (Eindhoven, The Netherlands, August 24-27, 2010) Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman
Microstructural Assessment of Rodent Behavior in the Hole-Board Experimental Assay

Maurizio Casarrubea* m.casarrubea@unipa.it

Andrea Santangelo a.santangelo@ymail.com **Filippina Sorbera** f.sorbera@inwind.it

Giuseppe Crescimanno crescima@unipa.it

Dept. of Experimental Biomedicine and Clinical Neurosciences Human Physiology Section, University of Palermo Corso Tukory 129, 90134, Palermo, Italy

ABSTRACT

The aim of the present methodological article was to assess whether a microstructure of hole-exploratory activity can be demonstrated in rats tested in the hole-board, an experimental assay widely used to study anxiety-related behaviors in rodents. Six never properly described holeexploratory behavioral components were identified: activedip, static-dip, rapid-dip, circular-sniff, point-sniff and central-sniff. Results, assessed by means of descriptive and multivariate approaches, reveal a complex microstructural organization of these behavioral elements. The hypothesis of an important role of the relationships between point-sniff and head dipping activities, in anxiety-related holeexploratory behavior, is proposed. It is suggested that the microstructural analysis can evidence even subtle behavioral changes induced by manipulations of animal anxiety condition.

Author Keywords

Multivariate analysis, stochastic analysis, hole-board, rat.

ACM Classification Keywords

I.2..1 Applications and expert systems: Medicine and science, G.3 Probability and statistics: multivariate statistics, H.5.2 User interfaces: theory and methods

INTRODUCTION

The hole-board (HB) is an exploration-based assay widely used to assess various features of anxiety-related behaviors in rodents. HB generally consists of a rectangular or square

* Corresponding author

arena provided with a variable number of holes in the ground. The behavioral component more frequently investigated in HB is head-dip, that is, insertion of the head into a hole. However, conflicting evidences surround this behavioral component. For instance, increases [18], decreases [14], or no modifications [16] of head-dip frequencies have been described following administration of anti-anxiety drugs. Thus, whether quantitative evaluation of head-dip alone may be considered suitable anxiety level indicator represents matter of discussion. It is our contention that contrasting findings concerning this behavioral component have been produced by the unfortunate association of two factors: first, since HB introduction almost 50 years ago [1], inter-relations among behavioral components in this experimental assay have been scantly investigated; second, as a consequence and even more important, only quantitative evaluations of headdip have been produced for a long time. Multivariate analyses (MVA) provide excellent tools to overcome such constraints because they make available useful and interesting information on the relationships among behavioral elements [3, 4, 5, 10, 11, 15, 17]. Concerning studies on anxiety and depression, various articles have provided, by means of MVA, reliable information on rodent's behavior in the elevated-plus maze [9], in the forced swimming test [12], in the sucrose preference test and in the open field [2]. As to hole-board assay, by means of MVA, we have demonstrated that edge-sniff, a specific sniffing activity of holes edge, has complex relationships with head-dip and with several components of the behavioral structure [6, 7, 8]. In brief, we hypothesized that the patterning between these two behavioral components could rely on rodent's motivational drives which, in turn, are likely influenced by anxiety level [6, 7, 8]. Thus, given the prospective functional meaning of head-dip/edge-sniff patterning, the study of hole exploration activities, at a microstructural level, might contribute to the knowledge of motivational drives underlying anxiety-related behaviors in

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

HB. On this subject, remarkable information has been provided, in the past, by analyzing the behavioral microstructure of different rodent's activities, such as grooming [11] or feeding [20]. The aim of the present research is to describe, from a methodological point of view, an alternative approach to the study of rodent behavior in HB, by focusing the analysis on hole-exploration activity in its microstructural complexity.

METHODS

Experimental Apparatus

HB apparatus consisted of an enclosed 50×50 cm arena made of white opaque Plexiglas with a raised floor (5 cm above a white opaque Plexiglas sub-floor) containing four equidistant holes (4 cm in diameter). Each hole centre was 10 cm from the two nearest walls so that holes were equidistant from adjacent corners [6, 7, 8].

Subjects and Housing

15 male, specific pathogen free, Wistar rats (Morini, Italy), 55-60 days old, weighing 220-250 g, were used. Each subject was individually housed in a room maintained at 23 \pm 1°C, under a normal 12 h light/dark cycle (light on: 7:00 a.m. – light off: 7:00 p.m.). Food (standard laboratory diet -Mucedola, Italy) and water were available "ad libitum".

Procedure

Rats were transported from housing room to testing room inside their home-cages to minimize transfer effect. To avoid possible visual and/or olfactive influences, animals were allowed to acclimate for 30 minutes far from observational apparatus. Environmental temperature was maintained equal to temperature in the housing room. Each subject, experimentally naïve at test beginning, was placed in the arena centre and allowed to freely explore for 10 min. After each observation, hole-board apparatus was cleaned with ethylic alcohol (70%) to remove scent cues left from the preceding subject. Experiments were recorded through a digital videocamera and video files stored in a personal computer for following analyses.

Ethical Statement

Experiment was approved by a local veterinary committee and conducted in accordance with the European Communities Council Directive 86/609/EEC concerning the protection of animals used for experimental scientific purposes.

Ethogram and Behavioral Coding

The first step in a behavioral study is generally represented by the construction of a suitable ethogram, that is, a formal description of the behavioral components taken into consideration. The ethogram here used is focused on rodent's activities of hole-exploration. Since behavioral coding is an extremely work-demanding aspect, utilization of a suitable software coder is very important. We analyzed digital video files by means of the software coder The Observer (Noldus Information Technology, The Netherlands) [19].

Descriptive Analyses

To discern frequent behavioral components from the infrequent ones per cent distributions have been calculated and presented by means of a pie-chart (Figure 1). Moreover, latencies of each behavioral element first appearance have been evaluated and graphically presented by means of a histogram (Figure 2).

Multivariate Analyses

Different multivariate approaches are available to assess animal behavior [5, 8, 10, 11, 15, 17]. In present research we used a MVA based on transition matrices elaborations. First of all, transitions from a behavioral element to another one were reported in a transition matrix (TM) by means of a relevant option available in the software coder. Then, TMs of each subject were summed obtaining a total TM. Within total TM, two important conditions should be respected: a) number of empty cells should be not more than 20% [4, 5, 6]; b) minimum number of transitions should be, at least, 5 times the number of components in the ethogram [5, 17]. Total TM represents the starting point for several analyses based on matrix elaboration such as stochastic analysis, cluster analysis or adjusted residuals analysis [5]. In this brief article, we present only the stochastic one. Multivariate stochastic analysis requires that a total TM is transformed into a table presenting probabilities of transitions among patterns (i.e. a probability matrix) where: a) transition probability from a pattern to all others is 1; b) each row must sum to 1, and c) each transition must be between 0 and 1 [4, 5, 6]. In present study individual TMs, along with resulting total TM, were processed by means of the specific software for matrix manipulation and analysis Matman (Noldus Information Technology, The Netherlands) [13].

MVA Graphical Representations

Basically, a transition matrix is a table filled with hundreds of numbers. In this form a transition matrix is quite difficult to interpret even for an experienced reader. For this reason it is important to express a transition matrix in an intuitive graphical fashion. In present article, the stochastic matrix has been graphically expressed by means a path diagram in which different transition probabilities were represented by connecting arrows of different thickness. By means of this approach both directions and transition probabilities can be simultaneously illustrated.

RESULTS

Ethogram of the Hole-Exploration Process

The ethogram encompasses six behavioral components: Point-Sniff (PoS): the rat sniffs a single point of the holeedge; Central-Sniff (CtS): the rat sniffs the hole center without inserting the head inside; Circular-Sniff (CiS): the rat sniffs the hole-edge in a continuous circular fashion;



Figure 1. Per cent distribution of rat's hole exploratory components. Abbreviations in the text.

Static-Dip (StD): the rat puts and maintains its head into the hole, the body is immobile; Rapid-Dip (RaD): the rat rapidly puts into and remove its head from the hole; Active-Dip (AcD): the rat puts its head into the hole; body movements are produced. Behavioral group "others" enfolds all remaining components not related with holeexploration [6, 7, 8].

Descriptive Analyses

Figure 1 shows that all sniffing activities, that is, pointsniff, circular-sniff and central-sniff, represent together the largest extent of hole-exploratory patterns. Mean latencies (Figure 2), evaluated on the basis of each behavioral component first appearance, demonstrate that point-sniff and static-dip are the first components of hole-exploratory behavior to appear, both within the first minute of observation. Rapid-dip, active-dip and circular-sniff are present after the first minute and within the third. Centralsniff is the last pattern to occur.

Multivariate Stochastic Analysis

Probabilities of transitions, illustrated by means of a path diagram (Figure 3), show that all dipping behaviors (i.e. rapid-dip, static-dip and active-dip) have high chances of transition exclusively toward point-sniff. This last, in turn, represents the only pattern linking the remaining components of the behavior (i.e. "others") with a high probability. All remaining transitions occur with medium to low probabilities.



Figure 2. Latencies \pm SE (s) of each behavioral component first appearance. Abbreviations in the text.



Figure 3. Path diagram illustrating three different transition probabilities. Abbreviations in the text.

DISCUSSION

This is the first research assessing, by means of descriptive and multivariate analyses, the behavioral process of holeexploration at a microstructural level. Six new behavioral components, never properly described, have been introduced and two main behavioral clusters of hole exploration activities in HB were identified: the sniffing one and the inside-hole exploration one. The former consisted of point-sniff, circular-sniff and central-sniff; the latter encompassed rapid-dip, static-dip and active-dip.

The pie-chart (Figure 1) shows a prevalence of all sniffing activities in comparison with the inside-hole exploratory ones. Noticeably, point-sniff and static-dip represent roughly the 70% of the whole behavioral repertoire of holeexploration. Latencies (Figure 2) demonstrate that pointsniff and static-dip are the first to appear, both within the first minute of observation. All the remaining components show higher latencies. Path diagram reveals that all headdipping activities have high transitional probabilities exclusively toward point-sniff (Figure 3) that, in turn, is the sole behavioral element of hole-exploration process showing high transition probabilities toward remaining components (i.e. 'others'). Taken together, such results suggest an important role of point-sniff, static-dip and of their patternings, both in terms of quantitative aspects and probabilistic arrangement of transitions as well. However, per cent distributions, latencies and, more in general, all quantitative assessments of behavior are able to emphasize only specific quantitative aspects, representing each component, although carefully quantified, only a disjointed element. For instance, albeit pie chart (Figure 1) clearly demonstrates that point-sniffing is characterized by a high per cent value and by the lowest latency (Figure 2), relationships among point-sniff and the whole behavioral structure are missing. Hence, the possibility to characterize each behavioral component through even thousands of numbers does not imply the possibility to use those numbers to figure out what the behavior is in its wholeness. As already showed in other experimental models of anxiety and depression [2, 9, 12], this is the pivotal aspect that distinguishes 'conventional' quantitative approaches from

multivariate ones: MVA offers useful tools to avoid reductive interpretation of isolated patterns, disjointed from the whole behavioral architecture, thus revealing behavioral dynamics otherwise undetectable by means of quantitative analyses alone.

CONCLUSION

Main outcome of the present research is that rat holeexploratory behavior in hole-board presents a complex microstructure where the interrelationships among its components play an important role. The multivariate approach, synergically used with traditional descriptive methods (such as evaluation of per cent distributions and latencies) offers an innovative tool to assess, at microstructural level, even subtle behavioral changes induced by manipulations of animal anxiety condition. Further studies are in progress in our laboratories to assess possible effects of diazepam, a reference anxiolytic drug, on the behavioral microstructure of rat's response to anxiety.

REFERENCES

- 1. Boissier, J.R., Simon, P., Lwoff, J.M. Use of a particular mouse reaction (hole board method) for the study of psychotropic drugs. *Thérapie*, *19* (1964), 571-583.
- 2. Brenes Sáenz, J.C., Villagra, O.R., Fornaguera Trías, J. Factor analysis of Forced Swimming test, Sucrose Preference test and Open Field test on enriched, social and isolated reared rats. *Behavioural Brain Research*, *169*, 1 (2006), 57-65.
- 3. Casarrubea, M., Sorbera, F., Crescimanno, G. Effects of 7-OH-DPAT and U 99194 on the behavioral response to hot plate test, in rats. *Physiology & Behavior*, 89, 4 (2006), 552-562.
- Casarrubea, M., Sorbera, F., Crescimanno, G. Multivariate analysis of the modifications induced by an environmental acoustic cue on rat exploratory behavior. *Physiology & Behavior 93*, 4-5 (2008), 687-696.
- Casarrubea, M., Sorbera, F., Crescimanno, G. Multivariate data handling in the study of rat behavior: an integrated approach. *Behavior Research Methods*, 41, 3 (2009), 772-781.
- 6. Casarrubea, M., Sorbera, F., Crescimanno, G. Structure of rat behavior in hole-board: I) multivariate analysis of response to anxiety. *Physiology & Behavior*, 96,1 (2009), 174-179.
- Casarrubea, M., Sorbera, F., Crescimanno, G. Structure of rat behavior in hole-board: II) multivariate analysis of modifications induced by diazepam. *Physiology & Behavior 96*, 4-5 (2009), 683-692.

- Casarrubea, M., Sorbera, F., Magnusson, M., Crescimanno, G. Temporal patterns analysis of rat behavior in hole-board. *Behavioural Brain Research*, 208, 1 (2010), 124-131.
- 9. Espejo, E.F. Structure of the mouse behaviour on the elevated plus-maze test of anxiety. *Behavioural Brain Research*, 86, 1 (1997), 105-112.
- Hemerik, L., Bukovinszky, T., Gols, R., van Lenteren, J. C. and Vet, L.E.M. Enter the matrix: How to analyze the structure of behavior. *Behavior Research Methods*, *38*, 3 (2006), 357-363.
- 11. Kalueff, A.V., Tuohimaa, P. Grooming analysis algorithm for neurobehavioural stress research. *Brain Research Brain Research Protocols*, *13*, 3 (2004), 151-158.
- Lino-de-Oliveira, C., De Lima, T.C., de Pádua Carobrez, A. Structure of the rat behaviour in the forced swimming test. *Behavioural Brain Research*, 158, 2 (2005), 243-250.
- 13. Matman. http://www.noldus.com/
- 14. Pellow, S., Chopin, P., File, S.E., Briley, M. Validation of open:closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. *Journal of Neuroscience Methods*, *14*, 3 (1985), 149-167.
- 15. Quera, V. RAP: a computer program for exploring similarities in behavior sequences using random projections. *Behavior Research Methods*, 40, 1 (2008), 21-32.
- 16. Sayin, U., Purali, N., Ozkan, T., Altug, T., Buyukdevrim, S. Vigabatrin has an anxiolytic effect in the elevated plus-maze test of anxiety. *Pharmacology Biochemistry and Behavior*, 43, 2 (1992), 529-535.
- 17. Spruijt, B.M., Gispen, W.H. Behavioral sequences as an easily quantifiable parameter in experimental studies. *Physiology & Behavior*, *32*, 5 (1984), 707-710.
- Takeda, H., Tsuji, M., Matsumiya, T. Changes in headdipping behavior in the hole-board test reflect the anxiogenic and/or anxiolytic state in mice. *European Journal of Pharmacology*, 350, 1 (1998), 21-29.
- 19. The Observer XT. http://www.noldus.com/

Zorrilla, E.P., Inoue, K., Fekete, E.M., Tabarin, A., Valdez, G.R., Koob, G.F. Measuring meals: structure of prandial food and water intake of rats. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 288, 6 (2005), R1450-1467.

A Novel Conditioning Paradigm Enables the Dissociation Between the Formations of Context- and Cue-Dependent Memories of Drug Reward

Yossef Itzhak

Department of Psychiatry & Behavioral Sciences and Neuroscience Division University of Miami Miller School of Medicine, 1011 NW 15th Street, Gautier Bldg Room 503 Miami, FL 33136, USA +1-305-243-4635 vitzhak@med.miami.edu

ABSTRACT

Specific context and cues that had been associated with drug reward elicit the development of drug-associated contextual and cued memories. Re-exposure to drugassociated context and cues may cause arousal and drugseeking behavior in humans who became addicted to drugs. This maladaptive behavior is dependent on classical Pavlovian conditioning. In animal models, the conditioned place preference paradigm has been used to investigate the motivational effects of drug reward. However, this paradigm does not allow dissociation between approach behavior to discrete contextual and cued stimuli that had been paired with drug reward. Recently, we had designed a novel conditioning paradigm that allows us to discriminate between the formations of context- and cue-dependent memories of drug reward. We used a commercially available two-compartment conditioned place preference apparatus; one compartment is painted white and the second is painted black. During the conditioning phase (4 days), mice received cocaine either in the black or in the white comportment of the cage. In our new design, we inserted a blinking light cue into the drug-paired compartment. Subsequently, we tested one group of mice for approach behavior to the context (in the original training context) in the absence of the light cue. Then, we tested a second group of mice for approach to the light cue in a novel context. Mice conditioned by cocaine developed robust approach behavior for both cocaine-associated context and cue. In control experiments, mice had neither preference nor aversion to the light cue. This new straightforward paradigm allows the dissociation between context- and cuedependent memories of drug reward. Therefore, this novel behavioral paradigm is significant for the investigation of mechanisms underlying the development of addictive behavior.

Author Keywords

Reinforcement learning, contextual and cued memory, Pavlovian conditioning, reward.

INTRODUCTION

In classical Pavlovian conditioning, pairing of an unconditioned stimulus (US) with a neutral context and cues, such as light or sound, confers conditioned stimulus (CS) properties to these entities. When a specific context is paired with reinforcing stimuli, the conditioned response is approach. Traditionally, this behavior was quantified in the conditioned place preference (CPP) paradigm. The acquisition of place conditioning requires first, an US that changes the affective state of the organism, and second, learning and memory processes [5]. The CPP paradigm has been used to investigate the motivational effects of drug and natural reward.

In a typical place conditioning apparatus, cues such as floor texture and wall color or pattern are embedded in the context, making it difficult to distinguish between cues and context. Thus the reinforcer-paired 'environment' acquires properties of a CS. Subsequent exposure to the CS elicits approach behavior to the drug-paired environment. This type of learning is also viewed as 'habit learning,' which has a major role in the development of drug addiction [3]. While the context of drug exposure may elicit drug craving, presentation of cocaine- and alcohol-related cues to cocaine and alcohol abusers elicits limbic activation, craving and physiological responses similar to the drugs' effects, suggesting the emergence of a conditioned response to drug-associated cues [2]. Given the pivotal role of cuedependent reinforcement learning in the development of drug addiction, our goal was to modify the traditional CPP paradigm to enable us to investigate not only contextual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

memories of drug reward but also discrete cued memories of drug reward. To this end, we investigated the development of approach behavior to discrete contexts and cue that were associated with cocaine administration, using a modified CPP paradigm in mice.

ANIMALS AND METHODS

Animals

Mice purchased from Jackson Laboratories (Bar Harbor, Maine, USA) were bred in our facilities at the University of Miami, Miller School of Medicine, Miami, FL, USA, as we described previously [1]. Both genotypes, wild type (WT) and neuronal nitric oxide (nNOS) knockout (KO), were generated on a mixed B6;129S genetic background. For the experiments described herein, adult (7-8 weeks old) WT and nNOS KO males were used. Animal care was in accordance with the Guide for the Care and Use of Laboratory Animals (National Research Council, National Academy Press, 1996) and approved by the University of Miami Animal Care and Use Committee. Mice were conditioned by 20mg/kg cocaine, a dose we found optimal in WT and nNOS KO mice for the acquisition of cocaine conditioned place preference [1].



Figure 1. Schematic presentation of the training (A) and novel (B) contexts. The training context consisted of two compartments, separated by a removable guillotine door, one comprising four black walls with a smooth black floor and the other four white walls and a floor covered with sandpaper. During training four blinking lights (mini bulbs; 2.5V, each) were suspended through the ceiling perforations (not shown) of the drug-paired compartment (e.g., black or white); the light served as the drugassociated cue. Drug-associated cue preference was tested in the novel context (B). One compartment was covered with white stars (2.5cm in diameter) on a black background, and the second compartment was covered with black and white strips (2cm wide x 20cm high).

Apparatus

Custom-designed Plexiglas cages (42L x 20W x 20H cm; Opto-Max Activity Meter v2.16; Columbus Instruments, Columbus, OH, USA) were used. The training context consisted of two compartments, separated by a removable guillotine door, one comprising four black walls with a smooth black floor and the other four white walls and a floor covered with sandpaper (Figure 1A). Each compartment was covered by a transparent ceiling, perforated with an array of 16 small holes (1cm diameter, in an array of 4 rows with 4 holes each) to allow ventilation.

We introduced four blinking lights (mini bulbs; 2.5V, each) through the ceiling perforations (Figure 1A) as a cue. The compound context-cue stimulus was always paired with cocaine administration. Because experiments were carried out in an unbiased design (half of the subjects were conditioned by cocaine in the black compartment and the other half in the white compartment) the light cue was present either in the black or in the white compartment of the training apparatus. Upon completion of the training (see below), approach behavior towards cocaine-associated context and cue was tested as follows: one group was tested in the training context in the absence of the light cue and a second group was tested in a novel context (Figure 1B) in the presence of the light cue. The novel context consisted of 2 different removable wall patterns and floor covers that we designed and had laminated. One compartment was covered with white stars (2.5 cm in diameter) on a black background, and the second compartment was covered with black and white strips (2cm wide x 20cm high) (Figure 1B). Each cage was equipped with 2 horizontal sensors mounted alongside opposing lengths. The two compartments (21 x 20 x 20 cm) were each scanned by 7 infrared beams at a rate of 10Hz (2.54 cm intervals). A null zone 8 cm wide was assigned at the interface of the two compartments to ensure that only full entry into each compartment was registered as 'real' time spent in each zone.

Training

Training and testing were carried out in dimmed lighting (30Watts; a reading lamp with two 18-inch white fluorescent bulbs, 15 Watts each, faced a wall) in a test room separate from the housing room. On the first day, between 12:00-14:00h, mice were habituated (20min) to the training context (Figure 1A) in the absence of the light cue; time spent in each compartment was recorded to determine preconditioning compartment-preference/aversion. To ensure a strictly unbiased training design, mice that showed initial preconditioning preference of more than 10-12% of the total time (20min) to either compartment were discarded. For the next 4 days (days 2-5) WT and nNOS KO mice (n=32-34 per group) were trained by a morning (10:00-12:00h) saline session and an afternoon (14:00-16:00h) cocaine (20mg/kg) session, each lasting 30min. For the unbiased design, training was counterbalanced: half of the subjects were trained with drug in the black compartment and the other half in the white compartment.

Cocaine was administered immediately before the animal was placed into the appropriate compartment, and the light cue was turned on 10min later. Animals thus experienced the presence of blinking lights for the final 20min of the drug-training session. Control groups of both genotypes (n=18-20 per group) received saline instead of cocaine in the afternoon session.

Testing

All tests were carried out between 12:00-14:00h, that is during the same time period in which the pretraining habituation had been recorded. Each test was performed in a drug-free state and it lasted for only 10min to minimize extinction learning. For each genotype, half of the subjects were tested for context preference and the other half was tested for cue preference. For context preference, the mouse was placed in the training context in the absence of the light cue (Figure 1A), and time spent in each compartment was recorded for 10min. For cue preference, the light cue was introduced into the novel context (Figure 1B) and it was blinking prior to entry of the mice into the testing room. Then each mouse was brought into the room, and placed in the center of the novel context (Figure 1B). Preference for the cue was recorded for 10min. In this group, half of the mice (n=8-9) were tested with the light cue in the same side of the cage in which the cue had been present during training. To avoid influence of spatial learning of the location of the light cue, the remaining mice from this group (n=8-9) were tested with the light cue located in the part of the cage corresponding to the spatial location opposite to that present during training. The outcome of these two tests was remarkably similar, suggesting that animals responded to the drug-associated cue and not to the spatial location of the light; hence, the results of the two tests were combined.

Statistical Analysis

The magnitude of preference for cocaine-paired context and cocaine-paired cue across the genotypes was analyzed by two-way ANOVA: genotype x context/cue preference (Figure 2). Specific differences between groups were analyzed by post hoc Newman-Keuls test. A p value less than 0.05 was considered significant.

RESULTS AND DISCUSSION

WT mice that had been conditioned by cocaine in a discrete context in the presence of a light cue showed significant preference for a) cocaine-paired context and b) cocaineassociated light cue (Figure 2). WT mice showed preference to the light cue that was paired with cocaine injections irrespectively of the spatial location of the light cue in the novel context. This finding suggests that following conditioning by cocaine, the light cue acquired the properties of reinforcing conditioned stimulus that elicits approach behavior. Notably, control mice that were trained only by saline showed neither preference nor aversion to the light cue.





Figure 2. Comparison between WT and nNOS KO mice for context and cue preference. There was no difference between WT and nNOS KO mice in the magnitude of preference for cocaine-associated context. In nNOS KO mice, a significant difference between the preference for cocaine-associated context and cue was observed (a; p<0.001). A significant difference between WT and nNOS KO mice in their preference for cocaine-associated cue was also observed (b; p<0.001).

Hence, we have shown that a modified conditioned place preference apparatus can be used to investigate both context and cue preference following conditioning by reinforcing drug. Interestingly, nNOS KO mice developed preference for the cocaine-paired context but not for the cocaine-paired light cue (Figure 2). It is unlikely that the impairment in the acquisition of preference to cocaine-associated cue is due to visual deficits in nNOS KO mice because they acquired normal context preference. The discrimination between the two contexts is visually dependent because one compartment is black and the other is white.

Several studies suggest that contextual memory is hippocampus-dependent and cued memory is amygdaladependent [4]. Thus, it appears that the nNOS gene is required for the formation of amygdala- more so than hippocampus-dependent memory of drug reward.

CONCLUSIONS

The present results demonstrate the utilization of a modified place-conditioning paradigm to investigate both contextand cue-dependent reinforcement learning. Unlike the traditional conditioned place preference design, this novel paradigm we developed allows dissociation between approach behavior to discrete contextual and cued stimuli that were associated with the motivational effect of drug reward. This novel paradigm will allow investigating the mechanisms underlying the development of addictive behavior.

ACKNOWLEDGEMENTS

This work was supported by award R21DA029404 from the National Institute on Drug Abuse, National Institutes of Health, USA.

REFERENCES

- 1. Balda, M.A., Anderson, K.L., and Itzhak, Y. Adolescent and adult responsiveness to the incentive value of cocaine reward in mice: role of neuronal nitric oxide synthase (nNOS) gene. *Neuropharmacology*, *51*, (2006), 341-349.
- Childress, A.R., Mozley, P.D., McElgin, W., Fitzgerald, J., Reivich, M., and O'Brien, C.P. Limbic activation during cue-induced cocaine craving. *American Journal* of *Psychiatry*, 156 (1999), 11-18.

- 3. Everitt, B.J., and Robbins, T.W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, 8 (2005), 1481-1489.
- 4. Phillips, R.G., and LeDoux, J.E. Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behavioural Neuroscience*, *106* (1992), 274–285.
- 5. White, N.M., and Carr, G.D. The conditioned place preference is affected by two independent reinforcement processes. *Pharmacology Biochemistry and Behavior*, 23 (1985), 37-42.

Automation of Continuous Spontaneous Alternation to Increase the Throughput for In Vivo Screening of Cognitive Enhancers. Optimization of the Ethovision System for the Y-maze Test in Mice

Eric Detrait Eric.detrait@ucb.com Catherine Brohez Catherine.brohez@ucb.com Etienne Hanon etienne.hanon@ucb.com Marc De Ryck Marc.deryck@ucb.com

UCB Pharma S.A., CNS Research, Chemin du Foriest, B-1420 Braine-l'Alleud, Belgium

ABSTRACT

Continuous spontaneous alternation in a symmetrical Ymaze is widely considered as a measure of spatial working memory in rodents, and is used as an entrance screening test for cognitive enhancers. The task consists of a single 8min session during which arm visits and sequence of arm entries are continuously recorded. Using EthoVision XT, we describe the implementation and validation of an automated scoring procedure for continuous spontaneous alternation in mice in a Y-maze. We optimized the design of the arena setting to avoid false arm entries, and developed an automated analysis program in order to extract from the EthoVision export file the sequence of arm entries and the number of alternations. We show that, after optimization, 100% agreement was reached between the automated output and human observers in terms of arm entries, sequence of entry and alternation percentage. We also demonstrate that several Y-mazes can be run in parallel without influencing mouse alternation behavior, and that the automated system is suitable for pharmacological testing.

Author Keywords

Y-maze, EthoVision, spontaneous alternation, drug screening.

INTRODUCTION

Spontaneous alternation in rats and mice refers to the natural tendency of rodents to spontaneously choose alternate arms in a Y- or T-maze. This behavior has been described about 80 years ago [1]. Spontaneous alternation

behavior has been ascribed to the operation of a variety of mechanisms, but regardless of this ethological function, it is evident that the animal must remember which arm it had entered on a previous occasion to enable it to alternate its choice on a following trial. Therefore, spontaneous alternation has been embraced by behavioral pharmacologists as a quick and relatively simple test of spatial working memory, devoid of fear, reward or reinforcers (review by Hughes, 2004 [2]).

Free-running or continuous spontaneous alternation in symmetrical Y-mazes was first applied to the study of drug effects about 30 years ago [3,4]. The basic procedure allowed individual rats and mice free access to the three arms of the maze for several minutes during which time sequences of entries of the arms were recorded. From these sequences, proportions of arms entered that were different from those previously visited were calculated to produce an alternation score. In the most recent form, which is widely used to study spatial working memory, animals are usually given an 8-min unrestrained trial in a symmetrical Y-maze [5]. Administration of scopolamine (intraperitoneally) or amyloid (intracerebrally) generates deficits of performance in the Y-maze. These deficits can be prevented by acetylcholine esterase inhibitors and by sigma-1 agonists [6,7].

In the present work, using EthoVision XT, we describe the implementation and validation of an automated scoring procedure for continuous spontaneous alternation of mice in a Y-maze.

MATERIAL AND METHODS

Animals

Young male Swiss mice were used (weighing 24-28 g, about 4-week old, purchased from Janvier France). They were group-housed in polypropylene cages (type III H Macrolon, with sterile aspen bedding) under the following standard conditions: 22 ± 2 °C, light/dark cycle 12 h/12 h - 6 am-6 pm, 55 ±15 % humidity, ultra filtered water and ad

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

libitum food (SDS RM1(E)SQC. The cages were located in an air-conditioned animal housing room (15-20 air changes per hour) in a type 1 conventional animal facility. All experiments were carried out according to the European guideline 86/609/CEE and 2003/65/CE and the Belgian legislation from August 14th, 1986 and its amendments. All experiments were approved by the UCB ethical committee.

Y-Maze Task

The Y-maze consisted of three arms of equal size. Each arm was a V-shape corridor made of white PVC. The arms measured 38.5 cm long, 3 cm wide and 13 cm high, and were oriented at 60° angles from each other. The arena was cleaned using Umonium (Huckert's international) between two trials in order to avoid odor trails. The Y-maze test was performed under moderate lighting conditions (200 lux), with moderately loud background music.

Mice began the single trial at the end of one arm, and were allowed to freely explore the Y-maze during 8 min. Number and sequence of arm visits were recorded either manually by the observer or by means of EthoVision XT. Alternation was defined as "a consecutive entry in three different arms". The alternation percentage was computed with the following formula: "number of alternations" divided by "total number of arm visits" minus 2.

RESULTS

Optimization of Arena Settings

EthoVision is a video tracking system, which first captures video, and then analysis the subject tracks using subject location and movement. The track analysis is based on arena settings. The arena settings define a variety of zones or locations inside the arena, for which a behavior will be measured. Behavior is here defined in terms of locomotion or position (e.g.: moving in, or around, a zone; being close, or far, from a location). The intuitive arena setting for a Ymaze task consisted of one central zone and three rectangular zones corresponding to each one of the three arms. Arm exploration was defined as entering at least the first third of an arm. Such settings did not reach satisfactory accuracy between automated and observer scoring, and provided on average 20% variation between the two scoring techniques. To optimize the accuracy of the automated scoring, additional zones were drawn as shown in Figure 1. The additional zones split the central zone of the maze between a central triangle located at the center of the three arms, and the proximal area of each arm. This proximal area covered approximately the first third of the arm. The reason to include the first third of the arm into the central zone was that part of normal rodent behavior consisted of repeatedly going around within the central zone of the Ymaze. In this case, rodents did not really explore any arm. Without the definition of the proximal arm area, the automated system considered any animal located in this area as exploring the arm, while it was not.

Furthermore, we designed an Excel-VBA macro to analyze the sequences of arm visits. EthoVision XT5 generated a track export file, which was first decoded by the Noldus Excel macro in order to provide the position of the animal relative to the zone areas defined in the arena settings. The position was given every 4/100 sec in a binary format, which was presence/ absence of the animal in any of the zone areas. Our Excel-VBA macro used the list of the positions of the center point as input. A new visit of an arm was recorded each time a visit of a distal area was followed by an entry in the central area. An arm entry was defined as center point of the mouse entered into the distal part of the arm (A, B, or C). Similarly, a new arm entry was only considered when the mouse center point left the proximal part of the arm (Ap, Bp, or Cp). The macro first detected the inter-zone transitions, zone transition being defined as a passage through the central area. In order to extract alternation from inter-zone transition, each visit to an arm was associated with a numerical code. The code took the values 1, 2 or 4 according to the arms visited. The successive transitions associated with a sum of codes equal to 7 corresponded to one alternation defined as a successive visit to the three arms of the Y-maze. For instance, the following sequence of mouse movements: "center - Ap - A -Ap - A - Ap - center - Bp - B'' was translated by the macro into the following arm transition : A - B.

The Optimized Arena Settings Increased the Accuracy of Automated Scoring

Mice were run during a 8-min session in the Y-maze. Their alternations were scored by EthoVision XT using the optimized arena settings or by visual observations. Figure 2 showed for each animal the alternation percentage obtained



Figure 1. Optimized arena settings to increase the accuracy of the automated scoring. The arm zone, defined as the distal 2/3 of each arm were labeled zone A, B, and C. The proximal areas of each arm were labeled zone Ap, Bp, and Cp. The central zone was therefore divided into the central triangle and three proximal zones.



Figure 2. Correlation between manual and automated scoring of alternation percentage. Each data point represents the alternation percentage of one animal. Total sample size = 28.

with both techniques. The alternation percentage was identical (100% accuracy and no variability) with the new settings. Similarly, and not surprisingly, the number of arm visits showed also 100% accuracy using the new arena setting and automated scoring (Figure 3).

Running Several Y-Mazes in Parallel Did Not Modify the Alternation Behavior of Mice

If the Y-maze test is to be used as an efficient screening tool, automated systems should not modify rodent performances when run in parallel within the same room. We compared the alternation percentage obtained by a single Y-maze, and by two mazes run in parallel in the same room by a single experimenter. The alternation percentage of vehicle-treated mice was 63.9 ± 3.1 with one maze (n=10) and 62.5 ± 3.0 (n=14) with 2 mazes. Student ttest did not find any significant statistical variation between the two settings (p=0.76), indicating that running several mazes in parallel did not modify the alternation behavior.



Figure 4. Alternation percentage of mice treated intraperitoneally with vehicle or 0.3mg/kg scopolamine 30min prior Y-maze trial. Statistical analysis: One sample t-test to compare the mean of each group to chance level (50%): ** = p<0.01; and Student t-test to compare the mean of both treatment group: # = p<0.05. vhe = vehicle-treated mice (n=14), scopo = scopolamine-treated mice (n=14).

Automated Y-Maze Task Is Suitable for Pharmacological Testing

The automated Y-maze system using two mazes in parallel was used to show a scopolamine-induced working memory deficit. Mice were blindly treated by intra-peritoneal injection of vehicle (1% methylcellulose-5% DMSO), or 0.3 mg/kg scopolamine. Thirty min later, they were run for a 8-min long Y-maze trial. Mice treated with vehicle showed an alternation percentage statistically above chance level (50% of alternation) whereas scopolamine-treated mice did not. Student t-test comparing both treatments showed that the alternation percentage of scopolamine-treated mice was statistically lower than that of vehicle-treated mice (see Figure 4).



Figure 3. Correlation between manual and automated scoring of the number of entries into each arm. Each data point represents the number of entries in arm A, B, or C.

CONCLUSION

We describe the optimization of an automated Y-maze task using EthoVision XT. To this end, we developed an arena setting spitting the central zone of the maze into proximal arm area and maze center. We also developed an Excel Macro to accurately extract the sequence of arm visits from the sequence of zone transitions in our new arena settings. This optimized system was able to detect arm transitions with 100% accuracy. Several Y-maze could be run in parallel in the same room by a single observer, without influencing mouse alternation behavior. The system was able to detect the deleterious effect of a pharmacological agent such as scopolamine on alternation performance.

REFERENCES

- 1. Tolman EC. Purpose and cognition: the determiners of animal learning. *Psychol. Rev.*, *32*, (1925), 285–97.
- 2. Hughes RN. The value of spontaneous alternation behavior (SAB) as a test of retention in pharmacological investigations of memory. *Neuroscience and Biobehavioral Reviews*, 28, (2004), 497–505.
- 3. Drew WG, Miller LL, Baugh EL. Effects of D9-THC, LSD-25, and scopolamine on continuous, spontaneous

alternation in the Y-maze. *Psychopharmacologia*, 32 (1973), 171–182.

- Kokkinidis L, Anisman H. Dissociation of the effects of scopolamine and d-amphetamine on a spontaneous alternation task. *Pharmacol. Biochem. Behav.*, 5 (1976), 293–297.
- Maurice T, Hiramatsu M, Itoh J, Kameyama T, Hasegawa T, Nabeshima T. Behavioral evidence for a modulating role of sigma-1 ligands in memory processes. I. Attenuation of dizocilpine (MK-801)induced amnesia. *Brain Res.* 647, (1994), 44–56.
- 6. Maurice T, Su T P, Privat A. Sigma1 receptor agonists and neurosteroids attenuate beta25–35-amyloid peptideinduces amnesia in mice through a common mechanism. *Neuroscience*, *83*, 2 (1998), 413–428.
- Meunier J, Ieni J, Maurice T. The anti-amnesic and neuroprotective effects of donepezil against amyloid beta25-35 peptide-induced toxicity in mice involve an interaction with the sigmal receptor. *Br. J. Pharmacol.*, *149*, 8 (2006), 998-1012.

Home Cage Testing of Decision-Making

Susanne Koot

s.koot@uu.nl

Ruud van den Bos

r.vandenbos@uu.nl

Division Neurobiology of Behaviour, Department of Animals in Science and Society, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 2, 3584 CM Utrecht, The Netherlands

Walter Adriani walter.adriani@iss.it Gianni Laviola gianni.laviola@iss.it

Behavioural Neuroscience Section, Department of Cell Biology and Neurosciences, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Roma, Italy

ABSTRACT

Testing rodents in their home cages has become increasingly popular. A new low-cost computer-controlled operant panel was designed, which can be placed inside the home cage. A pilot study was carried out, using a decisionmaking protocol, which was adapted from the original maze rodent Iowa Gambling Task (r-IGT). Male adult rats were tested in their home cages, containing the operant panel provided with nose-poking holes. Nose-poking was associated with rewards of different value and probability. A tryptophan-free diet was fed to investigate the effect of lowering central serotonin concentration on performance in the r-IGT. The data suggested that control rats behave in a way similar to rats tested in the original r-IGT; that is, they tend to choose the option with the best long-term payoff more often as the test progresses. Tryptophan depleted rats showed a weaker improvement across trials than controls.

Author Keywords

Rat, Home cage Testing, Iowa Gambling Task, Serotonin

INTRODUCTION

Recently, much effort has been devoted to developing new methods for behavioral phenotyping. Especially testing animals in their home cages has become increasingly popular. Behavior can be recorded undisturbed and continuously for prolonged periods of time, while human intervention is minimized [5], [7]. A new low-cost computer-controlled home cage operant panel (HOP) was developed by joint effort of the Istituto Superiore di Sanità, Rome, and PRS Italia, Rome. This apparatus is designed to be placed inside a standard-size (Macrolon IV) cage enabling the rodent to operate it 24 h/day. The HOP provides an alternative for yet existing home cage systems, which are either too complex or not able yet to carry out operant-learning tests. In order to test the functionality of this panel, a pilot study was carried out focusing on decision-making.

Decision-making plays an important role in everyday life of both human and non-human animals. The Iowa Gambling Task (IGT) is the most frequently used task to assess decision-making performance under uncertainty [2],[3], in which a conflict between the immediate and the long-term payoff options is represented. Decision-making is often disturbed in psychiatric conditions affected by the common human serotonin transporter promoter length polymorphism (5-HTTLPR). This fact raises the hypothesis that decisionmaking is indeed modulated by the serotonergic system [6], possibly by maintaining a choice option once established [10]. Furthermore, stress affects decision-making in a cortisol-dependent fashion [12]. The availability of a rodent version of the IGT (r-IGT) makes it possible to study the underlying mechanisms [11]. Testing rodents in their home cage could provide a less stressful environment compared to other methods, due to a reduction in human intervention, handling, and transport. Here, we report the results of an experiment in which the HOP was used in a modified r-IGT protocol. In addition, to study the effect of serotonin on IGT

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

performance, rats received a tryptophan (TRP) deficient (or control) diet to reduce brain serotonin synthesis.

MATERIALS AND METHODS

All experimental procedures were approved by Institutional Animal Survey Board on the behalf of Italian Ministry of Health, and were in close agreement with European Communities Council Directive (86/609/EEC) and Italian law. All efforts were made to minimize animal suffering, to reduce the number of animals used, and to use alternatives to in vivo testing.

Subjects

Twelve male adult Wistar rats (Charles River, Italy; mean bodyweight 365.5gr) were kept in an air-conditioned room (temperature $21 \pm 1^{\circ}$ C) on a 12-hr reversed light-dark cycle (lights on at 7.00 pm). Prior to the experiments animals were housed in pairs in Macrolon III cages, but from the start of the training/test protocol animals were singly housed. Water was available ad libitum, whereas food (Altromin-R, A. Rieper S.p.A., Vandoies, Italy) was available ad libitum until the start of the protocol. After four weeks of acclimation, rats were randomly assigned to one of two experimental groups: one group received a TRPfree diet (T-), while the other group received a control diet. The TRP-free diet (DP/1069 mod., A. Rieper S.p.A., Vandoies, Italy) was a standard diet, but with the complete lack of TRP. The control group (T+) was fed the same diet, except that TRP (2,8 gr/kg diet) was added.

Apparatus

The operant-testing apparatus, consisting of four prototype computer-controlled panels (HOP; PRS Italia, Rome), one for each of the subjects, was placed in a Macrolon IV cage with sawdust bedding. The panel contained two nose-poking holes, hole lights, a chamber light, two feeder devices, a food magazine where pellets (sugar pellets F0042 and quinine [4.44 g/kg diet] pellets F06498, Dustless Precision Pellet 45 mg; Bio-Serv, Frenchtown, NJ) were delivered, a little trapdoor to remove uneaten pellets, and a magazine light. The panel was connected through an interface to a PC, where specific software (Sca020; PRS Italia, Rome) controlled and recorded all events.

Nose poking in the holes of the panel resulted in the delivery of sugar or quinine pellets (see Protocol for ratio and amount). After nose-poking (adequate nose-poking) and before food delivery, the hole light was turned on for 1sec. Following food delivery, the magazine light was turned on for 15sec, during which nose-poking (inadequate nose-poking) was recorded but was without any scheduled consequence (time-out, TO). The trapdoor was opened 2sec before the end of the TO. The magazine light was then turned off, the chamber light was turned on, and the system was ready for the next trial.

The following variables were recorded automatically: adequate nose-poking (nose-pokes after a TO, resulting in the delivery of the pellets), inadequate nose-poking (nosepokes during a TO interval, which were recorded but were without any consequences), and time needed to complete the session. For adequate nose poking, the dependent variable was the choice preference (calculated as the percentage of adequate nose-pokes at the "bad" hole over total amount of all adequate nose-pokes). For inadequate nose poking, the dependent variable was the raw frequency of inadequate nose-pokes per trial.

Protocol

One week before the start of the training/testing protocol, rats were handled for 2min daily, their bodyweight was taken, and they were familiarized with the sugar pellets in the home cage (two pellets per animal per day). Five days before the start of the training/testing protocol the normal food was removed and animals received ad libitum the Tor control diet (animals were familiarized with the diet by receiving approximately 4gr per rat per day in the home cage, two days before the change of diet).

On the morning of the first day of training, rats were placed individually in the test cages with water available ad lib (no food), where they were left undisturbed for an hour before the first session started. Two sessions were run per day, which took place around 9.00AM and 5.00PM respectively (for arguments, see [9]). The training phase consisted of three sessions of magazine training and one session of nosepoke training, i.e. two days in total. During the magazine training, two sugar pellets were dropped automatically in the magazine with an interval of 60sec. After a variable time window, the trapdoor was opened to remove the uneaten pellets; this window was kept fixed for each session and was progressively decreased across subsequent sessions (time to eat, TTE, 20-15-10sec). Once all rats were reliably consuming the pellets, they were trained to nosepoke on each nose-poking hole (right/left alternating each trial, 10 trials per hole): they learned to collect two sugar pellets per adequate nose-poke from both holes, in an attempt to prevent a biased preference for either hole. In the magazine training phase, nose-poking holes were closed by covering Plexiglas plates, while the food magazine was always covered by an aluminium plate (to protect the underlying mechanics against sawdust entering the magazine) which was only removed during sessions.

Once the rats were nose-poking reliably and eating the sugar pellets within 10sec, the test phase started. The IGT was based on the r-IGT performed in a maze as described previously [6], [11], modified to be adapted in the home cage operant panels. Rats received 40 trials per session for a total of 240 trials; thus this phase lasted 3 days. Each session started with turning on the chamber light accompanied by the free delivery of two sugar pellets. The TO was 15sec, including a TTE of 13sec before the uneaten pellets were removed. Rewards were represented by sugar pellets, punishments were represented by quinine pellets that were unpalatable but not uneatable. The "bad" hole presented occasionally big rewards (four sugar pellets,

probability 30%) among series of quinine pellets (4, probability 70%), i.e. 12 sugar pellets per 10 trials. The "good" hole presented regularly small rewards (two sugar pellets, probability 80%) among quinine pellets (2, probability 20%), i.e. 16 pellets per 10 trials. The total number of nose-pokes for "good" and "bad" holes was calculated for each block of 40 trials to obtain the IGT score. This choice parameter (see Figure 1) was the percentage of adequate nose-pokes.

Before each session, rats received 0.5gr of T- or control diet respectively, and after each session they received the rest of their diet needed to maintain them on 95-98% of free feeding bodyweight. Daily after the AM-session, animals' bodyweight was taken. Rats were food restricted from the first day of training throughout the entire protocol in order to increase their motivation to work for food delivery. The entire training/test protocol lasted 5 days.

Data Analysis

Exclusion criteria were (1) rats that did not reliably eat the rewards, (2) rats that did not reliably nose-poke for pellet delivery, and (3) rats that ate quinine pellets. Throughout the r-IGT, repeated measures ANOVAs were performed on the preference for the "bad" hole (within subjects factor: block; between-subjects factor: treatment). Other data were analysed by ANOVAs (between-subjects factor: treatment) to detect group differences in the number of inadequate nose-pokes per trial, as well as in time needed per trial and in bodyweight (% of ad libitum bodyweight). Differences were considered significant when p = 0.05; NS = non-significant. All statistics are two-tailed.

RESULTS

Six T- and five control rats were tested in a modified version of the rodent Iowa Gambling Task [11], adapted to measure the animals' behavior in their home cage. One control subject was excluded from further testing as he did not reliably nose-poke for pellet delivery. Figure 1 shows that, independent of treatment, all subjects improved their performance over blocks (trial block: F(5,9) = 7.576, p < 0.001: trial block x treatment: F(5.45) = 0.958. NS). Visual inspection of the data suggests that T- rats do show a weaker improvement across trials than control rats. This is confirmed by separate repeated measures per group: control F(5,20)=7.63, p<0.001, vs T- F(5,25)=2.339, p<0.07. Tsubjects showed significantly more inadequate nose-pokes per trial (mean \pm SEM = 3.18 \pm 0.26) than controls (1.80 \pm 0.19; treatment F(1,9) = 7.071, p = 0.026). T- rats needed less time to complete their trials (minutes needed per trial, mean \pm SEM = 0.34 \pm 0.01) than controls (0.55 \pm 0.04) as the test progressed (trial block x treatment F(5,45) = 4.244, p = 0.003; treatment F(1.9) = 49.557, p<0.001). No difference was found in bodyweight between the groups (treatment F(1,9) = 0.079, NS).

DISCUSSION

The present study was a first attempt to translate the maze r-IGT [11] to a home cage setting using an operant panel. The data thus far indicate that rats, like humans and rodents in the original r-IGT, do seem to choose the best long-term option more often as the test progresses. However, the learning curves for present rats (see Figure 1) and rats tested in the original maze r-IGT [11] with respect to this decision-making task are not exactly similar: the present rats seem to show a slower improvement across trials than rats tested in the maze r-IGT. This may be partly due to the fact that we used a slightly different ratio of sugar and quinine pellets for the two different options, as we noted that the original ratio used in the maze was too easy in the HOP.

It has been reported in both humans and rats that serotonin transporter dosage affects long-term decision-making in the IGT, thereby substantiating an important modulatory role of the serotonergic system in decision-making [6]. In line with this, T- rats (with diet-induced TRP depletion reducing brain serotonin synthesis) seemed to show a weaker improvement across trials than controls. It is clear that more rats need to be tested to substantiate this observation. Still in line, T- rats showed more inadequate nose-pokes, which provides an index of inability to wait [1], and were faster in completing their trials than control rats. Both findings are signs of increased motor impulsivity, caused by an altered function of the serotonergic system which modulates premature and impulsive responding [4], [8].

The present findings show that in principle it is possible to run the r-IGT in a home cage setting. Future experiments are directed at further validation of this approach and refinement of the protocol.



Figure 1. Performance of T- (n = 6) and control (n = 5) rats tested in the home cage r-IGT. Shown are the mean $(\pm SEM)$ proportions of choices for the "bad" hole per block of 40 trials.

REFERENCES

- Adriani, W., Caprioli, A., Granstrem, O., Carli, M., Laviola, G. The spontaneously hypertensive-rat as an animal model of ADHD: evidence for impulsive and non-impulsive subpopulations. *Neuroscience & Biobehavioral Reviews*, 27 (2003), 639-651.
- Bechara, A., Damasio, A.R., Damasio, H., Anderson, S.W. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50 (1994), 7-15.
- Bechara, A., Damasio, H., Damasio, A.R., Lee, G.P.J. Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience*, 19 (1999), 5473–5481.
- 4. Bizot, J., Le Bihan, C., Puech, A.J., Hamon, M., Thiébot, M. Serotonin and tolerance to delay of reward in rats. *Psychopharmacology*, *146* (1999), 400-412.
- De Visser, L., van den Bos, R., Kuurman, P. W. W., Kas, M. J. H., Spruijt, B. M. Novel approach to the behavioral characterization of inbred mice: Automated home cage observations. *Genes, Brain & Behaviour, 5* (2006), 458-466.
- 6. Homberg, J.R., van den Bos, R., den Heijer, E., Suer, R., Cuppen, E. Serotonin transporter dosage modulates

long-term decision-making in rat and human. *Neuropharmacology*, 55 (2008), 80-84.

- Knapska, E., Walasek, G., Nikolaev, E., Neuhäusser-Wespy, F., Lipp, H. P., Kaczmarek, L., Werka, T. Differential involvement of the central amygdala in appetitive versus aversive learning. *Learning & Memory*, 13 (2006), 192-200.
- 8. Soubrié, P. Reconciling the role of central serotonin neurones in human and animal behavior. *Behavioral and Brain Sciences*, 9 (1986), 319-364.
- 9. Strubbe, J.H., Woods, S.C. The timing of meals. *Psychological Review*, *111* (2004), 128-141.
- 10. Van den Bos, R., Houx, B.B., Spruijt, B.M.,. The effect of reward magnitude differences on choosing disadvantageous decks in the Iowa gambling task. *Biological Psychology*, *71* (2006), 155-161.
- Van den Bos, R., Lasthuis, W., den Heijer, E., van der Harst, J., Spruijt, B. Towards a rodent model of the Iowa gambling task. *Behavior Research Methods*, 38 (2006), 470–478.
- 12. Van den Bos, R., Harteveld, M., Stoop, H. Stress and decision-making in humans: performance is related to cortisol reactivity, albeit differently in men and women. *Psychoneuroendocrinology*, 34, 10 (2009), 1449-1458.

Trainable, Vision-Based Automated Home Cage Behavioral Phenotyping

Hueihan Jhuang hueihan@mit.edu Estibaliz Garrote estibaliz.garrote@gmail.com Nicholas Edelman edelmann@mit.edu Tomaso Poggio tp@ai.mit.edu

McGovern Institute of Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Andrew Steele Broad Fellows in Brain Circuitry Program, Division of Biology, California Institute of Technology, Pasadena, CA, USA steelea@caltech.edu

Thomas Serre McGovern Institute of Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA serre@mit.edu

ABSTRACT

We describe a fully trainable computer vision system enabling the automated analysis of complex mouse behaviors. Our system computes a sequence of feature descriptors for each video sequence and a classifier is used to learn a mapping from these features to behaviors of interest. We collected a very large manually annotated video database of mouse behaviors for training and testing the system. Our system performs on par with human scoring, as measured from the ground-truth manual annotations of thousands of clips of freely behaving mice. As a validation of the system, we characterized the home cage behaviors of two standard inbred and two nonstandard mouse strains. From this data, we were able to predict the strain identity of individual mice with high accuracy.

Author Keywords

Computer vision, behavior recognition, rodent, mouse, phenotyping.

ACM Classification Keywords

I.5.4 Pattern Recognition: Computer Vision

INTRODUCTION

Automated quantitative analysis of mouse behavior will play a significant role in comprehensive phenotypic analysis – both on the small scale of detailed characterization of individual gene mutants and on the large scale of assigning gene functions across the entire mouse genome [1]. One key benefit of automating behavioral analysis arises from inherent limitations of human assessment: namely cost, time, and reproducibility. Although automation in and of itself is not a panacea for neurobehavioral experiments, it allows for addressing an entirely new set of questions about mouse behavior, such as conducting experiments on time scales that are orders of magnitude longer than traditionally assayed. For example, reported tests of grooming behavior span time scales of minutes whereas an automated analysis will allow for analysis of grooming behavior over hours or even days.

Most previous automated systems [3, 6] rely on the use of non-visual sensors (i.e. infrared beam) or video tracking techniques to monitor behavior. Such systems are particularly suitable for studies involving spatial measurements such as the distance covered by an animal or its speed. The physical measurements obtained from these sensor-based and tracking-based approaches limit the complexity of the behavior that can be measured. In particular, these approaches are not suitable for the analysis of fine animal behaviors such as grooming or micromovements of the head. A few computer-vision systems for the recognition of mice behaviors have recently been described, including a commercial system (CleverSys, Inc) and two prototypes from academic groups [2, 9]. These computer-vision systems have not vet been tested in a realworld lab setting using long, uninterrupted video sequences containing potentially ambiguous behaviors. In addition, the systems have not been comprehensively evaluated against large, human annotated video databases containing different animals and different recording sessions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

In this paper, we describe a trainable, general-purpose, automated, and potentially high-throughput system for the behavioral analysis of mice in their home cage. Developed from a computational model of motion processing in the primate visual cortex [4], our system computes a sequence of feature descriptors for each input video based on the motion and position of the mouse. In the learning stage, a classifier is trained from manually annotated labels (behaviors of interest) and used to predict an output label for every frame of the video sequence. We compare the resulting system against human labeling and existing commercial software. We also discuss a range of applications demonstrating the flexibility of this approach.

EXPERIMENTS

All experiments involving mice were approved by the MIT and Caltech committees on animal care.

Behaviors of Interest and Definition

We annotate 8 types of common behaviors of inbred mice: drinking (defined by the mouse's mouth being juxtaposed to the tip of the drinking spout), eating (defined by the mouse reaching and acquiring food from the food bin), grooming (defined by the fore- or hind-limbs sweeping across the face or torso, typically as the animal is reared up), hanging (defined by grasping of the wire bars with the fore-limbs and/or hind-limbs with at least two limbs off the ground), rearing (defined by an upright posture and forelimbs off the ground), resting (defined by inactivity or nearly complete stillness), walking (defined by ambulation) and micro-movements (defined by small movements of the animal's head or limbs). Figure 1 illustrates these typical behaviors.

Video Datasets

In order to train a set of motion templates that are useful for discriminating between behavior categories, we manually collected a dataset (*clipped dataset*) consisting of 4,200 clips with the best and most exemplary instances of each behavior (each clip contains one single behavior). This set contains different mice (differing in coat color, size, gender,



Figure 1. Snapshots taken from representative videos for the eight home cage behaviors of interest.

etc.) recorded at different times during day and night over 12 separate sessions.

Currently, the only public dataset for mice behaviors is limited in the scope [2]. In order to train and test our system on a real-world lab setting where mice behaviors are continuously observed and scored over hours or even days, we collected a second dataset (*full dataset*). This set contains 12 continuous labeled videos, in which each frame is assigned a behavior of interest. Each video is 30-60 minutes in length, resulting in a total of over 10 hours of continuously annotated videos. As in the *clipped dataset*, these videos are chosen from different mice at different times to maximize generalization of the dataset.

A team of 8 trained investigators ('Annotators group 1') manually annotated the videos. Two annotators of the 'Annotator group 1' performed a secondary screening on these annotations to correct mistakes and ensure the annotation style is consistent throughout the whole database. In order to measure the agreement between human labelers, we asked 4 of the original 8 investigators ('Annotators group 2') to label a subset of the already labeled videos (*doubly annotated dataset*). The *doubly annotated dataset* consists of many short video segments, which are randomly selected from the *full dataset*. Each segment in the *doubly annotated dataset* is 5-10 minutes long for a total of about 1.6 hours of video.

Training and Testing the System

The system computes two types of features for recognizing behaviors: the motion features developed by Jhuang et al. [4], as well as position and velocity features. Combining these two feature sets, the system learns a classifier that maps these features to the behaviors of interest.

Training based on the *clipped dataset* and the *full dataset* is done in two stages. In the first stage, we compute a set of 12,000 motion features on the clipped dataset. To reduce these 12,000 features to a more computationally tractable subset, we applied a feature selection technique called a zero-norm SVM [8] to select a subset (approximately 300) of the features that are most useful for discriminating between the behaviors categories. In the second stage, we compute the approximately 300 motion features, and the position and velocity features for the full dataset. The performance on the *full dataset* is evaluated using a leaveone-video-out cross-validation procedure: use all but one of the videos to train a classifier and the video not used in the training to evaluate the system. This process is repeated 12 times, once for each video. The system predictions for all the videos are then used to compute the accuracy as the percentage of frames correctly predicted by the system. Here a prediction or a label is 'correct' if it matches ground truth made by 'Annotators group 1'.

| | Our System | CleverSys Commerical System | Human ('Annotator Group 2') |
|---|---------------|-----------------------------------|-----------------------------------|
| Doubly annotated dataset (1.6 hours of video) | 76.6% | 60.9% | 71.6% |
| Full dataset (10 hours of video) | 77.6% | 61.0% | N/A |

Table 1. Accuracy of our system, human annotators andHomeCageScan 2.0 CleverSys system evaluated on thedoubly annotated dataset and the whole set of full dataset forthe recognition of 8 behaviors.

For the system's classifier, we used the Hidden Markov Support Vector Machine (SVMHMM) [5]. The SVMHMM models the temporal dependence between behaviors (i.e. drinking is unlikely to occur directly following hanging), and thus performed much better than a simple SVM classifier that evaluates each frame in isolation.

Comparison with Commercial Software and Humans

Here we evaluate the system performance on the doubly annotated dataset and the full dataset. The system is compared against commercial software (HomeCageScan 2.0, CleverSys, Inc) for mouse home cage behavior classification and against human manual scoring. The doubly annotated dataset benchmarks the agreement between human annotators. Table 1 shows the comparison. Overall, we found that our system achieves 76.6% agreement with human labelers ('Annotators group 1') on the doubly annotated dataset. The agreement is significantly higher than the 60.9% agreement between HomeCageScan 2.0 system and 'Annotators group 1'. Our system agreement is on par with the 71.6% agreement between human labelers, defined as the agreement between 'Annotators group 1' and 'Annotators group 2' on the doubly annotated dataset. Two online videos demonstrating the automatic scoring of the system are available at http://techtv.mit.edu/videos/5561 and http://techtv.mit.edu/videos/5562.

Running Time of the System

For performance reasons, much of the system is implemented on a graphical processing unit (GPU). After preprocessing, feature computation and classification run in nearly real time (30 frames per second).

Identifying Strain Based on Behavior

To demonstrate the applicability of this vision-based approach to large-scale phenotypic analysis, we characterized the home cage behavior of 4 strains of mice, including the wild-derived strain CAST/EiJ, the BTBR strain (a potential model of autism [7]), and two of the most popular inbred mouse strains, C57BL/6J and DBA/2J. We recorded video of seven mice of each strain during one 24hour session, encompassing a complete light-dark cycle.

From these videos, we computed patterns of behaviors for each mouse. We segmented the predictions for each 24hour video into four non-overlapping 6-hour long segments (corresponding to the first and second halves of the night, and the first and second halves of the day). For each segment, we calculated the histogram of each behaviors type (walking, hanging, etc.). The resulting 8-dimensional (one for each behavior) vectors of the four segments were then concatenated to obtain one single 32-dimensional vector (8 dimensions x 4 segments) as the pattern of behavior for each animal. The pattern of behavior corresponds to the relative frequency of each of the eight behaviors of interest, as predicted by the system, over a 24hour period. Using a leave-one-animal-out cross-validation procedure, we found that the resulting support vector machine (SVM) classifier predicted the strain of all animals with 90% accuracy.

Training the System to Handle More Complex Behaviors

To train and evaluate the performance of the system we chose the eight behaviors described above to capture standard home cage behaviors. We next asked if the system could be extended to other, more complex behaviors based on the same motion features.

We collected a new set of videos for an entirely new set of behaviors corresponding to animals interacting with "low profile" running wheels. This wheel-interaction set contains 13 fully annotated one-hour videos taken from six C57BL/6J mice. The four actions of interest are as follows: running on the wheel (defined as all 4 paws on the wheel and the wheel to be rotating), interacting with the wheel but not running (any other behavior on the wheel), awake but not interacting with wheel, and rest outside the wheel. These actions are shown in the video available at <u>http://techtv.mit.edu/videos/5567</u>. Using the leave-one-video-out cross-validation procedure as in the *full dataset*, the system achieves 92.8% of accuracy.

Although we certainly cannot generalize to all types of behaviors, the wheel results demonstrate that for many typical mouse behaviors no additional features need to be designed: the system learns new actions from annotated examples of new behaviors.

We have on-going work in the monitoring and analysis of abnormal behaviors in the context of neurological disorders. In particular, we are currently training the system to automatically detect and rate the severity of dyskinetic movements in the context of Parkinson's disease. Results will be presented at the meeting.

CONCLUSION

We have applied a biological model of motion processing to the recognition of mice behaviors. For common behaviors of interest, the system achieves performance on par with human observers. The system demonstrates the promise of learning-based and vision-based techniques in complementing existing approaches towards a complete quantitative phenotyping of complex behavior.

ACKNOWLEDGMENTS

This research would not have been possible without the dedicated work of our data collectors and annotators: Jessi Ambrose, Andrea Farbe, Cynthia Hsu, Alexandra Jiang, Grant Kadokura, Xuying Li, Anjali Patel, Kelsey Von Tish, Emma Tolley, Ye Yao and Eli T Williams.

This research was sponsored by grants from DARPA (IPTO and DSO), NSF-0640097, NSF-0827427, IIT, and the McGovern Institute for Brain Research. Andrew Steele was funded by the Broad Fellows Program in Brain Circuitry at Caltech. Hueihan Jhuang was funded by the Taiwan National Science Council (TMS-094-1-A032).

REFERENCES

- J. Auwerx and A. El. The European dimension for the mouse genome mutagenesis program. *Nature Genetics*, 36, 9 (2004), 925-927.
- 2. P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. *Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2005.
- 3. E. H. Goulding, A. K. Schenk, P. Juneja, A. W. Mackay, J. M. Wade, and L. H. Tecott. A robust automated

system elucidates mouse home cage behavioral structure. *Proceedings of the National Academy of Sciences USA*, 105, 22 (2008) 20575–20582.

- 4. H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. *IEEE International Conference on Computer Vision*, 2007.
- 5. T. Joachims. http://www.cs.cornell.edu/People/tj/svm_light/svm_hm m.html. Version V3.10, 2008.
- 6. L. P. Noldus, A. J. Spink, and R. A. Tegelenbosch. EthoVision: a versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, Computers, 33*, 3 (2001), 398-414.
- H. G. Mcfarlane, G. K. Kusek, M. Yang, J. L. Phoenix, V. J. Bolivar, and J. N. Crawley. Autism-like behavioral phenotypes in BTBR T1tf/J mice. *Genes, Brain and Behavior*, 7 (2008), 152-163,.
- 8. J. Weston, S. Mukhrejee, O. Chapelle, M. Pontil, T. Poggio, V. Vapnik. Feature Selection for SVMs. *Advances in Neural Information Processing*, 2001.
- 9. X. Xue and T. C. Henderson. Feature fusion for basic behavior unit segmentation from video sequences. *Robotics and Autonomous Systems*, 57, 3 (2009), 239-248.

Refinement Through Better Animal Monitoring: How Behavioural Researchers Can Contribute

Penny Hawkins

Research Animals Department, Science Group RSPCA, Wilberforce Way, Southwater, West Sussex, RH13 9RS United Kingdom phawkins@rspca.org.uk +44 870 7540 231

ABSTRACT

Researchers who measure animal behaviour have already contributed to the ability to assess animal welfare and recognise indicators of pain, suffering and distress. There are a number of areas where more research is needed to enable further progress to be made. A "wish list" of behavioural research topics can be drawn up on the basis of working group reports and the views of animal users and care staff. This list aims to inform further research directions and collaborations for behavioural researchers.

Author Keywords

Welfare assessment, pain assessment, refinement.

INTRODUCTION

New techniques in behavioural research have provided unprecedented insights into animals' responses to their housing and care, scientific procedures, interactions with humans and with one another. This has helped to provide a sound, scientific basis for the refinement of husbandry and procedures to reduce suffering and improve welfare, which is an essential component of humane science [1].

Successful refinement also depends upon the ability to assess animal welfare effectively, and detect any signs of pain or distress as rapidly as possible, so that any suffering can be alleviated. Researchers who measure animal behaviour can play an important role in further developing welfare assessment, especially if they collaborate with those who are responsible for monitoring and assessing animals [2].

AN IDEAL WELFARE ASSESSMENT SYSTEM

A survey of researchers, veterinarians, animal technologists and care staff in the UK identified a number of obstacles to the effective assessment of pain, suffering or distress [3]. These included a strong element of subjectivity when assessing welfare and a lack of time to observe animals properly. On this basis, there is a need for behavioural indicators of welfare that will:

- Be objective it should be possible to clearly describe and define the indicators, e.g. flank twitch in rodents [4].
- Require minimal training for observers to be able to recognise them reliably [4].
- Complement the clinical judgement of experienced and empathetic staff.
- Save time the ideal welfare assessment system would include a small number of indicators that occur relatively frequently.

Behavioural researchers have already contributed to defining indicators that fit these criteria. However, there is still an unmet need for more indicators that are tailored to individual species, strains and type of procedure.

A "WISH LIST" OF RESEARCH TOPICS

Several working groups have addressed the issue of pain and distress alleviation in laboratory animals [2,5-8]. Their reports include many common themes relating to pain and distress assessment such as the role of ethics and/or animal care and use committees, training, and ensuring that there are sufficient financial and human resources to observe and monitor animals effectively.

The "wish list" of suggested research topics relating to pain and distress assessment below is drawn from recent working group reports [2,5-8], a survey of animal users and care staff [3], and recommendations in a new Joint Working Group on Refinement (JWGR)¹ report on defining and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

¹ The JWGR comprises the British Veterinary Association Animal Welfare Foundation, Fund for the Replacement of Animals in Medical Experiments, Royal Society for the Prevention of Cruelty to Animals and Universities Federation for Animal Welfare.

http://www.rspca.org.uk/sciencegroup/researchanimals/implementing3rs/refinement

implementing protocols for the welfare assessment of laboratory animals [1].

- Define behaviours that are indicative of discomfort, pain or distress and that occur frequently and are easy for human observers to assess.
- Develop possible distress predictors to be used as outcome scores for laboratory animals, i.e. to predict severity in clinical outcomes and help to implement humane endpoints.
- Evaluate possible associations between disease behaviours or abnormal behaviours (e.g. sterotypies) and pain, suffering or distress.
- Evaluate physiological correlates of suffering that may provide useful adjuncts to behavioural observations, for example where heart rate or body temperature are available from telemetry transmitters that have been implanted for experimental purposes.
- Study the influence of an animal's characteristics (e.g. gender, age, genetic makeup) on indicators of pain or distress.
- Determine which easily assessable parameters provide the most useful information for the welfare assessment of genetically altered (GA) animals.
- Identify refinements in euthanasia methods.
- Determine optimum times for welfare assessment for different species and strains with respect to circadian rhythms, e.g. what difference it makes to the quality of the assessment when rodents are disturbed during their sleep phase or observed when they would usually be active.
- Define behaviours that could be used as indicators of positive welfare.

These suggested topics could all be progressed by researchers with expertise in measuring animal behaviour. As with the requirement for more generic indicator evaluation studies above, it should be possible to conduct these studies using animals involved in ongoing research programmes, leading to benefits for animal welfare and science without imposing any significant additional harms to experimental subjects.

CONCLUSION

The need for good collaboration between different research disciplines as a way of facilitating better welfare assessment was a common theme within the working group reports. For example, the US National Research Council committee stated that "Animal welfare scientists, and researchers and scientists who use animal models, should communicate with each other more frequently in order to compare objectives and progress and to identify opportunities for collaboration" [2]. This presentation aims to stimulate further dialogue and discussion so that behavioural researchers wishing to carry out applied research can help to fill outstanding knowledge gaps in the field of pain and distress assessment, enabling better science and animal welfare.

REFERENCES

- 1. Hawkins, P. et al. Defining and implementing protocols for the welfare assessment of laboratory animals: a BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement guide. *Laboratory Animals* (submitted).
- 2. National Research Council. *Recognition and Alleviation of Distress in Laboratory Animals*. National Academies Press, Washington DC, 2008. http://www.nap.edu/catalog.php?record_id=11931
- 3. Hawkins, P. Recognizing and assessing pain, suffering and distress in laboratory animals: a survey of current practice in the UK with recommendations. *Laboratory Animals 36* (2002), 378-395. http://tinyurl.com/yc37duj
- 4. Roughan J.V., and Flecknell, P.A. Behavioural effects of laparotomy and analgesic effects of ketoprofen and carprofen in rats. *Pain*, *90* (2001), 65-74.
- 5. Brown, M. et al. Report of the Working Group on Animal Distress in the Laboratory. *Lab Animal Europe*, 6 (2006), 41-45.
- 6. GA Mouse Welfare Assessment Working Group. Assessing the Welfare of GA Mice. *Laboratory Animals*, 40 (2006), 111-114. http://tinyurl.com/y9g8nku
- 7. Karas, A.Z. Barriers to assessment and treatment of pain in laboratory animals. *Lab Animal*, *35* (2006), 38-45.
- 8. Paul-Murphy, J. et al. The need for a cross-species approach to the study of pain in animals. *JAVMA*, 224 (2004), 692-697.

Natural Colour Preference in the Zebrafish (Danio rerio)

Avdesh Avdesh^{1, 3, 4} chaudharyavdesh@gmail.com Mathew T. Martin-Iverson^{1, 2, 3}

Alinda Mondal^{4, 5}

Mengqi Chen^{1,4}

Giuseppe Verdile^{1, 4, 5}

Ralph N. Martins^{1, 3, 4, 5}

 ¹Centre of Excellence for Alzheimer's Disease Research and Care, Edith Cowan University, Perth, Western Australia, Australia,
 ²School of Pharmacology and Medicine, University of Western Australia, Perth, Western Australia, Australia
 ³Centre for Clinical Research in Neuropsychiatry, Graylands Hospital, University of Western Australia, Perth, Western Australia, Australia
 ⁴School of Exercise, Biomedical and Health sciences, Edith Cowan University, Perth, Western Australia, Australia
 ⁵McCuskar Foundation for Alzheimer's Disease Research, Edith Cowan University, Perth, Western Australia, Australia

ABSTRACT

The zebrafish (Danio rerio), traditionally used in genetics is now becoming increasingly popular in behavioural neuroscience. Zebrafish is a small teleost (bony) freshwater fish originating from northern India, used worldwide in research in most stages of development from embryo to adult. While, a couple of papers on zebrafish related to visual colour discrimination and learning have been published [1, 2], its natural colour preference has received little attention [2]. Of particular interest is the receiver bias towards a specific colour. Natural colour preference towards a specific colour may lead to changes in visual discrimination learning, memory and decision making of an animal. Zebrafish has been shown to have colour vision with peak absorbance in ultraviolet (362nm), blue (415nm), green (480nm) and red (570nm) [2]. In the present study we used four different colours (red, yellow, green and blue) to test the natural colour preference of zebrafish. A conditioned place preference (CPP) apparatus using six different colour combinations of gravels was used to measure the time spent in each coloured compartment. This study might be helpful in developing further colour based learning and memory paradigms in zebrafish.

Author Keywords

Zebrafish, natural colour preference, learning and memory.

INTRODUCTION

Zebrafish (*Danio rerio*) is now becoming popular in behavioural neuroscience [5]; although it's natural colour preference has received little attention. In the present study we have used four different colours (red, yellow, green and blue) to test natural colour preference of the zebrafish. To carry out the investigations, fish were divided into six groups (six colour combinations) and tested for a place preference in a conditioned place preference apparatus. In our laboratory we are trying to develop novel learning and memory paradigms for zebrafish. For this purpose we needed to study the probable natural preference and/ or aversion for a specific colour.

MATERIALS AND METHODS

The fish were tested in a two-chambered conditioned place preference (CPP) box as described previously [3] with slight modifications shown in Figure 1. Briefly, the test apparatus was 23cmX15cmX15cm and was filled with water up to 12cm from bottom to minimize stress. The apparatus was divided into two equal halves with a perforated wall that allows complete albeit impeded movement. System water maintained at 28°C and pH of 7.3 (maintained by sodium bicarbonate) was used for the experiments. The apparatus used in the experiment is shown in Figure1A and B.

Animal Care and Maintenance

The principles and procedures were conducted in accordance with National Health and Medical Research Council, Australia (NHMRC), Guidelines for the Care and Use of Laboratory Animals.72 mixed sex wild type (AB) zebrafish were used for the experiment. Zebrafish were housed in an environmentally controlled room with 14:10h

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Conditioned place preference apparatus with colored gravels (A) and with a perforated wall (B) that allows complete, albeit impeded, movement of fish from one compartment to another.

light: dark cycle in aquaculture research facility. Fish were fed twice daily with live brine shrimp in the morning and dry food during evening. Red ocean salt was added to the system water at a concentration of 0.3 mg/ml. Other maintenance conditions were followed as described previously [4]. Fish were separated in 3 litres tanks (12 fish in a tank) two days prior to habituation in the testing apparatus. All the fish tested were 25 weeks old and were F3 generations of AB strain.

Procedure

All the fish were habituated in groups in the testing apparatus for four days. Since zebrafish is a shoaling fish, the number of fish per group was slowly reduced in subsequent habituation trials. On the first day, fish were given 5 minutes to fully explore the apparatus. In the next two consecutive days, fish were restricted to each side for 4 minutes using a transparent divider. On the fourth day, the apparatus was partitioned into two equal halves with a perforated wall that allows complete, albeit impeded, movement from one chamber to another. Fish were allowed to explore for 4 minutes. On the following day, the apparatus was equally filled with two different colours. Each fish was tested individually (n=12 per group). Since four colours were used, six different colour combinations were tested (red + yellow, red + green, red + blue, yellow + green, yellow + blue, green + blue). Each fish was placed in one side of CPP box and time spent in each compartment was recorded. The position and side of entry of each fish was counterbalanced. Number of entries into each compartment was recorded.

Reflectance spectra of these different coloured gravels were also measured using a spectrometer (Ocean optics, USB, 2000 with a halogen lamp light source. This was done to measure the wavelengths of the coloured gravels to compare the wavelengths range with the photo pigment spectral sensitivity of zebrafish.

Analysis

Data were analysed with a two-tailed t-test for the preference of one colour over another in each combination using R.2.10 software.

RESULTS

We have found a strong aversion of the blue colour relative to all other colours (red, yellow and green) when tested in





Figure 2. Results showing the time spent (seconds) in the coloured compartments when fish were tested for a combination of colours in CPP.

combinations. No preferences were found amongst the other red, yellow and green. Results are shown in Figure 2.

CONCLUSION

The present results are helpful in choosing colours to use in future colour-based learning and memory paradigms by considering natural colour preferences and aversions of zebrafish. Reds, yellows and greens apparently are equally pleasant or aversive, and are good choices for appetitive experiments. Blues are more aversive than all others, and might be useful for experiments involving aversion, anxiety or fear.

Ethical Statement

All the experimental procedures have been approved by Edith Cowan University Animal Research Ethics Committee.

REFERENCES

- Colwill, R. M., Raymond, M. P. et al. Visual discrimination learning in zebrafish (Danio rerio). Behav Processes 70,1(2005), 19-31.
- 2. Spence, R. and Smith, C. Innate and learned Colour Preference in the Zebrafish, Danio rerio. Ethology 114 (2008), 582-588.
- Darland, T. and Dowling, J. E. Behavioral screening for cocaine sensitivity in mutagenized zebrafish. Proc Natl Acad Sci U S A 98,20 (2001), 11691-11696.
- 4. Westerfield, M., Ed. The Zebrafish Book. A Guide for Laboratory Use of the Zebrafish (Brachydanio Rerio). University of Oregon Press (2007).
- 5. Sison M, Gerlai R. Associative learning in zebrafish (Danio rerio) in the plus maze. Behav Brain Res. 207 (2010), 99-104.

Estimating Slope and Level Change in N=1 Designs

Antonio Solanas

Department of Behavioral Sciences Methods University of Barcelona Pg. Vall d'Hebron 171, 08035 Barcelona, Spain antonio.solanas@ub.edu Patrick Onghena Centre for Methodology of Educational Research Katholieke Universiteit Leuven Tiensestraat 102 - bus 03700, B-3000 Leuven, Belgium patrick.onghena@ped.kuleuven.be **Rumen Manolov**

Department of Behavioral Sciences Methods University of Barcelona Pg. Vall d'Hebron 171, 08035 Barcelona, Spain rrumenov13@ub.edu

ABSTRACT

The study proposes a new procedure referred to as Slope and Level Change (SLC) for separately estimating slope change and level change between two adjacent phases in single-case designs, for instance between a baseline phase (denoted by A) and a treatment phase (denoted by B). Firstly, SLC eliminates baseline trend from the data series using a differencing method, instead of the commonly used regression. Secondly, the method used for estimating trend is also used for estimating slope change. After the slope change has been controlled for, the pure level change is estimated. The steps necessary to obtain the estimates are presented in detail, explained, and illustrated by means of a fictitious data example. A simulation study is carried out to explore the bias and precision of the estimators and compare them to an analytical procedure matching the data simulation model. The results suggest that the SLC estimates are unbiased for all levels of autocorrelation tested and control effectively for trend. An R code was developed to make the application of the procedure automatic.

Author Keywords

Single-case designs, level change, slope change, trend, autocorrelation.

INTRODUCTION

The need to complement the widely extended p values with measures of the strength of association between an independent and a dependent variable has been highlighted decades ago [6,13], since the former only focus on the null hypothesis. In recent years it has become widely accepted that not only group-design but also single-case (N=1) design studies need to provide scientific evidence on the

interventions applied [9,15]. The use of magnitude of effect measures would allow subsequent N=1 studies to have a solid foundation and would also make possible to perform meta-analyses.

In N=1 studies, effect sizes have been commonly expressed as amount of variability in the behavior accounted for by treatment introduction [16] or as the amount of overlap between data pertaining to different conditions [14]. Evidence suggests that the procedures based on regression analysis perform less than optimally [3,4]. Specifically, they do not distinguish sufficiently between presence and absence of treatment effect and the R^2 values they provide are distorted when data are sequentially related. The procedures based on data overlap do not seem to be that affected by autocorrelation, but the presence of general trend in data remains problematic [10].

The present study arises from the need to develop a procedure that quantifies precisely the amount of behavioral change present in data. In addition, SLC responds to the call for two separate estimates of level and slope change in single-case studies [2].

PROCEDURE PROPOSED

Rationale

The objective of SLC is to estimate level change and slope change eliminating baseline data linear trend whenever it is present. It is conjectured that the procedure may also deal with positive serial dependence, since the presence of large positive autocorrelation in data can be represented by an upward or a downward trend. In contrast, when measurements are negatively autocorrelated, data present greater variability rather than trend. The slope and level change estimates obtained for the detrended data express the shifts in terms of the measurement units. For instance, if the frequency of behavior is measured, a slope change of 3 would mean that at each measurement time during phase B the experimental unit produces an average of three behaviors more than in the previous observation point. A level change of 3 would imply that with the introduction of the intervention (i.e., with the change in phase) the experimental unit produces an average of 3 behaviors more

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

in the treatment phase than in the baseline phase. It has to be remarked that this average change in level is computed after eliminating the previously estimated slope change.

Steps Required to Compute the Procedure

Since SLC is designed to control general trend prior to assessing intervention effectiveness, an initial data correction step involves eliminating phase A (baseline) trend from data. Trend is estimated only for the baseline phase, since in that way it is possible to avoid confusion between trend and potential intervention effects taking place in phase B (treatment) [1]. The phase A trend is estimated as the mean of the differenced phase A measurements. These steps can be expressed as differencing, Equation (1), and averaging the differenced data, Equation (2):

$$\Delta A_t = A_{t+1} - A_t$$
 Equation (1)

$$\overline{\Delta A} = \sum_{t=1}^{n_A - 1} \Delta A_t / (n_A - 1) = \widehat{\beta}_A \qquad \text{Equation (2)}$$

where A_t represents the original phase A measurement at time t, ΔA_t represents the differenced phase A data, and n_A is the number of observations in the baseline phase. Consider a fictitious example of data series consisting of the following values 1, 2, 3, 4, and 5 for phase A, and 7, 9, 11, 13, and 15 for phase B. The phase A trend estimate $\hat{\beta}_A$ is used to correct data (i.e., remove trend) and the abovementioned measurements the differenced phase A data will contain only 1's and their mean (the estimate of trend) would also be one. Trend is eliminated from A data via Equation (3), and from phase B data via Equation (4):

$$\widetilde{A}_{t} = A_{t} - \widehat{\beta}_{A} \cdot t \qquad \text{Equation (3)}$$

$$\widetilde{B}_{t} = B_{t} - \widehat{\beta}_{A} \cdot t \qquad \text{Equation (4)}$$

where \widetilde{A}_i and \widetilde{B}_i represents detrended phase A and phase B data, respectively, and t is the measurement time. After the correction the phase A data will be all 0's and the phase B data: 1, 2, 3, 4, 5. This method for detrending data has been shown to be useful for dealing not only with trend but also with autocorrelation [12].

The second step involves estimating slope change as the trend present in the phase B data, from which baseline trend has already been removed. The detrended phase B data is differenced according to Equation (5):

$$\Delta \widetilde{B}_t = \widetilde{B}_{t+1} - \widetilde{B}_t \qquad \qquad \text{Equation (5)}$$

The differenced and already detrended phase B data of the example consists of four 1's. The mean of these differenced

measurements is computed as shown in Equation (6) obtaining the phase B trend estimate \widehat{SC}

$$\overline{\Delta \widetilde{B}} = \sum_{t=n_A+1}^{n_A+n_B-1} \Delta \widetilde{B}_t / (n_B-1) = \widehat{SC} \qquad \text{Equation (6)}$$

where n_B is the number of observations in the treatment phase. The average value is assumed to represent an estimation of slope change, \widehat{SC} , considering that the phase A trend has been previously removed. In the fictitious case the average of four 1's is 1 and so the slope change is estimated as 1.

The third step consists in the estimation of level change. Firstly, the already estimated change in slope is eliminated from the treatment phase data, without removing the intercept. That is, the phase B slope is eliminated from the detrended phase B data, while maintaining potential shifts taking place at time n_A+1 . This is achieved through Equation (7):

$$\widetilde{\widetilde{B}}_{t} = \left(\widetilde{B}_{t} - \widehat{SC} \cdot (t-1)\right) \qquad \text{Equation (7)}$$

When from the detrended phase B data (1, 2, 3, 4, and 5) the slope change of 1 is removed, the following values are obtained 1, 1, 1, 1, and 1 and represented the phase B data with trend and slope change eliminated.

Level change (LC) is estimated subtracting the detrended baseline data mean from the detrended and slope-changecontrolled treatment data mean, as shown in Equation (8):

$$\widehat{LC} = \overline{\widetilde{B}} - \overline{\widetilde{A}} = \sum_{t=n_A+1}^{n_A+n_B} \widetilde{B}_t / n_B - \sum_{t=1}^{n_A} \widetilde{A}_t / n_A \qquad \text{Equation (8)}$$

Through this expression level change is estimated in terms of average level of behavior in both phases. The phase A data after removing trend is represented by five 0's, while the phase B data after removing trend and slope change consists of five 1's, so level change is equal to 1-0=1. Both the slope and the level change estimates represent precisely the parameters used to construct (without random variability) the fictitious data set,.

The procedure described is not restricted to AB designs and can be applied to any combination of a baseline and treatment phase which is included in more complex design structures (e.g., multiple-baselines designs, ABAB designs).

METHOD

Data Generation

In order to test the performance of SLC for data series with random fluctuations and with known features, Monte Carlo simulation was used. The design structure studied was AB with the following series lengths $(n_A + n_B)$: 5+5, 5+10, 7+8, 10+10, 15+15, 20+20. The model used for data generation was the one presented in [8]: $y_t = \beta_0 + \beta_1 \cdot T_t + \beta_2 \cdot LC_t + \beta_2 \cdot LC_t$

 $\beta_3 \cdot SC_t + \varepsilon_t$, where y_t is the value of the dependent variable at moment t, β_0 is the intercept (set to zero), β_1 , β_2 , and β_3 are the coefficients associated with trend, slope change, and level change, respectively, T_t is a dummy time variable (taking values from 1 to $n_A + n_B$), LC_t is a dummy variable for level change (equal to 0 for phase A and to 1 for phase B), SC_t is a dummy variable for slope change being equal to 0 for phase A, and taking values from 0 to (n_B-1) for phase B, and ε_t is the error term. The beta parameters related to trend, level and slope change were set to 1 and 10 to represent a small and a large effect, respectively.

The error term (ε_t) was generated following the models assumed to represent adequately the greater part of behavioral data [7]: a) the first-order autoregressive model AR(1) $\varepsilon_t = \varphi_1 \cdot \varepsilon_{t-1} + u_t$, with φ_1 ranging from -0.9 to 0.9 with steps of .1; and b) the first-order moving average model MA(1) $\varepsilon_t = u_t - \theta_1 \cdot u_{t-1}$ with 19 values of θ_1 , which according to the relationship between MA(1) and AR(1) processes leads to parameter φ_1 values ranging from 0.4972 to -0.4972.

For both models the random variable u_t was generated following three distributions (exponential, normal, and uniform) in order to study the effect of skewness and kurtosis on the performance of SLC.

Data Analysis

For each experimental condition the mean and variance of the estimators was computed on the basis of 100,000 samples. The bias of the estimators was obtained as the difference between the simulation parameters and the estimates for slope and level change. The variance of the estimators was computed as an indicator of efficiency and a comparison was performed between SLC and a simultaneous multiple regression (SMR) procedure whose model corresponds exactly to the data parameters used [8].

Simulation

The simulation was carried out by means of Fortran programs – one for each combination of data generation process [AR(1) or MA(1)] and distribution of the random variable u_t term (exponential, normal, or uniform). Each program consisted of the following steps: 1) series length selection; 2) specification of the value of φ_1 or θ_1 ; 3) specification of the β_1 , β_2 , and β_3 parameters; 4) 100,000 iterations of steps 5 to 8; 5) generate the error term ε_t according to the generation process and the u_t distribution; 6) obtain the values for the dummy variables T, LC, and SC; 7) obtain y_t ; 8) apply SLC and SMR; 9) obtain the mean and variance of the estimates for both estimators of each procedure.

RESULTS

For both independent and serially related data with and without general trend, SLC and SMR provide unbiased estimates of the treatment effect. Therefore, the procedure proposed controls effectively for trend and is also unaffected by the lack of independence in data. This finding is general for both data generation processes and for the three error distributions.

Regarding the variability of the SLC estimators, it is increased by high negative autocorrelation, when data generated by a MA(1) process, and when there are less data points in the series.

CONCLUSION

SLC is unbiased both for first-order autoregressive and moving average processes and regardless of the distributional shape of the random variable. The SLC estimators are generally more efficient (i.e., less variable) than the SMR ones for positively autocorrelated data. Considering the large effects typical present in single-case data [5], the variability of the estimators does not seem excessive.

SLC is less efficient than the technique matching perfectly the data simulation model only for high negative serial dependence and when treatment effect is expressed as level change. Regarding the former, high negative autocorrelation is not frequent in N=1 data [12], while in relation to the latter it has to be remarked that in psychological studies an abrupt and sustained (level) change in the behavior is less likely to occur than a progressive change representing a gradual improvement of the individual or group treated. As the performance of the procedure proposed was tested using data simulated using Monte Carlo methods, further evidence on the ease and meaningfulness of SLC's application can be provided using real behavioral data.

Further research is needed in order to explore whether the data correction present in SLC can attenuate the effect of nonlinear trends. Future efforts may also focus on estimating the sampling distribution of the slope and level change estimators, due to its utility for obtaining statistical significance and, more importantly, confidence intervals.

REFERENCES

- 1. Allison, D.B. and Gorman, B.S. Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy 31*, 6 (1993), 621-631.
- 2. Beretvas, S.N. and Chung, H. A review of metaanalyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention* 2, 3 (2008), 129-141.
- 3. Beretvas, S.N. and Chung, H. An evaluation of modified R²-change effect size indices for single-subject experimental designs. *Evidence-Based Communication Assessment and Intervention* 2, 3 (2008), 120-128.
- 4. Brossart, D.F., Parker, R.I., Olson, E.A. and Mahadevan, L. The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification 30*, 5 (2006), 531-563.

- Campbell, J.M. Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification* 28, 2 (2004), 234-246.
- 6. Cohen, J. The earth is round (p < .05). American Psychologist 49, 12 (1994), 997-1003.
- 7. Harrop, J.W. and Velicer, W.F. A comparison of alternative approaches to the analysis of interrupted time-series. *Multivariate Behavioral Research* 20, 1 (1985), 27-44.
- Huitema, B.E. and McKean, J.W. Design specification issues in time-series intervention models. *Educational* and Psychological Measurement 60, 1 (2000), 38-58.
- Jenson, W.R., Clark, E., Kircher, J.C. and Kristjansson, S.D. Statistical reform: Evidence-based practice, metaanalyses, and single subject designs. *Psychology in the Schools* 44, 5 (2007), 483-493.
- Manolov, R. and Solanas, A. Comparing N = 1 effect size indices in presence of autocorrelation. *Behavior Modification 32*, 6 (2008), 860-875.

- 11. Manolov, R. and Solanas, A. Percentage of nonoverlapping corrected data. *Behavior Research Methods* 41, 4 (2009), 1262-1271.
- Parker, R.I. Increased reliability for single-case research results: Is bootstrap the answer? *Behavior Therapy 37*, 4 (2006), 326-338.
- 13. Rosnow, R.L. and Rosenthal, R. Statistical procedures and the justification of knowledge in psychological science. *American Psychologist* 44, 10 (1989), 1276-1284.
- 14. Scruggs, T.E., Mastropieri, M.A. and Casto, G. The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education 8*, 2 (1987), 24-33.
- 15. Shadish, W.R., Rindskof, D.M. and Hedges, L.V. The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention* 2, 3 (2008), 188-196.
- White, D.M., Rusch, F.R., Kazdin, A.E. and Hartmann, D.P. Applications of meta-analysis in individual subject research. *Behavioral Assessment* 11, 3 (1989), 281-296.

DyaDA: An R Package for Dyadic Data Analysis

David Leiva dleivaur@ub.edu Antonio Solanas antonio.solanas@ub.edu

Department of Behavioural Sciences Methods. University of Barcelona Pg. Vall d'Hebron, 171, 08035-Barcelona, Spain

Han de Vries Department of Behavioural Biology Utrecht University Padualaan 8, 3584 CH Utrecht, The Netherlands J.deVries1@uu.nl David A. Kenny Department of Psychology. University of Connecticut 406 Babbidge Road, Unit 1020 Storrs, CT 06269-1020, USA david.kenny@uconn.edu

ABSTRACT

The current research presents a new statistical software to conduct a wide range of statistical analyses for dyadic interaction data. Specifically, an R package, called DyaDA, is being developed to measure and test different characteristics concerning dyadic interaction. The main aim of the DyaDA package is to provide a free statistical software that includes a wide set of statistical methods developed in fields as ethology, social psychology, communications, family studies, and sociology. These statistical methods cover the study of social phenomena as dominance, social reciprocity, mutual influence and dyadic nonindependence. The common feature of all statistical methods included in the package is that dyads are taken the units of analysis. The statistical software here presented can be useful for those social researchers interested in studying groups from a dyadic perspective.

Author Keywords

Dyadic data analysis, statistical tests, R statistical software, R packages

INTRODUCTION

Social interaction can be defined as any sort of behavior that manifests itself in a group context involving at least two individuals that have an influence on each other (i.e., a dyad). Two different approaches have been proposed to study data emanating from such social interactions: the dynamic and static analyses of social interaction [8]. The dynamic approach is concerned with sequences of behavior over time and seeks to identify patterns and relations between different kinds of events. On the other hand, the static approach assumes stability of behavior over time (i.e., established relationships). The present study is primarily concerned with the static approach because all the statistical procedures here presented assume that patterns of behavior remain unchanged over time.

In the above definition of social interaction the concept of the dyad was mentioned. Dyad is the most basic structure in a group. It represents the unit of analysis adopted in most social interaction studies to explain phenomena as social dominance, social reciprocity and interpersonal perception. In this regard, social scientists often use sociomatrices when analyzing dyadic data.

In the field of ethological research, statistical methods have been proposed to study social dominance, rank order in (near-)linear hierarchies, correlation between sociomatrices, and social reciprocity. Regarding dominance hierarchy, several indices and statistical tests have been developed to measure and test this attribute in groups [3,5]. Other statistical methods have been proposed to rank individuals as a function of the outcomes of dyadic dominance encounters [4,5]. As for reciprocity, interchange, and other social interaction patterns, rowwise matrix correlation indices between two matrices, have been recommended to analyze interaction data. Procedures based on permutation tests have been developed to make statistical decisions [2,6].

In the context of social psychology, interdependence is, beyond all doubt, the main feature of dyadic data analyses. Several statistical methods have been proposed to evaluate nonindependence for different dyadic data structures, including the standard dyadic designs [1] in which each person has one interaction partner and the Social Relations Model [SRM; 10]. Focusing on the SRM, social interaction data can be decomposed into its variance components and, subsequently, actor, partner, and relationship effects can be

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

estimated by means of a two-way ANOVA. This statistical model also enables social researchers to estimate dyadic and generalized reciprocity as correlation coefficient values [15]. More recently, other measures of social reciprocity based on dyadic discrepancies have been developed to measure and test this social aspect at group, dyadic and individual levels [14].

The main aim of the present study is to develop statistical software that includes some of the statistical methods mentioned above. The performance of the DyaDA package is shown by means of some examples from applied research. This software can be useful for social scientists interested in dyadic data analysis.

THE DYADA PACKAGE

The primary aim in designing the DyaDA package is to provide free statistical software for measuring and testing several types of social structure in dyadic interaction data. An excellent way to achieve this is by writing an R package for these measures and tests. R [12] is a free, open-source statistical computing language and a statistical platform that can be run on Linux, MacOS and Windows. It can be downloaded from the Comprehensive R Archive Network (CRAN) at http://CRAN.R-project.org/. The DyaDA package can be obtained from the authors on request. This package will be also available at the CRAN web page. Once the DyaDA package has been installed, users can load it into the current R work session and use it very easily. One of the main reasons of creating DyaDA as an R package is that social researchers can benefit from the wide variety of powerful tools for statistical computation and graphics integrated in R.

The DyaDA package was firstly thought to consist of statistical methods to study the following social phenomena:

- 1. Social dominance hierarchies.
- 2. Rank order in linear hierarchies.
- 3. Correlation between sociomatrices.
- 4. Social reciprocity.
- 5. Dyadic nonindependence.

As regards social dominance hierarchies, the DyaDA package includes functions to measure and test social dominance at different levels of analysis. Specifically, it integrates several procedures to quantify dominance at dyadic (e. g., number of one and two-way dyads), group (e. g., h and h' indices [3]), and individual levels (e. g., David's Scores [5]).

To find a rank order consistent with a linear dominance hierarchy, several functions are implemented in the R package (e.g., I&SI method [4]).

The package also incorporates a number of different functions to assess the association between two

sociomatrices by means of the Pearson, Kendall or Spearman rowwise matrix correlation coefficients [2,6]. It also includes statistical methods to test the correlation between two sociomatrices when the effect of a third variable is partialled out [2,7]. Some of the previous correlation indices have been proposed as social reciprocity measures [6], but in the package are also implemented social reciprocity measures based on dyadic discrepancies (e.g. asymmetry indices at group, dyadic and individual levels [14]). Finally, as regards quantifying dyadic interdependence, DyaDA incorporates a set of functions to describe and make statistical decisions for the most used dyadic data structures. For instance, methods to be used in standard dyadic designs [1] and statistical procedures for other structures as round robin and one-with-many designs [9,10].

AN EXAMPLE

In this section an illustrative example is provided in order to demonstrate how the DyaDA package functions. In so doing, data drawn from psychological research are analyzed using this package. Specifically, a sociomatrix containing the interpersonal perceptions of six undergraduate students regarding their partners' contributions to the group performance [13] is used to quantify and test SRM variances and covariances. Then, users can use several DyaDA functions to obtain actor, partners and relationship effects and variances (e.g., SRM.effects and SRM.variances functions), as well as, estimating statistical significance for several variances and covariances the (e.g., SRM.within.groups.tTest function). The following table contains the interpersonal perceptions of the six students as well as the actor and partner effects estimated by means of SRM.effects function included in DyaDA:

| | | A | B | Partner | D | E | F | Actor Effects |
|----------------|----------|------|------|---------|-----|------|-----|------------------|
| | Α | - | 6 | 3 | 2 | 6 | 4 | 1.125 |
| | В | 5 | - | 2 | 4 | 6 | 5 | 1.583 |
| Actor | С | 6 | 6 | - | 6 | 5 | 1 | -1.667 |
| | D | 4 | 6 | 2 | - | 4 | 3 | -1.042 |
| | Е | 6 | 6 | 2 | 3 | - | 4 | .917 |
| | F | 6 | 5 | 4 | 2 | 5 | - | 917 |
| Partn Effec | er ts | .125 | .417 | .167 | 708 | .083 | 083 | |

Table 1. Interpersonal perceptions in the group of six individuals.

These effects show the tendencies of the students to elicit interpersonal perceptions regarding the importance of their partners' contribution to group performance, as well as their tendencies to receive this assessment regarding their own contribution to the group performance.

The relationship effects can also be obtained by applying the *SRM.effects* function of the DyaDA package. The following table shows the dyadic adjustment for the interpersonal perceptions of contributions to group performance:

| | | | | Partner | | | |
|-------|---|--------|--------|---------|------|------|--------|
| | | А | В | С | D | Е | F |
| | А | - | 842 | .408 | 717 | .492 | .658 |
| | В | 008 | - | 050 | .825 | .033 | 800 |
| Actor | С | .242 | -1.050 | - | .075 | 717 | 1.450 |
| | D | -1.383 | .325 | 2.575 | - | 342 | -1.175 |
| | Е | .658 | .367 | 383 | 508 | - | 133 |
| | F | .492 | 1.200 | -2.550 | .325 | .533 | - |

 Table 2. Matrix of estimated relationship effects for the dyads in the group of six undergraduate students.

Once the various SRM effects have been estimated, the researcher might well wish to estimate variances and covariances for the interpersonal perception data in the group of six students. It can be done by means of the *SRM.variances* function. DyaDA package also includes *SRM.within.groups.tTest* function that allows researchers to estimate statistical significance for the different variances and covariances obtained by means of the SRM (table 3).

| | Estimated | Standard Error | t statistic | Two- tailed p value |
|-----------------------------|-----------|-------------------|-------------|---------------------------|
| Actor variance | 139 | .115 | -1.203 | .283 |
| Partner variance | 1.578 | .999 | 1.579 | .175 |
| Relationship variance | 1.406 | .416 | 3.380 | .020 |
| Actor-partner covariance | .261 | .254 | 1.028 | .351 |
| Dyadic covariance | 178 | .416 | 427 | .687 |

Table 3. Matrix of variances and covariances for the round
robin data of six undergraduate students. Statistical
significances for the variances and covariances are estimated
by means of a within-groups t test [11].

Note that the differences in the interpersonal perceptions of the partners' contributions to the group performance can only be explained by the relationship variance, i.e., by the differences in the individuals' adjustments within dyads regarding the evaluation of their partners' contributions.

Generalized and dyadic correlations can also be obtained using the nonindependence package. However, for this illustrative example they are expected to be non-significant, since actor-partner and dyadic covariances are nonsignificant at the 5% level.

CONCLUSIONS

DyaDA is a free, open-source statistical package for dyadic data analysis. It includes several statistical methods developed in the last years by psychologists, ethologists and other social scientists. Some of these statistical methods are not available in any statistical software. The package is open to incorporate more new statistical procedures than those originally included.

Social researchers can benefit from the powerful of DyaDA package when carrying out their dyadic data analyses and, additionally, they can also use the wide variety of packages and functions that already exists for statistical computation and graphics under the R environment.

Although using the DyaDA package by means of the R command prompt (i.e., Command Line Interface) is fast and highly flexible, it may be difficult to handle for novice users. For this reason we should also develop a friendly interface for using the DyaDA package (i.e., Graphical User Interface). The ultimate goal of this research is, therefore, to make this package more generally available to a wide audience of social scientists interested in dyadic analysis.

REFERENCES

- Alferes, V. R., & Kenny, D. A. SPSS programs for the measurement of nonindependence in standard dyadic designs. *Behavior Research Methods*, 41 (2009), 47-54.
- 2. de Vries, H. The rowwise correlation between two proximity matrices and the partial rowwise correlation. *Psychometrika*, 58 (1993), 53–69.
- 3. de Vries, H. An improved test of linearity in dominance hierarchies containing unknown or tied relationships. *Animal Behaviour*, *50* (1995), 1375–1389.
- 4. de Vries, H. Finding a dominance order most consistent with a linear hierarchy: a new procedure and review. Animal Behaviour, *55* (1998), 827–843.
- de Vries, H., Stevens, J. M. G., & Vervaecke, H. Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*, 71 (2006), 585–592.
- 6. Hemelrijk, C. K. Models of, and tests for, reciprocity, unidirectionality and other social interaction patterns at a group level. *Animal Behaviour*, *39* (1990a), 1013–1029.

- Hemelrijk, C.K. (1990b). Hemelrijk, C. K. A matrix partial correlation test used in investigations of reciprocity and other social interaction patterns at group level. *Journal of Theoretical Biology*, 143 (1990b), 405-420.
- Kenny, D. A. Interpersonal perception. A social relations analysis. New York, NY: Guilford Press (1994).
- 9. Kenny, D. A., Kashy, D. A., & Cook, W. L. *Dyadic data analysis*. New York, NY: Guilford Press.
- Kenny, D. A., & La Voie, L. J. (1984). The social relations model. In L. Berkovitz (ed.), *Advances in experimental psychology*. San Diego, CA: Academic Press (2006), 142-182.

- Lashley, R. B., & Bond, C. F. Jr. Significance testing for round robin data. *Psychological Methods*, 2 (1997), 278-291.
- 12. R Development Core Team *R: Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna: Austria (2009).
- 13. Sierra, V., Andrés, A., Solanas, A., & Leiva, D. (in press). Agreement in interpersonal perception as a predictor of group performance. *Psicothema*.
- Solanas, A., Leiva, D., Sierra, V., & Salafranca, Ll. Measuring and making decisions for social reciprocity. *Behavior Research Methods*, 41 (2009), 742-754.
- Warner, R. M, Kenny, D. A., & Stoto, M. A new roundrobin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, 37 (1979), 1742-1757.

The Algorithm for Detection of Fuzzy Behavioral Patterns

Valera V. Vishnevskiy valera.vishnevskiy@yandex.ru **Dmitry P. Vetrov** vetrovd@yandex.ru

Moscow State University, Vorobyevy Gory, 1, 119992 Moscow, Russia

ABSTRACT

In this paper we present a new algorithm for the detection of fuzzy patterns in discrete time series. It generalizes the known approach by M. Magnusson to T-patterns detection. In contrast to the latter, our algorithm is able to find patterns where some elements can be absent in some occurrences of pattern. This makes possible to find soft stereotype in data which seems to be more natural in behavioral analysis.

Author Keywords

T-Patterns, behavior, fuzzy patterns, elementary behavioral acts.

INTRODUCTION

The problem of stereotypes detection in the behavior of humans and animals is extremely important in cognitive research since it allows to measure the complexity of behavior in quantitative terms, to monitor behavioral changes, etc. Here we focus on one possible approach to measuring behavior which is based on pattern detection. The behavior is represented as a sequence of events from a finite set of event types (e.g. the beginnings of behavioral acts) which occur at some moments of time. One or more events can occur at one moment of time. A pattern is a chain of events which occur one after another quite often. Such pattern allows to detect the repeated fragments of behavior.

A popular approach to pattern detection was proposed in [1] by Magnusson where the notion of T-pattern was established. The main drawback of this approach is the fact that patterns are assumed to be crisp, i.e. if at least one elements from the chain of events is missing the pattern is absent.

This complicates the process of searching the patterns in noisy chaotic data. It seems natural to assume that behavior is more complex than just a chain of events and depends on many factors that cannot be observed directly. In mathematics the traditional way to deal with unknown factors which cause influence on the studied process is to remove deterministic model with probabilistic one thus allowing the dependencies to be fuzzy and to contain the element of randomness. In the paper we establish probabilistic approach to pattern detection.

PROPOSED METHOD

General idea of our approach is based on the algorithm, proposed in [1]. It is iterative method, which consists of following repeated actions:

- Test every two patterns from the pattern set, whether there is significant co-occurrence among them (second pattern often occurs after the first one). If so, then these two patterns are joined together and added to the pattern set.
- Remove all duplicates and incomplete patterns from the pattern set.

That process goes on, until no more patterns are found. At first iteration the pattern set is a set of pseudo patterns, i.e. patterns of length 1.

Determining Data Types

Similarly to Magnusson's model, behavioral data is coded during an observation period[1, N_t]. At every time moment one or more events (behavioral acts) *can* take place. The set of time moments, when event E_i had appeared, we will denote as *ind*(E_i). Formally speaking, we are searching for the temporal patterns in a discrete signal.



Figure 1. Example of pattern A[0;1]B[μ_B;σ_B]C[μ_C;σ_C].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

To describe a pattern of length *N*, we use *N* pairs of parameters, which describe correspondent events: expected shift from previous event occurrence and its standard deviation (μ and σ). For the first event $\mu = 0$ and $\sigma = 1$. We will denote the pattern *P*, that consists of events $E_1, E_2, ..., E_N$ in the following manner :

$$P = E_1[\mu_1, \sigma_1]E_2[\mu_2, \sigma_2] \dots E_N[\mu_N, \sigma_N].$$

For each pattern *P* of length *N*, for every time moment $\varepsilon \in [1, N_t]$, we compute the *likelihood function* L_P in the following way (see Figure 2):

$$L_{P}(\varepsilon) =$$

$$= \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi}\sigma_{i}}\right) f_{Loss}(N_{-,}N) \prod_{i=1}^{N_{+}} \exp\left(-\frac{\delta_{i}^{2}}{2\sigma_{i}^{2}}\right),$$

$$\delta_{i} = \min_{x \in ind(E_{i})} |\varepsilon + \sum_{j=1}^{i-1} [\mu_{j} + \delta_{j}] + \mu_{i} - x|,$$

$$f_{Loss}(x, N) = \begin{cases} \exp\left(-\frac{\lambda x}{N}\right), & x < N, \\ 0, & x = N, \end{cases}$$

$$N_{-} + N_{+} = N.$$

where N_+ is the number of events, that occurred in the pattern at current time moment, N_- is the number of events, that are missing in pattern at current time moment, δ_i is distance between expected and observed position of *i*-th event occurrence(see Figure 1). Event is treated as missed, if $\exp\left(-\frac{\delta_i^2}{2\sigma_i^2}\right) < \exp\left(\frac{\lambda}{N}\right)$, i.e. when δ_i is too big. Here we assume that the position of the first event is fixed at ε .

We can control the level of pattern fuzziness, by changing λ : decreasing it, would allow more event gaps in pattern (see Figure 3).



Figure 3. Loss function of pattern of length N = 8and different λ .

The value of likelihood function at time moment ε can be interpreted as level of confidence that given pattern starts at that time moment. Finding significant maximums of the likelihood function, we can define the moments, when pattern begins.

Note, that we can compute the likelihood function fixing any event we want. For example, the likelihood of the pattern *P*, counted w.r.t. *m*-th event, we will denote as

$$L_{P,m}(\varepsilon) = L_P(\varepsilon + \sum_{i=1}^m \mu_i).$$

Detecting Co-Occurrences

On that step we consider whether two patterns P_L and P_R should be merged to larger pattern. By computing the likelihood functions for P_L from the end and for P_R from the beginning, we find significant maximums of these likelihoods. Let $\{\alpha_i\}, \{\beta_i\}$ be the values of the likelihood



Figure 2. Example of likelihood for pattern A[0;1]B[84;3.5]C[54;2.0]D[120;3.2]E[82;2.0]F[16;2.3]

maximums and $\{x_{L,i}\}, \{x_{R,j}\}$ indexes of these maximums for P_L and P_R respectively. We consider the distances between the occurrences of each pattern which are less than some predefined threshold *M*:

$$\rho = \{ x_{R,j} - x_{L,i} | 0 \le x_{R,j} - x_{L,i} \le M \},\$$

and used weights to increase the influence of those pattern occurrences that better correspond to the statistical model of pattern:

$$w_l = \log\left(1 + a_i\beta_i\right).$$

Next, we consider the following sum (see Figure 4):

$$k = \sum_{l=1}^{Q} w_l g_{\mu,\sigma}(\rho_l),$$
$$g_{\mu,\sigma}(\rho_l), = exp\left(-\frac{(\rho_l - \mu)^2}{2\sigma^2}\right),$$
$$S(\sigma) = \sigma\sqrt{2\pi},$$
$$Q = |\rho|.$$

Here $g_{\mu,\sigma}(\rho_l)$ is statistical model of co-occurrence. We are trying it on with different μ and σ , testing if there is a significant co-occurrence. Note, that because of computational complexity and the assumption that the distance between events in pattern should be small, only the co-occurrences that are shorter than *M* are considered.



Figure 4. The distribution of distances between patterns and $g_i(\mu, \sigma)$ which maximizes the expression (1).

To test the significance of co-occurrences establish nullhypothesis that the two patterns are independent. Let

$$Y = \sum_{l=1}^{Q} X_l,$$
$$X_l = w_l g_{\mu,\sigma}(\rho_l) = w_l exp\left(-\frac{(\rho_l - \mu)^2}{2\sigma^2}\right).$$

Under null-hypothesis:

- w and ρ_l are independent random variables,
- ρ_l is uniformly distributed. $\rho_l \sim U[0, M]$.

Then it can be shown that

$$Y \sim \mathcal{N}\left(\frac{\sum_{i=1}^{Q} w_{i}}{M}S, \frac{1}{M^{2}}\left[MS\sqrt{2}\sum_{i=1}^{Q} w_{i}^{2} - \frac{\left(\sum_{i=1}^{Q} w_{i}\right)^{2}}{Q}S^{2}\right]\right).$$

In order to perform test, we maximize the following value, using methods from [4]:

$$\frac{k - \mathbb{E}Y}{\sqrt{\mathbb{D}Y}} \to \max_{\mu, \sigma} \tag{1}$$

If maximum value is greater, then the quantile of normal distribution with predefined significance level $\omega \in [0, 1]$, and the two patterns P_L and P_R co-occur sufficiently frequently, then we make decision that the statistically significant co-occurrence $P_L[\mu;\sigma]P_R$ takes place. And therefore, constructed pattern is added to pattern set. Parameter ω stands for significance of found patterns: the closer ω to 1, the more significant patterns are found.

While speaking "co-occur sufficiently frequently" we mean, that the sum of significant maximums of the likelihood function $L_{P_L[\mu;\sigma]P_R}(\varepsilon)$ is greater than η .

Removing Patterns

Similarly to Magnusson's approach, our method can construct duplicate and incomplete patterns. That is why we need some mechanism to eliminate those patterns on each step.

Duplicated Patterns

The problem is that, one pattern can be constructed from different subpatterns. For example, pattern ABCD can be detected both by uniting (AB) and (CD), or (A) and (BCD). Generally they result to the same patterns, but due to complicated process of uniting, they could have slightly different likelihood functions.

Incomplete Patterns

While constructing patterns from subpatterns, it's possible, that subpattern only appears as a part of constructed pattern. Therefore we don't need to consider such subpatterns independently. For example, if in pattern ABCD, AB just doesn't occur out of ABCD, then we don't need pattern AB in pattern set, and likelihood functions of ABCD and AB should be very similar.

Considering described above examples, simple procedure of pattern elimination was proposed. First, let's define the following values:

$$\overrightarrow{L_{P,i}} = \left(L_{P,i}(1), \dots, L_{P,i}(N_t)\right) - \text{vector } 1 \times N_t,$$

where $L_{P,i}(\varepsilon)$ — is the likelihood function of pattern P, at time moment ε , computed with respect to the *i*-th event. N_t is the length of the time period of the observation.

$$cor(\overrightarrow{L_1},\overrightarrow{L_2}) = \frac{\overrightarrow{L_1}\overrightarrow{L_2}^T}{\sqrt{\overrightarrow{L_1}\overrightarrow{L_1}^T}\sqrt{\overrightarrow{L_2}\overrightarrow{L_2}^T}} \ \epsilon[0,1],$$

is the correlation coefficient between likelihoods, therefore, the closer it is to 1, the more similar the likelihoods are.

Procedure of Elimination

We test every pair of non-pseudo patterns(P_L and P_R), from the pattern set, considering P_L as a duplicate or incomplete copy of P_R . Note that we don't consider pseudo
patterns(single event types), because they may be necessary for constructing new patterns.

The first test deals with duplicate patterns: if P_L and P_R consist of the same events, and $cor(\overline{L_{P_L,1}}, \overline{L_{P_R,m}}) > \nu$ (*m* is index of the first event of P_L in P_R), and $\|\overline{L_{P_L,1}}\| \leq \|\overline{L_{P_R,m}}\|$, then we remove P_L from the pattern set. The exclusion of incomplete copies is done in a similar manner.

In this section we introduced new parameter ν . In our experiments, $\nu = 0.7$ usually worked well.

Structural Parameters

The algorithm we derived in the paper has some parameters (see Table 1), that should be set manually. However, during experiments we discovered that the default values are often working well, or alternatively they can be set, according to prior information about considered behavioral time series, which would improve the performance leading to more interpretable patterns



Figure 5. Comparison of the longest detected pattern (in the actual data), using T-Patterns (above) and fuzzy patterns (below). Gaps are illustrated as circles. Note that fuzzy patterns are longer and are observed more often than their crisp variants. That happens because of insufficient number of pattern occurrences without gaps.

| Parameter | Possible values | Default value | Has influence on |
|-----------|--------------------|------------------|--|
| ω | [0, 1] | 0.95 | Significance of patterns |
| ν | [0,1] | 0.7 | How much similar patterns should be to be eliminated |
| М | $[0, N_t]$ | None | Length of relations that connect patterns. |
| η | [0,+∞] | 3 | Minimal pattern occurrences |
| λ | [0, +∞] | 6 | Fuzziness of patterns |

Table 1. Main algorithm parameters.

EXPERIMENTS AND COMPARISION WITH ANOTHER METHOD

To test implemented algorithm on real data, we used hamster behavioral data from open field test and recordings of grooming. Also we compared proposed algorithm with Magnusson's T-Pattern approach [1].

Each dataset was presented by set of pairs: event type (behavioral act) and time moment at which that event had started. Both grooming and open field data had, on average, 15-30 event types and each event type occurred 20-80 times. Every following event occurrence defined the end of previous event. Figure 5 contains an example of discovered fuzzy pattern and the closest T-pattern generated by Magnusson's algorithm on open field test data.

In general, the set of patterns found by our method contained¹ almost all patterns that were discovered using T-Patterns technique. At the same time, it didn't contain too much noisy patterns, which meant, that Fuzzy Patterns extended T-Patterns framework in a reasonable way. The typical example of the difference between two methods is shown in Figure 6. Moreover, fuzzy patterns that corresponded to some T-Pattern had greater likelihoods and fuzzy analogues of longest T-Patterns were always detected by our method.

In some cases we observed the situation when there were many fuzzy patterns that were fuzzy variations of the same T-Pattern. This effect could be eliminated by fine parameter tuning. Also the longest fuzzy pattern is the extension of the discovered T-Pattern, which seems reasonable.



Figure 6. Number of patterns found using different methods on several real datasets. On the average, Fuzzy patterns method finds 93.6% of T-Patterns.

¹ In general we cannot say definitely that some fuzzy pattern corresponds to the specified T-pattern, because of different pattern representations.



Figure 7. Histogram of pattern lengths.

CONCLUSION AND FUTURE WORK

Our proposed method for behavioral time patterns discovery, based on fuzzy pattern detection has shown promising results. It worked well on synthetic data (both when the distances between elements of pattern were generated from Gaussian and uniform distributions), and on actual data, detecting only those patterns, that were present in time series. The experiments show, that our algorithm is able to detect longer significant patterns in time series, than the algorithm based on T-Pattern detection. Figure 7 shows the distribution of the length of patterns found in grooming behavioral data. Due to method's statistical roots, some patterns can be treated as noise. Also our method is computationally complex. The current version, implemented on MATLAB works approximately 100 times longer, than the algorithm, based on T-Pattern detection. One of the directions for future work is parallel implementation of the algorithm on multiprocessor computers or on Graphical Processing Units (GPUs).

We would like to thank Irina Zarayskaya for providing us behavioral data. The work was partly supported by Russian foundation for basic research (grant #08-01-00405), Russian President grant (MK3827.2010.9) and by Federal Target Program "Scientific and scientific-pedagogical personnel of innovative Russia in 2009-2013" (contact no. P1265).

- Magnusson, M.S. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers 23* (2000), 93-110.
- 2. Anolli, L., Duncan, S., Magnusson, M.S. Understanding Social Interaction: Discovering Hidden Structure with Model and Algorithms. In *The Hidden Structure of Interaction: From Neurons to Culture Patterns*. IOS Press (2005), 2-22.
- Anolli, L. The Detection of the Hidden Design of Meaning. In *The Hidden Structure of Interaction: From Neurons to Culture Patterns*. IOS Press (2005), 23-50.
- 4. Boyd, S., Vandenberghe, L. *Convex Optimization*. Cambridge University Press (2004), 466-475.

Machine Learning Techniques for Parameter Selection and Automated Behavioral Classification

Dimitris Fekas

Noldus Information Technology Nieuwe Kanaal 5, 6709 PA Wageningen, The Netherlands dimitris@noldus.nl Raymond. C. de Heer raymond.deheer@ deltaphenomics.com

Marco Mellace

marco.mellace@ deltaphenomics.com

Delta Phenomics BV PO Box 80086, 3508 TB Utrecht, The Netherlands

Berry M. Spruijt

Delta Phenomics BV, P.O. Box 80086, 3508 TB Utrecht, The Netherlands berry.spruijt@deltaphenomics.com Cajo J.F. ter Braak Biometris, Wageningen University and Research Centre, Box 100, 6700 AC Wageningen, The Netherlands cajo.terbraak@wur.nl

ABSTRACT

In this paper, we address two different problems: efficient parameter selection for complex data sets and automated behavioral classification. Different strains of rodents were monitored in the PhenoTyper and various observable features have been quantified. We applied several machine learning techniques in order to explore the structure of the data and we managed to identify the most important behavioral parameters. In addition, we compared different classification techniques with respect to their accuracy and robustness in determining behavioral differences between the strains.

Author Keywords

Variable selection, behavioral classification, support vector machines, random forests, multidimensional scaling, multivariate analysis.

INTRODUCTION

In the area of behavioral research, there is a great need for standardized, automated experimental set-ups which allow for repeatable experiments and more reliable results. In order to tackle these issues, Noldus Information Technology has developed both hardware (PhenoTyper) and software (EthoVision XT) that allow automated video tracking of animals. Such an approach greatly improves our ability to quantify observable behaviors. However, the problem is deriving meaningful interpretations of the vast amount of data acquired by the tracking system. In this research, we compare the behavior of two different strains of mice with respect to various locomotive parameters.

MATERIALS

The PhenoTyper is a specially designed cage which allows automated observation of mice or rats. On top of each cage there is a unit containing the hardware needed for video tracking. In addition, other hardware devices are available, such as computer-controlled lights, a sensor to detect when the animal is drinking and a pellet dispenser. EthoVision XT is the video tracking software which detects the animal in the PhenoTyper. It also provides the option to the user to determine zones inside the cage according to the experiment and a Trial and Hardware Control module to control external devices such as lights, sounds and the pellet dispenser. EthoVision XT continuously samples the coordinates of the center of gravity of the animal from which the trajectory and velocity of the animal can be calculated. By using Trial and Hardware Control, it is possible to automate cognitive tasks and further discover behavioral characteristics such as anxiety, learning capability and memory. These tools allow for continuous recordings without much human interventions, and are used by behavioral researchers worldwide [6,9,12,13].

The animals used in this experiment were obtained by Harlan laboratories (males, 8-weeks-old): They were of 2 different strains: C57BL/6JOlaHsd and DBA/2JOlaHsd. The experiment was conducted at the research facilities of Delta Phenomics and it lasted for a period of one week. It

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

was initially designed to validate automated behavioral tests, which are not directly within the scope of the presented research, because our main focus is on locomotive behavior.

METHODS

For our analysis, a choice needed to be made for selecting a biologically meaningful set of parameters, as input for the classification algorithms. The raw EthoVision XT data consists of almost continuous measurements at a rate of 12.5 samples per second. Missing samples were interpolated and the entire set was smoothed using the LOWESS algorithm [5]. From the smoothed data, we can directly calculate some parameters for each animal, such as: distance moved, velocity, measures of the animal's elongation, mobility and angular velocity, duration and frequency of stops, as well as of visits to certain predefined zones within the PhenoTyper. During the experiment, a light/dark phase of 12/12 hours was used. Therefore, the above-mentioned parameters were either averaged or summed up for that time window. For the "continuous" parameters that were averaged, apart from the mean, the coefficient of variation was also used to provide some more information about their distribution in that rather long time bin.

For the final data set used, we collected these summary statistics for each of the parameters, each of the time bins and each of the animals and thus, we created a summary file for the entire experiment. After some standard preprocessing, such as removal of highly correlated parameters, logarithmic transformations (where appropriate) and scaling, we explored how the data was organized and we visualized that using methods like principal components analysis [7] and (non-metric) multidimensional scaling [3].

The main goal of this research was to discover which parameters mostly contribute to the between-strain differences in behavior. For variable selection, we used two methods: logistic regression [15] and support vector machine [10]. The results of both methods will be compared and behavioral implications of these parameters will be discussed. The second goal of this paper is to build a classifier, which can assign each animal to its respective strain, using behavioral parameters. With respect to this, our research continues as follows. Some of the animals are treated as if we did not know to which group/strain they belong to. The rest of the data is used as a training set and different algorithms that try to "learn" the machine how do animals from each strain behave. The classifier's performance can be measured by checking whether the animals that have been excluded from the training set are assigned correctly to their respective classes. The classification algorithms that were used were: random forests [2], support vector machines [11,14], and partially squared discriminant analysis [1]. The performances of the different classifiers were compared with respect to their accuracy and robustness using cross-validation.

CONCLUSIONS AND FUTURE WORK

The methods techniques mentioned above, are nowadays applied in numerous diverse scientific fields, but their use within behavioral science is to our knowledge quite limited [8] and hence there is much scope for novel developments. This is an on-going research and even though the importance of many behavioral parameters has been discovered, it may well be the case that there is more information in the data. For instance, some particular time bins may be more important than others and valuable information is lost when using a 12 hour time bin. Also, possible day-to-day differences could perhaps be explained by the cognitive tests that take place. This has not yet been taken into account, but it will be done in the near future.

ETHICAL STATEMENT

The aforementioned experiment was conducted after getting approved by the DEC of the University of Utrecht (DEC number: 05111201).

- Barker, M. and Rayens, W., Partial least squares for discrimination. *Journal of Chemometrics*, 17 (2003), 166-173.
- 2. Breiman, L., Random forests. *Machine learning*, 45 (2001), 5-32.
- 3. Borg, I. and Groenen, P. *Modern Multidimensional Scaling: Theory and Applications.* Springer, 2005.
- 4. Cristianini, N. and Shawe-Taylor, J. An introduction to support vector machines: and other kernel-based learning methods. Cambridge University Press, 1 edition, March 2000.
- 5. Hen, I., Sakov, A., Kafkafi N., Golani, I., Benjamini, Y. The dynamics of spatial behavior: how can robust smoothing techniques help? Journal of *Neuroscience Methods*, 133 (2004), 161-172.
- Hodges, A., Hughes G., Brooks, S., Elliston, L., Holmans, P., Dunnett, S. and Jones, L. Brain gene expression correlates with changes in behavior in the r6/1 mouse model of Huntington's disease. *Genes*, *Brain and Behavior* 7 (2008), 288-299.
- 7. Jollie, I.T. *Principal Component Analysis*. Springer, second edition, October 2002.
- 8. Martiskainen, P., Jarvinen, M., Skon, J., Tiirikainen, J., Kolehmainen, M. and Mononen J. Cow behaviour pattern recognition using a three- dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science*, *119* (2009): 32-38.
- 9. Noldus, L., Grieco, F., Loijens, L., and Zimmerman, P. (Eds.). *Measuring behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research, Wageningen, The Netherlands, 2005.*

- Rakotomamonjy, A. Variable selection using svm based criteria. *The Journal of Machine Learning Research*, 3 (2003), 1357–1370.
- 11. Scholkopf, B. and Smola, A. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, 2002.
- 12. Spruijt, B.M. and De Visser, L., Advanced behavioural screening: automated home cage ethology. *Drug Discovery Today: Technologies*, *3* (2006), 231-237.
- Van Vliet, S., Vanwersch R., Jongsma M., Van der Gugten, J., Olivier, B. and Philippens, I. Neuroprotective effects of modanil in a marmoset

Parkinson model: behavioral and neurochemical aspects. Behavioural Pharmacology, 17(5-6):453{462, 2006. Professional ethics. Social Science Computing Review *10*, 2 (1992), 453-469.

- 14. Vapnik ,V. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- 15. Zou, H. and Hastie, T., Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 67 (2005), 301.

Using Ideas of Kolmogorov Complexity for Studying Animal Behavioural Patterns

Sofia Panteleeva psofia@mail.ru

Zhargal Danzanov danzanov.zhargal@gmail.com Zhanna Reznikova zhanna@reznikova.net

Institute of Systematics and Ecology of Animals Frunze, 11, Novosibirsk, Russia, 630091

ABSTRACT

We suggest a method for evaluating the complexity of animal behavioural patterns based on the notion of Kolmogorov complexity, with ants' hunting behaviour as an example. We compared complete (successful) and incomplete hunting stereotypes in members of a natural ant colony and in naïve laboratory-reared ants. We represent behavioural sequences as "texts", and compress them using a data compressor. Behavioural units (10 in total), singled out from video records and denoted by letters, served as an alphabet. Successful hunting stereotypes appeared to be characterized by smaller complexity than incomplete ones. A few naïve "born hunters" which enjoy "at once and entirely" complete hunting stereotypes are characterized by a lower level of complexity of hunting behaviour. We conclude that innate complete stereotypes have less redundancy and are more predictable, and thus less complex.

Author Keywords

Kolmogorov complexity, data compressor, redundancy, behavioural sequence, ants, hunting.

INTRODUCTION

The concept of complexity of animal behaviour is still mainly intuitive. First of all, one has to distinguish between the complexity of flexible and stereotypic behaviour. In the first case we mean levels of complexity of problems to solve and decisions to make, whereas in the second case we mean the inner coordination and regularity of speciesspecific repertoire [5]. Surprisingly, despite many attempts to examine the organizational complexity of signal repertoires [1], there are no reliable tools for studying the complexity of animal behavioural patterns. Ants can serve a good example here because these insects exhibit diverse behavioural patterns, both flexible and instinctive.

In this study we apply the ideas of Kolmogorov complexity to compare species-specific hunting stereotypes in experienced and naïve ants. Informally, Kolmogorov complexity is the length of a minimal program generating this sequence [a formal exposition in: 2]. In our case the alphabet is a whole set of species-specific behavioural units, and a sequence of letters corresponds to a displayed behaviour. Intuitively Kolmogorov complexity of a text is the length of this text compressed by an "ideal" data compressor; an ideal data compressor is a compressor that finds every possible regularity in the data. As ideal compressors do not exist, we use real ones. Real data compressors can find many regularities that commonly occur in textual data. The concept of Kolmogorov complexity is widely applied in different fields of science [3]. However, although many authors have tried to apply ideas of information theory to animal communication [4], an experimental study of ant "language", that is, ants' ability to "compress" messages based on the "alphabet" consisting of two letters R (right) and L (left), still remains the sole application of Kolmogorov complexity to animal behaviour [7].

METHODOLOGICAL APPROAHES AND MATERIALS

Let us define the notions that we use to describe behaviour. We selected elementary movements and postures as minimal units of behaviour ("behavioural elements", for brevity). We call a "behavioural sequence" an arbitrary sequence of successive behavioural elements, and a "behavioural stereotype" - a relatively stable chain of behavioural elements. The display of behavioural stereotypes in the context of species-specific behaviour is somewhat stochastic: some elements can be duplicated or dropped out; they also can be "diluted" with accessory behavioural elements ("noise"). Fixed action pattern (FAP) is defined as an instinctive behavioural sequence that is invariant and indivisible [8]. It is possible that a behavioural sequence includes a behavioural stereotype, and the latter includes a FAP.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

| N⁰ | Complete (successful) stereotypes | Incomplete stereotypes |
|----|---|----------------------------------|
| 1 | RACACKCRUACKCKCC CCCP. | WARURW. RTARW. |
| | RARUKACKCKCKCP. | RARTRRW. |
| 2 | RUACCKKCWTTWCUCP RUACRUTURACRURCP. RURBRARRACKCKCCP. | RTRACTTRAUT RAC. |
| 3 | - | WBBAT. |
| 4 | RARWTTRACCKCKCCK KCKCTP. RURAAACTCKCKCCKC KCCCP. RACKCKCCCCP. | RARRURAARW. WRSA. RTRUA. |
| 5 | - | RART. |
| 6 | RAUURRTRACP. | RARAK. RUWRAW. RARRUAR. |
| 7 | RACKSSWSSSCCCCCCP. RACKCKCP. RACCCP. RURARCKCKCCKCCCCP | RARAK. WCKCKR. |
| 9 | RTRACCKCP. RARACP. | RWTAAU. |
| 11 | RARARRACKKCCP. RTACKCCKCP. | RTARATW. SWATAW. RARUARAK. |
| 13 | - | RTAAUAU. RUARAW. |
| 14 | RACWKSSKCCP. RAACCKKKCCP. | - |

 Table 1. Behavioural sequences denoted by letters in 14 'wild' ants.

Legend: W – waiting/ stopping; S - slow walk; R – running; T – turning ; U – turning around; B - belligerent posture (an ant stands on four legs raising two legs highly over its head); A – attack (falling on a victim); C – capturing: S – stinging; T – transporting a prey; point means the end of a stereotype; dash – lack of a stereotype. The №№ 8, 10, 12 wild ants did not hunt.

We analysed a hunting stereotype of *Myrmica rubra* ants towards jumping springtails. This stereotype includes determining the victim, approaching it, and then performing the FAP that we call "tip-and-run attack": the ant attacks the prey, bend the abdomen and head to the thorax, jumps to the springtail, falls on it abruptly, and stings. Then the ant intercepts the victim and transports it to the nest. In laboratory we compared two groups of highly genetically variable ants: members of a natural colony ('wild' for brevity) and naïve (laboratory reared) ants of age from 3 to 12 days. It is worth to note that not all naïve ants were tested. It was shown in [6] that within ant colonies about 5% of all members are "born hunters", and they display the hunting stereotype "once and entirely". In this study we intentionally chose the most active ants, and from 12 individuals there were 6 that demonstrated the completed hunting stereotype ended with killing the prey (see Table 1 and 2). All ants were placed one by one into glass containers with 30 live springtails, and each individual was tested once. To analyze ethograms from video records, we used the Observer XT 7.0 (version: 7.0.214, Noldus Information Technology). Using an "alphabet" of 10 behavioral units, we expressed the hunting stereotypes as text files. Every sequence (file) was constructed manually (by the researchers) from the corresponding video fragment. As the starting point of a hunting stereotype we considered the ant's approach to the victim and the display of purposive movements; transportation of a killed victim was considered the end of the complete stereotype. All cases of loss of a victim and switching to another one were considered ends of incomplete stereotypes. We reduced files to equal initial length, compressed them with the use of KGB Archiver (v.1.2) and compared ratios of compression in different stereotypes. The length of a compressed "text" can be thought of as an estimate of its Kolmogorov complexity.

RESULTS AND DISCUSSION

We obtained 4 files which included: 19 complete and 20 incomplete hunting stereotypes in "wild" ants and, correspondingly, 20 and 31 stereotypes in "naïve" ants (see Table 1 and 2). The successful (complete) hunting stereotype in "wild" ants appeared to be characterized by smaller complexity than the incomplete stereotype: the compression ratio is 63.27% for the first file, and 70.07% for the second one. It is likely caused by a greater frequency of key elements and less "noise" in complete stereotypes. The complete stereotypes also have less redundancy and are more predictable. The same was found in naïve ants: 56. 4% and 68.03%, correspondingly. Of particular interest is that naïve "born hunters" which enjoy "at once and entirely" hunting stereotypes are characterized by a lower level of complexity of hunting behaviour. Their hunting stereotypes are possibly most "laconic" and clear. This supports the hypothesis of Reznikova and Panteleeva [6] that a few carriers of the whole hunting stereotype to be spread within an ant colony, can serve as catalysts of social learning for other individuals which possess only fragments of relevant behavioural patterns.

Evaluation of the complexity of behavioural stereotypes has already enabled us to reveal discrete behavioural variability within an ant colony. In perspective, the use of ideas of

| № | Complete (successful) stereotypes | Incomplete stereotypes |
|----|---|---|
| 1 | - | STAW. STAR. RWARU. WARW. |
| 2 | - | WCCCCW. |
| 4 | RTRACKUCCP. WTACKCCCCP. RACKCP. | RURAU. RARRTAR. RAUR. RAUW. |
| 5 | RTACKUP. RACTTTP. RTACKRTRRCCP. RTRACKCP. | RATWRAT. RTRAT. RTRAT. WTAK. |
| 6 | RACKCCKCP. RACKCKCP. RACKCCKCCCP. RACKCP. RACKCCP. | WTARTUR. WAKW. RTACKCKCR. |
| 7 | RUACKCP. RACKCCCCCCCCCCC CCP. RTATACCUKCCCCCP. RACKP. | RTACKRU. RAUR. RACRU. RUACKCKCR. RACKKTCCCCCCC CCCR. RACKCCKCCR. RACKRARACKUR. |
| 9 | RACKCP. | - |
| 12 | SACCCCP. STACCKCP. STAACKCP. | WAW. STAR. STATS. WUAUTS. SARURW. SAKS. SATUTS. SASU. |

| Table 2. Behavioural sequences denoted by letters in 12 | |
|--|--|
| naïve ants. The №№ 3, 8, 10, 11 naïve ants did not hunt. | |

Kolmogorov complexity for studying animal behaviour is a promising tool to be used in different areas of behavioural and evolutionary research. In particular, this method can help to extract "basic" (completely innate) behavioural patterns by comparing behavioural sequences of different levels of complexity and flexibility. This is particularly important for evolutionary and ethological studies in the field.

- 1. DeVoogd T. J. Neural Constraints on the Complexity of Avian Song. *Brain, Behaviour and Evolution*, *63* (2004), 221–232.
- 2. Kolmogorov, A.N. Three approaches to the quantitative definition of Information. *Problems in Information Transmission 1* (1965), 1-7.
- 3. Li M., Vytanyi P. An Introduction to Kolmogorov Complexity and Its Aplications. Springer Verlag. 2008.
- 4. McCowan B., Hanser S.F., Doyle L.R. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal behaviour* 57, 1999, 409-419.
- 5. Reznikova, Zh. Animal Intelligence: From Individual to Social Cognition. Cambridge University Press: Cambridge, MA, USA, 2007.
- Reznikova Zh., Panteleeva S. An ant's eye view of culture: propagation of new traditions through triggering dormant behavioural patterns. *Acta Ethologica 11*, 2 (2008), 73-80.
- Ryabko B., Reznikova Zh. Using Shannon Entropy and Kolmogorov Complexity To Study the Communicative System and Cognitive Capacities in Ants. *Complexity*, 2 (1996), 37-42.
- 8. Tinbergen N. *The Study of Instinct*. London and New York: Oxford University Press. 1951.

Ball Recovery in the Handball Tournament of the 2008 Beijing Olympic Games: Sequential Analysis of Positional Play as Used by the Spanish Team's Defence

António Lopes

Faculdade de Educação Física e Desporto, Universidade Lusófona, Portugal antonio.m.lopes@gmail.com

M^a Teresa Anguera

Department of Behavioural Sciences Methodology, University of Barcelona, Barcelona, Spain mtanguera@gmail.com

ABSTRACT

In sport there is a great need to obtain as much information as possible about the factors which affect the dynamics of play. This study uses sequential analysis and temporal patterns (T-patterns) to examine the evolution of defence (against an equal number of attackers) as used by the Spanish handball team at the Beijing 2008 Olympic Games. The aim is to help handball coaches (during their training and gathering of professional experience) to understand the importance of the structure of defensive systems. This can be achieved through observational processes that reveal the evolution and adaptation of these defensive systems according to different variables: the match score, the response of the opposing team and progress through the tournament.

Author Keywords

Team sport observation, T-patterns, observation, handball

INTRODUCTION

Recently there has been an increase in the number of studies applying observational methodology [1] to handball, both in theoretical terms (the conceptual development of models) and in the applied sense (recording and analysis of data derived from tactical and technical aspects). The form of positional play most widely used by teams is that based on attack, which raises the question as to whether this is the

Oleguer Camerino

Human Motricity laboratory, INEFC-University of Lleida, Spain ocamerino@inefc.es

Gudberg K. Jonsson

Human Behaviour Laboratory, University of Iceland, Reykjavik, Iceland gjonsson@hi.is

most characteristic method of handball play as regards ball recovery. At all events, studies that regard the attacking process as a key moment in team preparation are more common than are those which stress the importance of defence [2]. However, the choice of defence as a form of positional play is justified. Indeed, some authors have considered it to be the most important aspect and it has been widely used in handball as a phase of increased player activity, this being something which characterises and distinguishes handball play from other strategies [3].

The present paper aims mainly to describe and characterise the defensive behaviour of one handball team with respect to the attacking strategy of another (in the context of equal numbers of defenders and attackers). This is done by analysing the games played by the Spanish handball team during the 2008 Beijing Olympic Games. Traditional methods for quantifying performance in sport are limited in their ability to describe the complexity which emerges during the game. Due to the multiple dimensions and unpredictability that characterise play in handball, there is a need for studies that consider the sequential interaction between variables pertaining to both defensive and attacking play, and which do so from a multidimensional perspective and with a methodology that is consistent with the process of player opposition/cooperation.

METHOD

An appropriate research strategy in this regard is observational methodology. This approach enables the observation of individual behaviours in the real world by means of an *ad hoc* observation instrument that combines field formats and category systems. Furthermore, this instrument offers the quality and flexibility required to record the constant and random changes in player behaviour, enabling the detection of different relationships.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Participants

The match material was obtained by recording the games broadcast on television. Spain was the team observed in each case.

Design

The observational design [4] was ideographic (one team as the unit), longitudinal (several matches played by the same team at different stages) and multidimensional (the dimensions correspond to the criteria established for the observation instrument). The use of this design determines a set of decisions with respect to participation, tools (adaptation and validation of a new instrument) and procedures.

Observation Instrument

The observation instrument chosen for this study was the (SODMO), which combines field formats with category systems. This observation instrument was developed on the basis of two sources of information: first, previously published studies on defensive play, and second, the inductive construction of categories by experts (high-performance coaches specialising in handball). After constructing the observation instrument (SODMO) a team of observers was then trained in how to use it. The instrument was accepted as reliable when the value of Cohen's kappa was greater than 0.8.

Recording Instrument

The recording instrument used was the *Match Vision Studio* software [5] (see Figure 1). This is a highly-flexible software package into which the user first introduces all the codes corresponding to each one of the changing criteria of the SODMO observation instrument. All the co-occurrences of codes are then recorded, each of which occurs in a frame (the time unit used), and this produces a recording formed by the successive co-occurrences; the duration in frames of each one is recorded.

| Stage | Teams involved | 1st half | 2nd half | Final overall | score / result |
|-----------------------|-------------------|----------|----------|------------------|-------------------|
| Preliminary round | Croatia- Spain | 16-11 | 15-18 | 31-29 | Lost |
| | China- Spain | 12-16 | 10-20 | 22-36 | Won |
| | France- Spain | 16-10 | 12-11 | 28-21 | Lost |
| | Brazil- Spain | 20-17 | 16-18 | 35-36 | Won |
| Quarter- finals | Korea- Spain | 13-14 | 11-15 | 24-29 | Won |
| Semi-finals | Iceland- Spain | 17-15 | 19-15 | 36-30 | Lost |
| 3rd place play-off | Croatia- Spain | 14-12 | 15-23 | 29-35 | Won |

Table 1. List of games observed.



Figure 1. Screen capture from the Match Vision Studio software.

Procedure

In order to study the defensive behaviour of the Spanish handball team in the Beijing Olympic Games we conducted a sequential analysis so as to obtain temporal patterns (T-patterns) [6,7,8] for each game and game half. This was done using the Theme software, which produces dendograms based on event configurations that correspond to the behaviours recorded, showing the temporal distance between occurrences [9].

The total number of configurations/game-events recorded using the Match Vision Studio software (for all seven matches and based on the changing criteria of the SODMO) was 3511. These files were then transformed into the format required by the Theme analytic software, i.e. a table with temporal information for the successive co-occurrences of codes recorded. The codes of the SODMO observation instrument were then imported into the Theme software [6,7,8].

The data were then imported into the Theme software with the aim of detecting temporal patterns. In this context, Tpatterns can help to reveal hidden structures and nonobservable aspects of sporting techniques. Here, these Tpatterns were obtained using the algorithm that is incorporated within the software package Theme v.5 [7].

RESULTS

The result of all seven matches was analysed by grouping them as follows: each game individually and then grouped according to games won and games lost. Due to the large amount of data generated by the seven matches observed, we only present the results of matches won and lost in the group stage of the Olympic tournament. Therefore, the analysis concerns four games distributed in two groups: games won (Brazil-Spain and China-Spain) and games lost (Croatia-Spain- and France-Spain) (see Table 1).



Sequential Analysis

The plotter graphics were analysed by describing the events during the match and observing the sequences and ball recovery methods according to the match score at a given point (see Figures 2 and 3).

Into the results obtained related to the games won we can see that the game between China and Spain was the third match in the handball tournament, and ended with the score at 22-36. The game between Brazil and Spain was the fifth and final game in the group stage (preliminary round) and ended 35-36. Over these two games Spain scored 72 goals and conceded 57.

Seven different types of sequences (shown in the figure by colour bars) can be identified from among the 97 that make up the game as a whole. They have the following event frequencies and intervals according to the match score for the Spanish team at a given point in time. This enables us to observe the different dynamics of defensive sequences at the end of each match.

Into the results obtained related to the two parties lost in the preliminary phase, we can see that the first match lost by the Spanish team was that against Croatia, which ended with the score at 31-29. The match between France and Spain was the fourth game in the group stage (preliminary round) and ended 28-21 (see Table 1). Over these two games Spain scored 50 goals and conceded 59.

The analysis of the defensive systems used in these two defeats reveals seven different types of sequence (shown by the colour bars in Figure 3) from among the 97 that make up the game as a whole. They have the following event frequencies and intervals according to the match score for the Spanish team at a given point in time. Here we can see greater variability among systems of play, which shift between 5x1 and 6x6 in an attempt to overcome the rival.



Figure 3. Event time plotter of games won in the group stage.

After analysing the transition between defensive configurations using the plotter graphics we then selected the most significant T-patterns by means of the dendograms produced by the Theme software. All criteria were used to analyse the interaction between defence and attack.

By way of an example we show here an interpretation of dendograms for the match between Croatia and Spain (see Figure 4). The Spanish team were behind for most of the game and resorted to a 6:0 (zonal) defensive system, which was modified mid-way through the second half to 3:2:1 (nine), when the match scoreboard showed a three-goal disadvantage (p3). They then switched to 5+1 (six) when they were only one goal behind (p1) and level (e), returning then to the initial 6:0 (zonal) system.



Figure 4. Dendogram of the match between Croatia and Spain.

CONCLUSION

It was possible to verify the existence of T-patterns in which a defensive system was more associated with the type of ball recovery (conceding or without conceding a goal). There were more event configurations in dendograms when the Spanish team used the 6:0 defensive system. The change of defensive system varied depending on both the score at any given time and the 'quality' of the opponent. Defensive systems with more defensive lines were used in unbalanced matches in which the observed team was losing, as well as against better opponents.

The application of this software has proved to be extremely effective for studying different aspects of the dynamics of play in various contexts of interaction between opposing teams. As stated above, the present study focuses on the dynamics of defensive play in the handball tournament of the 2008 Beijing Olympic Games.

- Gomes, F., & Volossovitch, A. Análise dos processos defensivos da França em função do sucesso - estudo realizado no campeonato da europa masculino de 2006. *13th Annual Congress of the European College of Sport Science - Sport by the Sea*, Estoril – Portugal (2008), 450-451.
- 2. Volossovitch, A. A análise da performance em andebol: Perspectivas e tendências. *Revista Técnica De Andebol, 1*, 1 (2005), 16-20.
- Ribeiro, M. & Volossovitch, A. Andebol 2. O ensino do jogo dos 11 aos 14 anos. Cruz Quebrada, Portugal. Edições Faculdade Motricidade Humana, 2008.

- Anguera, M. T., Blanco Villaseñor, Á, & Losada López, J. L. Diseños observacionales, cuestión clave en el proceso de la metodología observacional. *Metodología de las ciencias del comportamiento*. Barcelona, España: Univ. Barcelona, Fac. Psicología, Dep. Metodología de Las Ciencias de Comportamiento (2001), 135-160.
- Perea, A., Alday, L., & Castellano, J. Software para la observación deportiva match visión studio. III Congreso Vasco Del Deporte. Socialización y Deporte/Kirolaren III Euskal Biltzarra. Sozializazioa Era Virola, Vitoria, 2004.
- 6. Magnusson, M.S. Hidden real-time patterns in intra- and inter-individual behavior. *European Journal of Psychological Assessment, 12*, 2 (1996), 112-123.
- 7. Magnusson, M. S. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods Instruments* & Computers, *32*, 1 (2000), 93-110.
- Magnusson, M.S. Understanding Social Interaction: Discovering Hidden Structure with Model and Algorithms. In L. Anolli, S. Duncan, M. Magnusson & G. Riva (Eds.), *The hidden structure of social interaction. From Genomics to Culture Patterns*. Amsterdam: IOS Press (2005), 51-70.
- Anguera, M. T. Registro y análisis de datos al servicio de la comprensión de la complejidad en deportes de equipo. In R. Martín Acero & C. Lago Peñas (Eds.), Deportes de equipa: Compreender la complejidad para elevar el rendimento. Barcelona: INDE (2005), 127-142).

DataPrism: a Tool for Visualizing Multimodal Data

Adam S. Fouse afouse@cogsci.ucsd.edu James D. Hollan hollan@cogsci.ucsd.edu

Distributed Cognition and Human-Computer Interaction Laboratory Department of Cognitive Science, University of California, San Diego 9500 Gilman Drive, La Jolla, CA, USA, 92093

ABSTRACT

We present DataPrism, a new interactive visualization tool to aid analysis of multimodal activity data. DataPrism enables analysts to visualize, annotate, and link multiple time-based data streams, including video, log files, and paper-based digital notes. Automated analysis is supported through a plug-in mechanism. A usage scenario is employed to describe DataPrism's current facilities.

INTRODUCTION

Advances in technology continue to improve the ability of researchers to study real-world behavior. A new generation of inexpensive digital recording devices and storage facilities are revolutionizing data collection, extending it into situations that have not typically been accessible and enabling examination of the fine detail of action captured in meaningful settings. This is important because in order to understand the dynamics of human activity we must understand its full context, and that can only be accomplished by recording and analyzing data of real-world behavior.

Consider three examples: affordable high-definition video and inexpensive digital storage have enabled recording video of activities from multiple angles; small wearable sensors have made it possible to capture detailed records of complex human behavior; and interaction with digital systems can produce time-stamped logs of performance. Each presents new opportunities to analyze activity in more systematic and precise ways. More generally, the ability to collect these and other forms of rich data provides an unprecedented opportunity for scientific scrutiny of activity as it occurs in real-world everyday situations.

Along with the exciting scientific possibilities created by new data collection abilities there are also clear analysis challenges due to the scale and richness of data that is now readily collected. For example, coding and careful analysis of video data remains extremely time-consuming. Other forms of data, such as sensor readings, while useful as separate data sources, may provide more valuable insights about activity when viewed and analyzed in combination. New visualization tools are needed to address the challenge of analyzing and synthesizing information from multiple data sources and at multiple time scales.

We have developed DataPrism, a tool that allows researchers to visualize time-based data from multiple sources and manually or automatically annotate the data. To elucidate its facilities, we present a scenario from a current use of DataPrism, involving analysis of the behavior of airplane pilots during simulated flight. Understanding their cognitive activity requires analysis of details of pilot behavior that can only be acquired through the collection of a rich set of multimodal data. While we focus here on a specific scenario, DataPrism is designed to support analysis of a wide range of multimodal time-based data. For example, it is currently being used in several other contexts, including analysis of children interacting with museum exhibits and of personal activity data captured with the Microsoft SenseCam [2].

Although several other systems have been developed to support various aspects of this visualization challenge, DataPrism represents a unique focus on the visualization of multiple diverse data sources. For example, numerous systems exist for coding and annotation of video, such as ELAN [6], VCode [1], and Diver [6]. While these systems have powerful coding capabilities, they are not designed for analysis or visualization of heterogeneous data, nor do they support easily extensible automatic analysis facilities. Other systems, such as The Observer XT [7], readily support multiple types of data, but are commercial products with limited open extensibility to visualize new types of data.

FEATURES

The primary goal of DataPrism is to support interactive visualization and analysis of multiple simultaneous data streams from the same activity. DataPrism supports data import from multiple sources, including video files, audio files, log files (encoded in a variety of formats), transcript data, and paper-based digital notes.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Visualization

The main DataPrism interface consists of a video player and one or more timelines arrayed below the video, as shown in Figure 1. A single timeline is shown by default, but multiple timelines can be added to visualize additional data, group annotations, or view different time scales. Each timeline has a visual marker (a vertical yellow bar toward the right on each of the timelines in Figure 1) to indicate the current position in the data, which is aligned with the current video location. This marker allows the researcher to relate a notable moment in one data stream to corresponding points in other data sets.

Video files can be viewed as with a typical video viewer. The viewer displays the current frame of the video as a researcher moves through the data and allows the video to be played at varying rates. In addition, frames from the video can be viewed on the timeline below the main video viewer. Frames from the video can also be associated with annotations and displayed in a pop-up window that appears when the researcher hovers the cursor over the annotations. If the video file has an audio track, a waveform of the audio can be shown on a timeline. Multiple video files can be shown in separate windows or on separate timelines.

Log file data can currently be shown in two ways: time series graphs and map depictions of geographic position. Time series graphs are displayed on timelines in the main DataPrism window, along with dynamic scales based on the current visible data. Geographic positions can be shown on



Figure 1. Main DataPrism Window, showing (from top): video, video frames, time series graph, and annotations.

3/2/2010 Adam flies, whitne Whitney briefs RCL whete boerd digara 114/0 looking out right window Potale on the note de in turin . Good alt. tak Somo

Figure 2. Digital notes (top) and map display with overlaid annotations (bottom).

a map that is retrieved from Google Maps based on latitude and longitude, as shown in Figure 2. As with timelines, the current time position in the data can be changed by clicking on the map. In this way, a notable movement in physical space can be used as a guide to explore data in the same way as a change in a time series graph.

Digital Pen¹ data from a researcher's paper notes can also be shown in two ways. First, markers can appear on the timeline that correspond to the time that a particular note was written. When the researcher hovers over the marker, an image of the note appears. Second, the notes as a whole are displayed, as shown in Figure 2. The pen marks are initially displayed in gray, but as the researcher moves

¹ We currently support Livescribe digital pens (<u>www.livescribe.com</u>) that record both pen strokes and associated audio.

through the data, the letters get darkened to indicate which notes correspond to the current time in the data. As with the timelines and the map, analysts can change the current time position in the data by clicking on particular notes to jump to the corresponding points in the rest of the data.

Annotations

Data annotations in DataPrism can be visualized on their own or overlaid on other data types. For example, a set of annotations might be created in relation to moments in a video that indicate particular types of activity. The corresponding segments of time can then be highlighted on a time series graph to depict how various variable values changed during those aspects of the activity. The annotations are color-coded with researcher-defined colors, and can be shown on the timelines and on maps.

Annotations can be used in multiple ways, including both continuous coding and identification of particular moments of interest in the data, as shown in Figure 3. Annotations can be simple assignments of categories or free-form text entries. Categories and possible subcategories can be defined before analysis by the researcher, and can be dynamically modified and extended during analysis. Annotations can be assigned to multiple categories and used to group, align, and filter annotations.

Synchronization

Synchronization can be a problem when separate devices are used to record multiple data sets. The ideal solution is to precisely calibrate system clocks in each of the recording devices, in which case data is automatically aligned when imported into DataPrism. Since this is not always possible with heterogeneous recording devices, interactive alignment of data is also available. With each supported data type, including multiple videos, the researcher can find a point that corresponds to another point and align the data simply by clicking on the identified point.



Figure 3. Different ways annotations can be coded in DataPrism. Top: Color-coded segments of data that have been manually defined to indicate phases of activity; Middle: Values with respect to a single variable assigned continuously over the course of the video, using vertical position to indicate value; Bottom: Instances of different types of activity, using vertical position to group by activity type.

Automatic Analysis

One important use of visualization is to present the results of automatic processing of data streams. DataPrism supports a data analysis plug-in system for scripts written in the Python language. These scripts can operate on any data that resides in DataPrism. The results are treated as annotations and can be visualized and filtered in the same way as normal annotations. This makes it easy, for example, for an analysis based on one set of variables to highlight a region of important activity across the entire collection of data. In the domain of flight, one type of data analysis that we have explored uses logic-based rules to identify flight segments based on well-defined flight parameters. Similarly, simple rules can be used to identify moments, such as when the airspeed is within a particular range, which might be of interest for a particular analysis.

USAGE SCENARIO

In this section, we step through a simplified description of the use of DataPrism to analyze data from a flight simulator. The scenario is based on real-world DataPrism sessions we observed. The data is from novice pilots flying the flight simulator, which created text logs of simulator variables at a rate of 5 Hz. Paper notes were made by an observer during the flight with a LiveScribe digital pen.

Analyses using DataPrism in this domain typically include the following data configuration:

- Video of pilot-simulator interaction, shown in primary video window
- Video of pilot-pilot interaction, shown in secondary video window
- Simulator log of Altitude, Airspeed, Vertical Speed, and Bank Angle variables, shown as time series
- Geographic position from simulator log, shown as a path on a map
- Notes made by the researcher, shown as annotations and as a full page

After loading the videos, simulator log, and digital notes, the researcher begins by aligning the data. When the videos were made, a clapperboard was used at the beginning to make sure that devices could be synced. The appropriate places in the movies are found, and then the data is matched to the videos by finding the point where the plane begins to move in the airspeed data and the video data. Finally, the digital notes are synced to the same point by matching a note made during the simulation at that time. This simulation run involved novice pilots, so there are many possible errors to identify in the data. On an initial visual scan of the data, the analyst notices a period of fluctuation in the vertical speed graph. By moving the current time position indicator, he gets visual reference marks of the corresponding time position in all of the other data sources. As he moves the time indicator, the position indicator on the map moves as well. This reveals that the vertical speed fluctuation occurred at the same time as a sharp turn in the flight pattern. The relevance of the fluctuations occurring

during a turn is reinforced by a sudden jump in the bank angle graph at the same time. The shape of the bank angle graph reveals a sharper turn than would be needed in this situation. The combined evidence points to a lack of skill in the novice pilot for maintaining proper flight parameters during turns. By examining the visualizations from these different sources, the researcher forms a hypothesis that the cause of the poor performance might be an inability to properly manage attention between instruments depicting bank angle and vertical speed. To help evaluate this hypothesis, he looks at the video of the flight and notices the pilot's eyes seem to be locked on the artificial horizon and rarely move to the right to the location of the altitude indicator. This short scenario provides an example of how DataPrism might be used to identify an instance of poor flight performance. The important things to note are the usefulness of being able to visualize multiple data sources and how access to linked visualizations can help evaluate hypotheses that arise during analysis.

FUTURE WORK

While the current capabilities of DataPrism make it an effective tool for analysis of multimodal behavioral data, we have plans for extending it to create more powerful analysis facilities. Our design goal is to maintain the simplicity of the current interface but enable a loose coupling via a plug-in architecture with additional facilities useful for particular domains or types of analyses.

One natural direction is to explore how additional automatic analysis features might be supported. For example, computer vision techniques have advanced in capabilities and reliability to the point that they promise to be highly useful tools for aiding analysis of video data. Object recognition is one particularly promising area. It would be very useful for digital video analysis if an algorithm could automatically label all (or even most) video frames in which a particular object is present. The use of scale invariant features (e.g., SIFT and related techniques [4]) is one promising approach we are exploring. We are especially interested in supporting interactive use of automated analysis facilities. For example, since many computer vision algorithms are probabilistic, users can modify the algorithm's threshold depending on the task. For example, the threshold for object detection could be set at a low value in which virtually every frame that contains the object is detected, with the price of increased false alarms. In this case, a small amount of user intervention would be required in order to cull the false alarms from true detections. On the other hand, the object detection threshold could be set at a high value in which case there would be virtually no false alarms, with the price that in some frames the object would be present but not detected. Integrating computational analysis methods with human interaction in mind allows such decisions to rest with the analyst, and be interactively adjusted to specific analysis needs. We are also actively developing methods for integrating DataPrism with additional facilities of digital pens. A digital copy of

freehand notes is a useful addition to analysis, but digital pens have several advantages over traditional pens that can be exploited during real-time observation. For example, since each stroke of the pen is time-stamped, this can eliminate the need to manually record time-stamps for later indexing into the data. In addition, there is considerable promise in linking paper notes and their digital counterparts using pen gestures [3]. We are developing a system to allow marks made in certain regions on paper (e.g., in specific areas in the margin or on customizable forms) to have special meaning and be automatically converted into annotations.

Finally, we plan to engage in further detailed studies of researcher interaction with DataPrism. While the scenario described in this paper is based on actual observations of real usage, conclusions about the efficacy of the interactive visualization techniques require additional data, including usage in other domains.

ACKNOWLEDGMENTS

NSF Grant 0729013 supported this research. We thank users of DataPrism and members of our lab for feedback and design suggestions.

- 1. Hagedorn, J., Hailpern, J., & Karahalios, K. VCode and VData: Illustrating a new Framework for Supporting the Video Annotation Workflow. *Ext Abstracts AVI 2008*, 317-321.
- Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., & Wood, K. SenseCam: a Retrospective Memory Aid, *Ubicomp 2006*, 177–193.
- Liao, C., Guimbretiere, F., Kinckley, K., and Hollan, J. Papiercraft: A Gesture-Based Command System for Interactive Paper. ACM Transactions on Computer-Human Interaction, 14 (2008) 1-27.
- 4. Lowe, D. Distinctive Features from Scale Invariant Keypoints. Intl J of Computer Vision, 60 (2004), 91-110.
- Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert, E. The Diver Project: Interactive Digital Video Repurposing. *IEEE MultiMedia*, 11, 1 (2004), 54-61.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. ELAN: a Professional Framework for Multimodality Research. *Proc. LREC 2006*, 1556-1559.
- Zimmerman, P.H., Bolhuis, J.E., Willemsen, A., Meyer, E.S., & Noldus, L.P.J.J. The Observer XT: A tool for the integration of and synchronization of multimodal signals. *Behavioral Research Methods*, 41, 3 (2009), 731-735.

Video Deep Tagging and Data Archiving in the Comparative Mind Database

George Kampis

History and Philosophy of Sci., Eötvös University 1117 Budapest, Pazmany P.s. 1/C, Hungary gk@hps.elte.hu

Zsofia Viranyi

Konrad Lorenz Inst. for Evol. and Cog. Research 3422 Altenberg, Adolf Lorenz Gasse 2, Austria zsofi.viranyi@gmail.com

ABSTRACT

In this paper we describe our efforts towards a Comparative Mind Database and discuss an available component, the video sharing and deep tagging system. We present the "Comparative Cognition" project of which the database component is a module, and present general discussions of data archiving and processing in various scientific contexts, from astrophysics to animal behaviour. Afterwards, we turn to the discussion of the developing system and various aspects of its benefits. Finally, we describe the fundamental features of the video system and its advanced, cutting edge functionalities to support the study of animal behaviour.

Author Keywords

Comparative cognition, comparative mind database, data archiving and sharing, video archive, deep tagging

INTRODUCTION

The Comparative Mind Database (<u>www.cmdbase.org</u>) is an activity of the project "The Evolution of Social Cognition: Comparisons and integration across a wide range of human and non-human animal species", an ESF Research Networking Program (<u>www.compcog.org</u>). The general objective of the parent project is to develop "real" comparative cognition across a wide range of vertebrate and invertebrate species (including humans) with a coherent theoretical background, unified terminology and standard methods, and to make it transparent for and integrated with other fields, like social sciences, genetics, physiology, animal welfare, robotics, and so on. The program brings

Adam Miklosi

Institute of Biology, Eötvös University 1117 Budapest, Pazmany P.s. 1/C, Hungary amiklosi62@gmail.com

Laszlo Gulyas

Aitia International, Inc. 1039 Budapest, Czetz János u. 48-50, Hungary lgulyas@aitia.ai

together 28 European laboratories from 11 countries, and runs for 5 years since May 2008.

Despite the existence of a large body of fragmented knowledge, we do not currently understand human and animal cognition. What constitutes the animal that we talk about, i.e. how can it be conceptualized? What are the terms to be used, and how they are backed up in the various disciplines that are relevant for the experimental and theoretical study of animal and human behaviour? How can different species and different experiments be compared with, if they measure different things (or even if they claim to measure the same thing)? What are justifiable conceptual and experimental methods that can support reproducibility, scalability, and generalizability? We believe these questions are strongly interrelated with what can be broadly called *documentation* and *conceptualization*.

The Comparative Mind Database is a module of the CompCog project that supports the above inquiry using innovative, advanced information technologies and methods from philosophy of science, statistics, experimental design and data/text mining. Its envisioned components include: Web2.0 style computational tools and methods for data acquisition and description, procedures for data and text mining for conceptual analysis, specific ontologies for animal cognition, integrated tools for the design and analysis of experiments (DoE), and work towards the standards of experiments, their communication and evaluation.

DATA ARCHIVING AND PROCESSING

The public archiving and sharing of data is becoming almost ubiquitous. Since the recent declaration of various platform statements, many journals have adopted an archiving and sharing policy. The platform statements in question range from the announcement of the PetaByte age [1, 4] through the Jerusalem declaration [3] to the notion of Web Science [7]. The essence of the Jerusalem declaration is: "public money – public data" for research, and this

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

formulates a very natural requirement for data sharing and archiving.

Many of the above topics point towards the usefulness of comprehensive, integrated databases. Databases can come in various formats from the simple deposits to complex systems that use suggestive tools for the acquisition, handling and analysis of data. The database could be also suggestive as a unified documentation, synthesis and repository tool for much of existing information. Such database initiatives include the "virtual observatory" of astrophysics. In astrophysics, a prior submission of all data before publication is required today [6, 10]: The virtual observatory enables a new way of doing astronomy, moving from an era of observations of small, carefully selected samples of objects to the use of multi-measurement data for millions, or billions of objects. Such datasets allow researchers to discover subtle but significant patterns in statistically rich and unbiased databases, and to understand complex astrophysical systems through the comparison of data to numerical simulations. The virtual observatory provides simultaneous access to multiple archives and advanced visualization and statistical analysis tools.

Recently, evolution and ecology journals also move towards a Journal Data Archiving Policy [9] and at this time the journal Animal Cognition is considering the use of a similar practice (Czeschlik, personal communication, 2010). These policies – similar in vain to those in use in astrophysics since almost a decade – put strict conditions on publications by requiring the publication (i.e. public archiving) of all relevant data.

THE COMPARATIVE MIND DATABASE INITIATIVE

The development of the Comparative Mind Database capitalizes on these prior approaches and results, and extends them for the specific purpose of the study of animal behaviour. Databasing animal behaviour poses new kinds of challenges, however. Comparative cognition aims at reconstructing the evolutionary history of various cognitive skills via investigating which animal species demonstrate or lack them. Such species comparisons require comparable methods utilized by different research labs keeping and/or working with subjects of one or few species. For greater importance even subtle variations in observational and experimental methods can differentially influence the animals in study as well as the interpretation of data. The current form of a paper's method section and the supplementary demo videos hardly provide enough data for the reproduction of studies on animal cognition. These challenges can be overcome by developing a suitable system for animal cognition related data archiving and sharing.

Over the last years there have been a few attempts to develop database(s) for the behavioural sciences that do not seem to have achieved their aims. One may argue that the fundamental problem lies in the complexity of the structure under study. But also, in contrast to other fields, like astrophysics or genetics where complexity might be a similar issue, the behavioural sciences have not developed a useful definition of "data". A recent review showed that for many researchers even the definition of the term "behaviour" was quite elusive [5]. Following their survey, these authors are inclined to define behaviour *as "the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes"*. Importantly, however, this definition avoids referring to those physical (visual, acoustic etc.) attributes of an "action" that should be "captured" for such a database, that is, for this definition the behaviour should be "describable" in some general ways.

Importantly, in contrast to astrophysics, genetics and cosmology, most of the behaviour data are still obtained more or less by direct human intervention (especially in the animal and human cognition domain), which is deficient in "objectivity". In contrast, physical parameters measured by specifically designed (and validated) equipments used by a scientific community (e.g. Hubble space telescope) or the units of the genetic code provide a clear case for data banking because the data have been collected reproducibly by standardised (non-human) instruments.

However, the behavioural sciences cannot avoid moving into the age of data banking. This process has already started. For example, software technology allows for automated collection of some specific behavioural data (movement in space, simple interaction with objects, etc.) in simple environments for laboratory species (e.g. EthoVision - Noldus), wild living animals can be equipped with GPS sensors, etc., but especially in the field of cognition most data are collected "by hand" utilising a diverse array of experimental approaches. Importantly, modern technologies emerge offering more objective collection of such behavioural data. For example, "looking time" is often used as an index for different cognitive processes. Traditionally, the human experimenter (observer) noted the duration of a particular head orientation by hand held timers, or more recently, by counting video frames. However, in some particular cases it is now possible to use eye tracking devices to obtain observer-independent data on "looking". In an ideal situation the change from a human-based coding to a machine based data collection would make such data more useful for data banking.

Despite such advances in technology objective, "en mass" data recording from freely moving, "behaving" systems will not be achieved in the near future. Nevertheless, in this time there is a possibility for making advances in areas in which some basic requirements for data banking are in place. Thus we suggest that a Comparative Mind Database (CMD) could have two important contributions for the behavioural sciences.

The Weak Form of CMD

The idea of weak CMD is based on the experience that despite sometimes lengthy descriptive method sections, more or less clear definitions of behaviour, ad hoc provided video files, and intra- and inter-observer reliability, the "objective" definition of the data is inadequate for data banking. The weak form of CMD can be seen as providing a suitable framework for collecting and sharing basic data in forms specific to the behavioural sciences. It could be useful as an aid (or eventually an obligatory tool) for publishing behavioural data in scientific journals, to teach students how to plan experiments, to share data and methods with fellow scientists, and to contribute to a common data base. The utilisation of video data together with tools that help to describe the "behaviour" could canalise data collection, eventually leading to data which have been collected in very similar ways.

The Strong Form of CMD

The strong form of CMD is aimed at providing a storage space, definitions and ways of sharing, that is, bringing behavioural data from different sources onto a common platform, providing interlinking and common reference. Note that the strong form of CMD is useless until the weak form of CMD is realized, because data from under-defined data sources are worse than no data at all.

SOLUTIONS FOR A COMPARATIVE MIND DATABASE

We started to realize these plans by working on a prototype of the CMD. Besides adopting existing technologies found in different domains mentioned earlier, our suggested new solutions include an innovative copyright handling, integration, raw/select data archiving, and online/offline processing. One of the notorious issues is copyright handling. The current Journal Data Archiving Policy foresees an embargo prior to publication (where the data is uploaded but not shared, and a moratorium of one year after publication during which the data owner has priority of publication using secondary analysis). This is an important step but may not be sufficient in all cases – our proposed system offers different copyright formats selectable by the owner at upload time.

Another recurrent issue is integration: many data archives are scattered and put their data in a different context than it was produced or processed. This raises several questions from the attractiveness to the usefulness of such archives. We are attempting to overcome this issue by endeavouring to develop a framework for the uniform handling of experiments (form design to evaluation and theoretically to paper writing), and the uses of advanced methodological tools for data analysis such as MEME (developed by one of our contributors, Gulyas 2010, [2, 8]). Usually, by "data" something distilled is meant – most archiving initiatives do not concern "raw data" in the form of lab notebooks or video recordings. However, for animal behaviour raw data is also of significance, both for secondary analysis and quality control. We are treating raw and processed data (below: video footage and edited shots) in the same way, interlinked with pointers and tags.

Finally (and also related to the copyright issue, but also to data safety and availability), many researchers might be reluctant to put their current experiments into the public domain, and especially to use Web 2 tools for strategic mass storage as the only form of archiving. As a solution to this, and handled together with the raw/edit data policy, our tools will be available online as well as in an offline version (with the online version providing more "goodies" such as integrated data processing and "paper writing" tools). This way the scientist can use the system in her laboratory as well as in the public domain and turn to the latter only at paper writing time, or for edited data only, while keeping (i.e. archiving) raw data in her local system. During its five years of operation, the CompCog/CMD project endeavours to develop a unified encompassing framework and toolkit for such data preparation, handling, archiving and processing task, for raw data (such as video footage) and processed data alike. The prototype CMD is focused on video representations, and is available in the form of a fully functional alpha version (announced in May 2010).

VIDEO ARCHIVING, SHARING AND DEEP TAGGING

The CMD video platform is a community platform that offers several different functions. These functions include: registration, video upload, video editing (in particular, tagging, annotation and deep tagging), timeline control, video (as well as annotation) search and video playback (in various sizes). It takes the form of a video sharing system on a backbone similar to that of YouTube. The system is available at www.cmdbase.org.

Video tagging is a common technique to assign metadata (such as author, title and time information) to videos in an archive. Deep tagging is a recent development that extends advanced tagging (i.e. the use of tags, bookmarks, and other descriptive elements, including attachments of various kinds) to the temporal domain, by permitting the data owner to enter the "depth" of the video. Deep tagging involves the use of temporal tags and a timeline that positions these tags over the duration of the video. Deep tagging became widely known in 2007 and is now available in various commercial products such as Momindum (www.momindum.com) (however, they are costly and not open source, hence not scalable and extendable, nor tailored to scientific needs).

Video annotation goes even one step beyond deep tagging. The idea is to allow users to place marks, even drawings (including simple shapes, text, and hand drawing/writing to frames or intervals of the video. The system features an interval editor that configures each tag and annotation in time: the minimum duration is 1 second; the maximum is the video's entire playtime. In edit mode, the timeline shows the timing and the overlay of different instances and the layers of all tags and annotations. Video annotation may have uses in different domains but currently few tools exist that can support it. In animal behaviour studies, together



Figure 1. Editing and deep tagging in the CMD system.

with deep tagging, video annotation may have a key role in the future for the identification (and documentation) of various behaviour elements, their relation to each other and the experimental design, and the mark-up of experimental settings, which includes the identification of various environmental cues. Our realized tool has a flexible management system that allows various policies for both tagging and annotation: from owner-only to general access. Currently no other available system supports the variety of these functions, ranging from archiving and sharing to deep tagging and video annotation. The tool is open source and freely extendable.

ANNOTATION AND DEEP TAGGING: AN EXAMPLE

The system has three main pages: namely, the main portal page, the editor and the player. A toy example is shown below for the use of the editor page to create video markup in an imitation experiment (found in the database at www.cmdbase.org under the name MouthFull.mpg). The editor page shows the video and a timeline where a relocatable line marks a current moment. Positioning the line to a desired moment, graphical annotations (here at 21. sec, an ellipse and two text boxes) can highlight critical parts of the picture or provide short written explanations. Their temporal duration is also shown in the timeline and can be controlled by dragging their markers' ends. Tags and bookmarks (at the upper right panel) support navigation on the video: critical moments can be marked and jumped on with one single click. In the example on Fig. 1, the animal is captured in the position where, in contrast to the preferred paw action of dogs, she is demonstrating an unpreferred paw action to manipulate an apparatus, which is justified by her mouth being occupied by a ball she carries during the demonstration. The experiment investigates whether observer dogs copy the model's paw action after watching 10 demonstrations. Rather subtle behaviours need to be noticed during the 8 test trials. In contrast with demo videos typically used in behavioural sciences, the new system can offer footages where all methodological details of the experiments can be followed in the archived experimental video, and at the same time tagging the video can highlight

the critical but subtle events in the animals' behaviour (see red arrows and text boxes at moments 0:06:02, 0:06:53 and 0:06:58). Attachments such as a ppt slide or a pdf document can also be associated with a given frame using the bottom right panel. (Attachment links are shown in the player and until a new attachment comes up.) They can either provide further details missing from the video (such as structure of the apparatus on the first ppt slide in this example (0:00:12)) or can directly connect the behaviour observed and the graphs presented in the publication PDF (e.g. 0:06:02).

CONCLUSION

We presented general discussions and ongoing work towards a comprehensive, domain specific database for comparative animal cognition with archiving, sharing and data processing capabilities and we presented an example component, a fully functional video sharing, annotation and deep tagging system, available in an alpha version.

- 1. Anderson, CH. 2007. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
- Gulyás, László, Bartha, Sándor, Kozsik, Tamás, Szalai, Róbert, Korompai, Attila and Tatai, Gábor, "The Multi-Agent Simulation Suite (MASS) and the Functional Agent-Based Language of Simulation (FABLES)", SwarmFest Conference, Santa Fe NM (2005).
- 3. Jerusalem Declaration (2008): http://ocs.unipa.it/Declaration20081026.pdf
- 4. Kampis, G. 2008: Science in the Petabyte Age. http://www.colbud.hu/programme/calendar/event.shtml? cmd%5B8%5D=i-8d235e57fb3477697cf98aabbc533e817
- Levitis, D.A., Lidicker, W.Z., Freund, G. Behavioural biologists do not agree on what constitutes behaviour, *Animal Behaviour* 78 (2009), 103–110.
- 6. NVO, 2002: Towards the National Virtual Observatory.A Report Prepared by the National Virtual Observatory Science Definition Team, http://www.usvo.org/pubs/files/sdt-final.pdf
- 7. Riera, D 2008: Science, the Web and Web Science, http://www.uoc.edu/uocpapers/7/dt/eng/riera.pdf
- 8. MEME 2010: http://mass.aitia.ai/intro/meme
- 9. Whitlock, M.C., McPeek, M.A., Rausher M.D., Rieseberg L., and Moore A. J. Data Archiving, *The American Naturalist 175*, 2 (2010), Editorial.
- 10. Williams, R. et al. (2009): The Role of the Virtual Observatory in the Next Decade, http://www.usvo.org/pubs/files/Williams_VO_TEC.pdf
- Zellweger, P.T., Bouvin, N.O., Jehøj, H., and Mackinlay, J.D. Fluid Annotations in an Open World. Proc. Hypertext 2001, ACM Press (2001), 9-18.

Using Postural Control System Measures to Detect Hypovigilance

Thomas Schnupp¹ t.schnupp@fh-sm.de Adolf Schenka¹ a.schenka@fh-sm.de

David Sommer¹ d.sommer@fh-sm.de

Martin Golz¹ m.golz@fh-sm.de

¹ University of Applied Sciences Schmalkalden Blechhammer, 98574 Schmalkalden, Germany

ABSTRACT

Posturography is a method to assess the postural control system quantitatively providing the possibility for testing vigilance. In this paper we present pilot experiments and analysis investigating the discriminatory abilities of posturography. A total of 10 young adults participated in postural assessment within a study with extended time awake. Acquired measurements were assigned to two classes "vigilant" and "hypovigilant" according to subject's continuous time-since-sleep (TSS). Two features sets were extracted from posturographical recordings. In time domain 7 kinds of features from other authors were extracted, including measures of sway velocity and sway area. In spectral domain features were extracted by estimating power spectral densities and subsequent averaging in equidistant spectral bands. In addition to static feature extraction this paper introduces analysis of temporal dynamics within the features of both domains. The ability to discriminate between both classes "vigilant" and "hypovigilant" was evaluated in terms of mean test set errors estimated using 25-fold delete-d cross validation. Different algorithms of computational intelligence including artificial neural networks (ANN) and Support-Vector Machines (SVM) were applied. SVM using Gaussian kernel functions performed best with achieved mean test set error rates of 9.0 ± 4.2 %.

Author Keywords

Posturography, Vigilance, Support-Vector Machines, Computational Intelligence, LSTM.

Jarek Krajewski² krajewsk@uni-wuppertal.de

 ² University Wuppertal, Experimental Industrial Psychology, Gaußstraße 20, 42097 Wuppertal, Germany

ACM Classification Keywords

I.5.4 Pattern Recognition: Signal Processing

INTRODUCTION

Posturography is a method of quantitative balance assessment and is used as a diagnostic tool in Neurootology. Posturography measures body sway during sitting or standing. For the latter subjects are instructed to stand upright and remain as motionless as possible. The center of pressure (COP) is calculated utilizing a three- or four-point force sensor platform.

Data analysis quantifies features of the medio-lateral and the antero-posterior components of the COP time series (Figure 1). It has been shown, that both components are impaired by hypovigilance [1, 2]. Therefore, posturography is a new candidate for vigilance assessment which may be cost-efficient, short-lasting, mobile, and easy-to-administer. Recent studies in our lab resulted in relatively high error



Figure 1. Stabilogram recorded after 14 hrs of continuous time awake under EC condition. Test duration was 120s. The gray polygon indicates the area of the convex hull – one of the utilized time-domain features.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

rates when classifying between two extremes: "vigilant" vs. "hypovigilant" [3]. One possible explanation may be the kind of data analysis. For each signal segment a set of different features were extracted. But it was not asked whether temporal dynamics between feature sets of different segments exist. The application of Recurrent Neural Networks (RNN), for example Long Short-Term Memories (LSTM) [4], provides means to answer this question.

Long Short-Term Memory

RNN are ANN characterized by feedback connections between artificial neurons. These connections conduct output variables from one neuron to its neighbors' and predecessors' inputs [5]. In contrast to non-RNN, they are able to process sequences of feature vectors such that the temporal localization of feature characteristics within the sequence gains relevance. RNN are capable to adapt to temporal dynamics within the input sequences. LSTM are trained by a variant of Real-Time Recurrent Learning [4].

An alternative approach is the application of Evolutionary Strategies to adapt input weights of network cells [6]. Schmidhuber et al. [6] demonstrated that combinations of RNN with multivariate output (Fig. 2) with non-linear classifiers have an improved ability to generalize a subsymbolic temporal memory in the feature space.

METHODS

Study Design

In this contribution a dataset of a partial sleep deprivation study is processed. A total of 10 young and healthy adults volunteered to participate in postural assessment. Posturography was performed during morning hours (10:00-12:00AM) and during eight test runs, hourly between 8:00PM and 4:00AM.

Each test run consists of two conditions. The first trial was performed with eyes opened (EO), focusing a marker on the opposing wall. The second test run was performed with eyes closed (EC) disabling the visual feedback to human balance control. This trial is more demanding and may therefore be sensitive to performance decrements at an earlier stage. Both trials lasted 120 seconds. Subjects were instructed to stand as motionless as possible and to maintain a uniform, practiced upright pose with both feet on the ground.

Pre-Processing

Samples were divided into two classes. The first class ("vigilant") contained test runs of the morning (time-since-sleep < 5 hours) and the second class ("hypovigilant") contained test runs of the late night (time-since-sleep > 14 hours).

During segmentation recorded data is split into nonoverlapping segments. Using only the first segment it is possible to simulate different test durations. The segment length was optimized empirically regarding the achieved



Figure 2. Signal Processing Chain: feature sequences of l elements consisting of n features each are mapped to a twodimensional space by LSTM. For linear decisions (TH, dashed line) classificiation errors may occur. Non-linear classifiers (SVM, dotted line) can adapt to more complex distributions.

mean test set error estimated via 25-fold delete-d cross validation [7]. The optimal segment length obtained this way is an indication for the required test duration.

Feature Extraction

In time domain methods utilized by several authors were applied, including, but not limited to amplitude range, measures of velocity and of sway area. In total 32 features were extracted.

In spectral domain power spectral densities have been estimated using Weighted Overlapped Segment Averaging (Welch's method). Estimated power spectral densities have been averaged in equidistant spectral bands. Band averaging parameters are lower and upper cut-off frequency, and band width. They were empirically optimized in terms of lowest test set errors. The number of features varies depending on band averaging parameters.

In addition to full-segment feature extraction a sliding window feature extraction was implemented resulting in sequences of feature vectors. The window size and the percentage of overlapping were both optimized empirically regarding the achieved test set error rates.

Classification

Different methods of computational intelligence have been applied in order to find generalized discriminant functions. We applied Fisher's Linear Discriminant Analysis (LDA), k-Nearest Neighbor (kNN), Learning Vector Quantization (LVQ), Long Short-Term Memories (LSTM), and Support Vector Machines (SVM). Free parameters of these methods were optimized empirically in analogy to feature extraction. The outcomes of sliding window feature extraction (LSTM, SVM_{window}) were compared to the outcomes of static approaches (LDA, kNN, LVQ, SVM).

CONCLUSIONS

This contribution introduces the utilization of sliding window features to Posturography in order to test for vigilance. Preliminary results of a relatively small data set of 10 partially sleep deprived subjects offered that further effort have to be made to improve classification accuracy. In future Evolutionary Strategies should be applied to LSTM

| Classifier | Error - Time Domain | Error - Spectral Domain |
|----------------------------------|---|--|
| kNN | $35.8 \pm 8.7 \%$ k=44 | 18.8 \pm 9.0 % k=8 |
| LVQ | 37.1 ± 8.9 % <i>n=2</i> | $19.0 \pm 9.0 \%$ n=41 |
| SVM | 22.4 ± 8.3 % <i>C=1e3.75 γ=1e-3.375</i> | 9.0 ± 4.2 % <i>C=1e6.375, γ=1e-2.188</i> |
| $\mathrm{SVM}_{\mathrm{window}}$ | 21.6 ± 7.4% <i>C=1e0.2813, γ=1e-2.325</i> | 10.4 ± 6.1 % C=1e6.875, γ=1e-4.875 |
| LSTM | 29.9 ± 8.0 % 6 blocks, 6 cells each | 12.0 ±-5.9 8 blocks, 8 cells each |
| LSTM + SVM | 21.1 ± 5.6 % 6 blocks, 6 cells each C=1e1.375, y=1e-1.75 | 15.3 ± 1.5 % 8 blocks, 8 cells each C=1e-1.625, y=1e0 |

Table 1. Comparison of classification errors.

training in order to optimize the many free parameters as well as the weight values. Further recurrent methods should

be utilized in order to investigate the potential of sequence learning in posturography. Results of Collins et al. [8] indicate that posture in quiet stance is not explainable be random walk models. There are short-term as well as longterm correlations which should by assessable also by subsymbolic sequence learning.

Despite this dataset being sufficient for first course analysis, the very important questions of inter- and intra-subject variability cannot be examined using this limited dataset. Especially when focusing on intra-subject variability a sufficient large amount of measurements for each subject and different levels of vigilance is necessary. In order to establish posturography as a method for fit-for-duty testing the limitations of this approach regarding inter- and intrasubject variability must be known. In Addition to the classification approach a regression of TSS using posturographical measures is possible [1]. The training of a regression model, e.g. using ANN, can include the whole dataset and is not limited to distinct TSS intervals.

Compared to other vigilance assessment approaches posturography still shows relatively high error rates. Nevertheless it has to be emphasized that these error rates are achieved based on short testing durations. Empirical parameter optimization of segment lengths shows that error rates decrease with increasing test duration. With minor decrements compared to the optimum test durations of 45 seconds seems applicable. Shorter test durations do not only increase the acceptance of the test but also limits testinduced vigilance reducing effects. So far test durations of approximately 10 minutes are common [9, 10].

- 1. Liu, Y., Higuchi, S. and Motohashi, Y. Changes in postural sway during a period of sustained wakefulness in male adults. *Occup. Med.* 51, 8 (2001), 490-495.
- Forsman, P., Hæggström, E., Wallin, A., Toppila, E. & Pyykkö, I. Daytime changes in postural stability and repeatability of posturographic measurements. *J. Occup. Environ. Med.*;49(6) (2007), 591-596.
- 3. Schnupp, T., Schenka, A., Edwards, D., Krajewski, J. and Golz, M. Is Posturography a Candidate for a Vigilance Test? World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings 25/IX, Munich, Germany, (2009), 388-389.
- 4. Hochreiter, S. and Schmidhuber, J. Long Short-Term Memory. *Neural Computation*, 9, 8 (1997), 1735-1780.
- 5. Mandic, D.P. and Chambers, J.A. *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures, Stability.* Wiley & Sons, London, 2001.
- 6. Schmidhuber, J., Gagliolo, M., Wierstra, D. and Gomez, F.J. Evolino for recurrent support vector machines. *ESANN*, 2006, 593-598.
- 7. Golz, M. and Sommer, D. Automatic knowledge extraction: fusion of human expert ratings and biosignal features for fatigue monitoring applications. In: Mandic DP, Golz M et al. (eds) *Signal processing techniques for knowledge extraction and information fusion*. Springer, New York, (2008), 299-316.
- Collins, J.J., De Luca, C.J., Burrows, A. and Lipsitz, L.A. Age-related changes in open-loop and closed-loop postural control mechanisms. *Exper Brain Res* 104, 1995, 480-492.
- Makeig, S. and Jolley, M. COMPTRACK: A compensatory tracking task for monitoring alertness. Technical Document 96-3C. San Diego: Naval Health Research Center, 1996.
- Wilhelm B., Widmann, A., Durst, W., Heine, C., Otto, G. Objective and quantitative analysis of daytime sleepiness in physicians after night duties. *Int J Psychophysiol* 72, 2009, 307-313.

Measuring Instant Emotions During a Self-Assessment Test: The Use of FaceReader

Vasileios Terzis bterzis@otenet.gr Christos N. Moridis papaphilips@gmail.com Anastasios A. Economides economid@uom.gr

Information Systems Department, University of Macedonia, 156 Egnatia Avenue, Thessaloniki 54006, Greece

ABSTRACT

Emotions are very important during learning and selfassessment procedures. Measuring emotions is a very demanding task. Several tools have been developed and used for this purpose. In this paper we evaluate the efficiency of the FaceReader during a self - assessment test. We compared instant measurements of the FaceReader with the researchers' estimations regarding students' emotions. The observations took place in a properly designed room in real time. Statistical analysis showed that there are some differences between FaceReader's and researchers' estimations regarding Disgusted and Angry emotions. Generally, results showed that FaceReader is capable of measuring emotions with an efficacy of over 87% during a self-assessment test, and that it could be successfully integrated into a computer-aided learning system for the purpose of affect recognition. Moreover, this study provides useful results for the emotional states of students during self-assessment tests and learning procedures.

Author Keywords

FaceReader, e-learning, self-assessment test, emotion recognition.

ACM Classification Keywords

H.5.1. Evaluation/methodology

INTRODUCTION

Measuring emotions could be crucial for many fields, such as psychology, sociology, marketing, information technology and e-learning. Consequently, several researchers have developed their own instruments to assess emotions [14]. Research evidence supports the existence of a number of universally recognized facial expressions for emotion such as happiness, surprise, fear, sadness, anger and disgust [5]. Therefore, estimating emotional experiences from objectively measured facial expressions has become an important research topic. Many facial recognition systems use single facial images instead of tracking the changes in facial expressions continuously [9]. Other facial recognition systems employ advanced videobased techniques [6] or measure the electrical activity of muscles with EMG (facial electromyography) [10].

Until now, machines using video cameras have been the predominant methods in measuring facial expressions [2, 8, and 11]. VicarVision and Noldus Information Technology launched FaceReader, a system for fully automatic facial expression analysis [13]. The FaceReader recognizes facial expressions by distinguishing six basic emotions (happy, angry, sad, surprised, scared, disgusted), plus neutral, with an accuracy of 89% [3]. Several studies have used FaceReader for different purposes [1, 12].

With regard to learning, there have been very few approaches for the purpose of affect recognition. A realtime analysis should be incorporated in human-computer interaction [7], especially concerning computer-aided learning systems. Previous studies in different fields showed that FaceReader is a reliable measuring tool. However, learning and self-assessment are procedures with particular characteristics. The aim of this paper was to evaluate the effectiveness of the FaceReader 2.0 during a self- assessment test. Accordingly, FaceReader's efficiency was measured in comparison to two experts' opinions.

METHODOLOGY

Participants were undergraduate students. The course was a basic IT (Information Technology) skills course and the syllabus included knowledge and techniques. The self-assessment test was optional. Students filled in an application form in order to participate in the self-assessment test. The test consisted of 45 multiple choice questions and the time limit was 45 minutes. 208 applications were collected. The next step was the arrangement of the appointments. Finally, 172 applicants out of the 208 came to their appointments.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Each student took the test alone in a properly designed room. The room had two spaces. There was a bulkhead between the two spaces. At the first space, there was the PC on which the self-assessments test took place. Moreover, the camera of the FaceReader was hidden in a bookcase. It is well known that people express themselves more freely when they feel that they are on their own.

In the second space were the two researchers. FaceReader was connected with another PC in that space, so the researchers were able to watch the facial expressions and the emotions of the participants in real time. Each researcher recorded the student's emotions measured by the FaceReader and his estimation regarding the student's emotion at the same time.

The purpose of this study was to examine the efficiency of the FaceReader during a self-assessment test. In order to accomplish this aim, the results of the FaceReader were compared to the researchers' estimations.

RESULTS

Firstly, it had to be examined whether the two researchers' estimations were statistically different. It was important to show that these estimations were free from the researchers' opinions. This means that any researcher would have a good chance to show the same results if the experiment was repeated. Thus, a contingency table was created. The 2 groups were the 2 researchers and the outcomes were the agreement and the disagreement with the FaceReader (Table 1). Pearson's Chi square was calculated in order to show the independence between the two groups. Chi squared equals 2.329 with 1 degree of freedom. The two-tailed P value equals 0.1270. Thus, the difference between the two researchers is not considered to be statistically significant.

Secondly, for the 172 students, we recorded 7416 different emotional states given by the FaceReader. Table 2 shows the results for each emotional state. Researchers and FaceReader had almost the same opinion regarding Neutral (99%) and Happy (90%) emotions. Moreover, Researchers and FaceReader had high agreement for Scared (87%), Surprise (82%) and Sad (79%) emotions. However, the agreement results were lower regarding Disgusted (70%) and Angry (71%) emotions. Nevertheless, there was a high agreement overall between the emotion measured by the FaceReader and the researchers' opinions.

| Groups | Agreement | Disagreement | Total |
|--------------|-----------|--------------|-------|
| Researcher 1 | 2910 | 415 | 3325 |
| Researcher 2 | 3530 | 561 | 4091 |
| Total | 6440 | 976 | 7416 |

| Table | 1. | Contingency | Table |
|-------|----|-------------|-------|
|-------|----|-------------|-------|

| Emotion | FaceReader and researcher agreement | Records for each emotion statement | Percentage |
|-----------|---|--|------------|
| Disgusted | 295 | 421 | 70% |
| Surprised | 215 | 262 | 82% |
| Neutral | 3561 | 3607 | 99% |
| Нарру | 263 | 292 | 90% |
| Angry | 1325 | 1870 | 71% |
| Scared | 195 | 223 | 87% |
| Sad | 586 | 741 | 79% |
| Total | 6440 | 7416 | 87% |

 Table 2. FaceReader and Researcher agreement for various emotional states.

| Emotion | FaceReader and researcher agreement | Records for each emotion statement | Percentage |
|------------------|--|---|------------|
| Disgusted male | 131 | 198 | 66% |
| Disgusted female | 164 | 223 | 73% |
| Surprised male | 82 | 93 | 88% |
| Surprised female | 133 | 169 | 78% |
| Neutral male | 1196 | 1205 | 99% |
| Neutral female | 2365 | 2402 | 98% |
| Happy male | 68 | 73 | 93% |
| Happy female | 195 | 219 | 89% |
| Angry male | 563 | 779 | 72% |
| Angry female | 762 | 1091 | 70% |
| Scared male | 62 | 63 | 98% |
| Scared female | 133 | 160 | 83% |
| Sad male | 200 | 272 | 74% |
| Sad female | 386 | 469 | 82% |
| Total male | 2302 | 2683 | 86% |
| Total female | 4138 | 4733 | 87% |

Table 3. FaceReader and Researcher agreement for various emotional states observed regarding each gender.

Moreover, Table 3 shows the agreement between researchers and FaceReader for emotional states observed for each gender. From 172 students, 60 were male (35%) and 112 were female (65%). This sample is large enough for gender differences to be studied. For Neutral, Happy and Angry emotions, FaceReader showed almost the same results in both genders. For Surprised and Scared emotions FaceReader showed better results regarding males than females. Finally, for Disgusted and Sad emotions,

FaceReader showed better results regarding females than males. Gender differences, concerning FaceReader performance, were observed in 4 out of 7 emotional states. Interpreting these differences is not part of this work. However, we plan to discuss these differences in another work in the near future.

In order to obtain the confidence interval for the agreement between researchers' opinions and FaceReader, a binomial proportion confidence interval was used (Table 4). The Adjusted Wald interval provides the best coverage for a specified interval.

DISCUSSION & CONCLUSIONS

Disgusted and Angry were the two emotions that FaceReader recognized less effectively. Examining the results revealed that Disgusted and Angry co-appeared frequently. Most of the times FaceReader measured simultaneously these two emotions, the researchers agreed only with the presence of an Angry emotion. Some movements of jaw, mouth and nose confused the FaceReader accuracy. Additionally, manv times FaceReader measured an Angry emotion simultaneously with a Neutral one, but neutral was the only emotion confirmed by the researchers. This particular disagreement was expected. When participants read the questions, many of them had clouded brow. People are taking this facial something expression when reading with great concentration. Zaman and Shrimpto-Smith (2006) came up to the same result. This is the reason why FaceReader measured, so frequently, an Angry emotion at the same time with a Neutral one. Moreover, FaceReader faced problems with participants that wore glasses or had piercing. Other problems were caused by special characteristics of some persons like big noses, bushy brows, small eyes or chins. Another difficulty were fringes reaching down to eyebrows.

| Emotion | 95% | 95% | 95% |
|-----------|------------------------|-------------------------|---------------------------|
| | confidence interval | conf. interv., males | conf. interv., females |
| Disgusted | 78% - 85% | 59% - 72% | 67% - 79% |
| Surprised | 77% - 86% | 80% - 93% | 72% - 84% |
| Neutral | 98.3% - 99% | 98.5% -99.6% | 97.8% - 98.8% |
| Нарру | 86% - 93% | 86% - 97% | 84% - 93% |
| Angry | 69% - 73% | 69% - 75% | 67% -72% |
| Scared | 82% - 93% | 91% - 99% | 77% - 88% |
| Sad | 76% - 82% | 68% - 78% | 79% - 86% |
| Total | 86%-87,5% | 84.4% - 87% | 86.5% - 88% |

| Table 4. The overall and the genders' confidence Interval for |
|---|
| the six emotions plus Neutral. |

Hopefully, these problems may be confronted because FaceReader will be upgraded. VicarVision and Noldus Information Technology support that they classify features which are located outside the modelled area of the face (e.g. hair) or features which are poorly modelled wrinkles, tattoos, piercing and birthmarks. Moreover, they will add person identification to the system [3].

Generally, results showed that FaceReader is capable of measuring emotions with an efficacy of over 87% during a self-assessment test, and that it could be successfully integrated into a computer-aided learning system for the purpose of affect recognition.

An instrument like FaceReader is very crucial for the amelioration of computer-aided learning systems. Educators will have the opportunity to give better and more effective emotional feedbacks in learning, self-assessment or CAT (Computer Adaptive Testing) systems [4].

To conclude, to our best knowledge this is the first study that evaluated FaceReader during a self-assessment test. Besides the evaluation of FaceReader, this study provides useful results for the emotional states of students during self-assessment tests and learning procedures.

- 1. Bența K.-I., Cremene M., & Todica V. *Towards an Affective Aware Home*, M. Mokhtari et al. (Eds.): ICOST 2009, LNCS 5597 (2009), 74–81.
- Cohen, I., Sebe, N., Chen, L., Garg, A., & Huang, T. S. Facial expression recognition from video sequences: Temporal and static modelling. *Computer Vision and Image Understanding*, 91, 1-2 (2003), 160-187.
- Den Uyl, M. J., & van Kuilenburg, H. The FaceReader: Online facial expression recognition. In *Proceedings of MB* 2005, 589-590.
- 4. Economides, A. A. Adaptive feedback characteristics in CAT (Computer Adaptive Testing). *International Journal of Instructional Technology & Distance Learning*, 3, 8 (2006), 15-26.
- 5. Ekman, P. *Emotion in the human face* (2nd ed.). Cambridge, MA: Cambridge University Press, 1982.
- 6. Essa, I. A., & Pentland, A. P. Coding, analysis, interpretation and recognition of facial expressions. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 7 (1997), 757-763.
- Moridis, C. & Economides, A. A. Towards computeraided affective learning systems: a literature review. *Journal of Educational Computing Research*, 39, 4 (2008), 313-337.
- 8. Oliver, N., Pentland, A., & Berard, F. LAFTER: A realtime face and lips tracker with facial expression recognition. *Pattern Recognition*, *33*, 8 (2000), 1369-1382.

- 9. Partala, T., Surakka, V., & Vanhala, T. Real-time estimation of emotional experiences from facial expressions. *Interacting with Computers*, *18*, 2 (2006), 208-226.
- 10. Partala, T.,& Surakka, V. The effects of affective interventions in human-computer interaction. *Interacting with Computers, 16,* 2 (2004), 295-309.
- 11. Smith, E., Bartlett, M. S., & Movellan, J. Computer recognition of facial actions: A study of co-articulation effects. In *Proceedings of the eighth annual joint symposium on neural computation*, 2001.
- Truong K. P., Neerincx M. A., & Van Leeuwen D. A. Measuring spontaneous vocal and facial emotion expressions in real world environments. *Proceedings of MB* 2008, 170-171.
- Van Kuilenburg H., Den Uyl M.J., Israel M.L. & Ivan P. Advances in face and gesture analysis. Proceedings of MB 2008, 371-372.
- 14. Zaman B. & Shrimpton-Smith T. *The FaceReader: Measuring instant fun of use*. NordiCHI 2006, 457-460.

Comparison of Eye Tracking Systems with One and Three Cameras

Christer Ahlstrom christer.ahlstrom@vti.se **Tania Dukic** tania.dukic@vti.se

Swedish National Road and Transport Research Institute (VTI), 581 95 Linköping, Sweden

ABSTRACT

When using eye movements to determine the state of a car driver it is important that the eye tracker is robust, unobtrusive, inexpensive and fully automatic. The objectives of this study are to compare the performance of a one-camera system with a three-camera system and to investigate if the accuracy and availability of the onecamera system is sufficient to monitor driver state. Data from 53 subjects were evaluated and the results indicate that there is not much difference between a single-camera system and a multi-camera system as long as the driver is looking straight ahead. However, with more peripheral gaze directions, the larger coverage that is provided by the additional cameras works in favour of the multi-camera system.

Author Keywords

Eye tracker, availability, accuracy, precision.

ACM Classification Keywords

H.3.4 Systems and Software: Performance evaluation, H.1.2 User/Machine Systems: Human factors

INTRODUCTION

Eye movement tracking is an important tool when measuring behaviour in different domains such as neuroscience, psychology, industrial engineering and human factors, marketing/advertising, and Computer Science [1]. In early days, eye tracking technology was both invasive and obtrusive, thus limiting the areas of application. However, today it is possible to use portable eye trackers that allow real-time eye tracking in realistic environments [2].

Eye tracking is mainly used in research to discover new knowledge in the above mentioned domains, but technology

is now spreading to consumer products where it is of interest to monitor human state. One example is the automotive industry where eye movements can be used to measure driver behaviour in many different ways. For example, the gaze direction has been used to measure visual distraction based on where the driver is looking [3], and the fixation duration [4], the gaze concentration [5] and the saccade amplitude [4] has been found to vary with cognitive distraction. Further, different blink parameters such as the blink duration and the blink velocity have been used to recognise drowsy driving [6]. Prerequisites for such consumer oriented products are that they are inexpensive, non-obtrusive, fully automated and highly robust. This means that it is necessary to use remote cameras (in contrast to helmet mounted eye trackers), and to lower the cost by using as few cameras as possible.

The objective of this paper is to compare two remote eye tracking systems mounted in the cockpit of a car. In particular, performance of the two systems will be compared in order to determine when it is sufficient to use a one-camera system and when a several-camera system is required.

METHODOLOGY

The two eye tracking systems were evaluated based on their availability, accuracy and precision. Availability (or tracking ratio) is defined as the total amount of logged gaze data divided by the maximum number of gaze data during some time period. This measure is related to the robustness of the system, but it says nothing about its accuracy. Accuracy is defined as the angular difference between the mean of all gaze log entries for a specific gaze target and the corresponding true gaze direction. This means that accuracy is related to the robustness of the system, but it says nothing about the robustness of the system. Finally, precision is defined as the standard deviation in angle between the tagged gaze log entries for a specific gaze target.

Participants

53 participants aged between 30 to 60 years, 50% males and 50% females, were included in the evaluation. The participants should have normal visual acuity without

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

wearing glasses (contact lenses were allowed), facial piercings were prohibited (earrings were accepted), different hair lengths were allowed (participants with long hair were asked to put their hair up) and the body length should lie between 155 and 193 cm (to fit the range of the camera).

Measurement Setup

The one-camera system used in the study was the Smart Eye AntiSleep system (Smart Eye AB, Gothenburg, Sweden). The system uses a single standard camera of VGA resolution together with IR flash illuminators. The IR illuminators and filters are tuned to frequencies with minimum interference of outdoor light, making the system robust to all natural illumination conditions in automotive applications. Smart Eye AntiSleep measures the driver's head position and orientation, gaze direction and eyelid opening at a rate of 60 Hz. The three-camera system used is the Smart Eye Pro (Smart Eye AB, Gothenburg, Sweden). The system has similar properties as the one-camera system, but also facilitates gaze direction in full 3D.

The data collection took place in a driving simulator at Saab AB in Trollhättan, Sweden. 41 dots at predetermined fixed locations were used as gaze targets, 14 in the cockpit of the car (Figure 1) and 27 outside the vehicle. The dots light up for 2 seconds so that the participant knows where to look. Each participant performed the experiment with and without eyeglasses. The eyeglasses had rather heavy frames and were the same model for each participant. The order of eyeglasses use (with or without mock-up eyeglasses) was randomized for all participants as well as for the sequence of the dots.

RESULTS

Out of the 53 participants, four were excluded from further analysis due to technical problems during data collection. Additional issues that were excluded from further analyses include incomplete fixation during the 2 seconds that the



Figure 1. The layout of the LED inside the cockpit (small dots: LED; large dots: eye tracking camera location for three-camera system; square: eye tracking camera location for one-camera system).

gaze target was active, erroneous camera calibration and bad initialization.

Availability

Availability per participant is presented in Figure 2. It can be seen that availability is really low for several participants. Availability is about 20 percent better for the three-camera system, something that is directly related to the coverage of the cameras (a single camera cannot see the eyes when the participant is looking at several of the peripheral dots).

An ANOVA analysis (p = 0.05) reveals that availability is affected by the location of the current dot and the age and gender of the participant. There is also an interaction between the angular distance of the current dot and the eye tracking system that is used, where availability decreases more with distance for the one-camera system as compared to the three-camera system. There are no significant



Figure 2. Availability per participant.

differences due to the eye glasses. Distance is defined as the angular distance from the active gaze target to the central gaze target that is located in front of the driver.

Availability results when dividing the gaze targets in three regions are presented in Table 1. The regions are defined by ellipses with horizontal radius r and vertical radius 2r/3, where $r \le 20^{\circ}$ for the central region, $20^{\circ} < r \le 120^{\circ}$ for the middle region and $r > 120^{\circ}$ for the outer region.

Accuracy

ANOVA results related to accuracy shows that the main differences in accuracy are due to the eye tracking system that is used and the distance from the central region. Figure 3 shows 2D maps of accuracy and availability for the onecamera and the three-camera system. The three-camera system has high accuracy and availability (coloured in blue) in a larger area compared to the one-camera system. The maps are constructed by interpolation between the measured availability and accuracy values in the 41 dots.

Precision

Figure 4 illustrates the precision and accuracy of the onecamera versus the three-camera system when the participants are not wearing glasses. In the central area the precision (ellipses) and accuracy (lines) are similar between the two systems, but the three-camera system performs better in peripheral areas. No systematic changes in precision due to glasses versus no glasses could be found.

DISCUSSION

The motive for this study was to evaluate eye tracking systems based on one or three cameras in terms of accuracy, availability and precision. The results show that both accuracy and availability decreases with the distance from the central region, and that the decrease is larger for one camera as compared to three cameras.

According to the statistical tests, both accuracy and availability deteriorates with distance from the central gaze target. There are also clear trends that this deterioration is larger for the one-camera system as compared to the threecamera system, and the same trends can be seen for precision. This means that with more cameras, you obtain higher accuracy, availability and precision over a larger area. There was also one unexpected finding, namely that eyeglasses did not significantly affect the tracking performance. This is a result which we cannot explain.

There are a few peculiarities related to the three performance indicators that are used to assess the eye tracking systems. The most important thing to remember is



Figure 3. 2D-maps of accuracy and availability encoded by color. The top four subplots illustrates when the participants are not wearing glasses and the bottom four subplots illustrates when they are wearing glasses.

that the three indicators should never be interpreted individually. Availability in itself can be misleading since it is not affected by where the participant is looking. This means that if a participant is looking straight ahead, where the eye tracking system has better performance, instead of at a peripheral gaze target where he/she is supposed to look according to the experimental protocol, this peripheral gaze target will get very high marks on availability. Similarly, accuracy (and precision) is calculated based on data with a certain quality. This means that only high quality data is used in the actual calculations. If the system is unable to measure the gaze direction for most of the time, the calculations will be very unreliable. It would have been possible to punish such unreliable data with a weighting function, but this was not done in this study.

| | | No glasses | | | | Glasses | | | |
|---------------------|------|------------|--------|-------|------|---------|--------|-------|--|
| | All | Centre | Middle | Outer | All | Centre | Middle | Outer | |
| One-camera system | 53.9 | 84.0 | 63.8 | 5.0 | 52.2 | 86.9 | 60.8 | 2.1 | |
| Three-camera system | 71.6 | 84.1 | 82.1 | 37.2 | 69.6 | 85.7 | 80.7 | 30.7 | |

Table 1. Overall availability (mean values) in different regions based on the distance from straight ahead.



Figure 4. Precision and accuracy of the measurements for each dot, where the ellipses indicate the average standard deviation across participants and the lines indicate the accuracy (distance from the reference point). Results are omitted in cases where the precision calculations are based on less than 15 % of the data.

Limitations to this study include that only eye trackers from one manufacturer were included in the evaluation and that the tests were conducted in a static setting in a simulator as compared to a dynamic real-world environment.

Thresholds defining when availability, accuracy and precision levels are acceptable are hard to provide. In general, algorithms for sleepiness warnings only require high quality data in the central region. This could also be sufficient for visual distraction algorithms, but then it is very important that the system only delivers data when tracking is available since the warning system will implicitly assume that lost tracking means that the driver is not looking in the central region.

CONCLUSION

Advantages with a one-camera system are that it is cheaper, easier to operate and easier to install in a vehicle. A multicamera system will, on the other hand, provide higher availability and accuracy for areas that are far from the road centre. A one-camera system is thus mostly suitable for invehicle applications such as systems that warn drivers for sleepiness or internal distraction while multi-camera solutions are preferable for applications such as visual distraction where it is necessary to know where the driver is looking.

ACKNOWLEDGEMENTS

We would like to thank VIP - Virtual Prototyping and Assessment by Simulation for financial support. We would

also like to thank our industrial partners in this project, Saab Automobile AB, SmartEye AB and Pixcode AB.

- 1. Duchowski, A.T. A breadth-first survey of eye-tracking applications. *Behavior Research Methods Instruments & Computers*, *34*, 4 (2002), 455-470.
- Hayhoe, M. and Ballard, D. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9, 4 (2005), 188-194.
- 3. Kircher, K. and Ahlstrom, C. Issues related to the driver distraction detection algorithm AttenD, in *1st International Conference on Driver Distraction and Inattention* (2009).
- 4. Reyes, M.L. and Lee, J.D. Effects of cognitive load presence and duration on driver eye movements and event detection performance. *Transportation Research* Part F, 11 (2008), 391-402.
- 5. Victor, T.W., Harbluk, J.L., and Engström, J. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8, 2 (2005), 167-190.
- Liu, C.C., Hosking, S.G., and Lenne, M.G. Predicting driver drowsiness using vehicle measures: recent insights and future challenges. J. Safety Res., 40, 4 (2009), 239-245.

Gaze Metrics for Efficient and Safe Operations of Hemodialysis

Hirotaka Aoki

Tokyo Institute of Technology, Japan 2-12-1 Oh-okayama, Meguro-ku, Tokyo 152-8552 aoki.h.ad@m.titech.ac.jp

ABSTRACT

This paper presents an application of eye tracking approach to cognitive task analyses of hemodialysis operation conducted at real working places in a Japanese hospital. In the study a series of observations was performed, where three clinical engineers' gazes as well as their behavior was recorded for five consecutive months. Two of the engineers were novices who had just begun their work at the hospital a few days before the observation started. Another engineer was an expert who had more than sixteen years of experience.

As a framework of cognitive task analysis, we developed a six gaze metrics that are closely connected with the cognitive aspects in efficient and safe operations of the hemodialysis. Trends of each novice engineer's information acquisition along with learning during a set-up task that is a part of hemodialysis operation were analyzed on the basis of the six gaze metrics. Several issues/implications on the novice engineers' adaptation to the hemodialysis operation task was discussed after the study.

Author Keywords

Gaze metrics, eye tracking, hemodialysis machine, learning processes.

ACM Classification Keywords

H.5.2 User interfaces

INTRODUCTION

Clinical engineers performing hemodialysis operation shift their gazes between many areas like small indicators, controllers, tubes and so forth as well as change in medical staffs. It can be expected that understanding of when and why such shift of gazes occurs provides great insights about

Satoshi Suzuki

Tokyo Women's Medical University, Japan 8-1 Kawada-cho, Shinjuku-ku, Tokyo 162-8666 ssuzuki@kc.twmu.ac.jp

skills or knowledge that are needed to carry out such tasks both efficiently and safely.

In the medical domain, some researchers have argued that the eye tracking approach involves great potentials to reveal skills and know-how (e.g., [1-4, 9, 10]). However, the application of eye tracking to issues in real environment is still rare. There are three primary reasons for this: First, recording eye tracking data in real-world environments seems to have some potential for adverse influence on performance. Caused by this, most of hospitals or clinics are reluctant to adopt eye tracking recording in their working places. Second, analyses of eye tracking data are basically labor-intensive (e.g., [8]). Such analyses require very time-consuming in-depth cataloging of fixations, fixation durations, fixated areas, and so forth from video recordings. Thirdly, though interpretation of eye-tracking data is of great difficulty, there are very few analysis methodologies that can be widely used for knowledgeelicitation or behavior analysis in various domains. This means that every investigator has to develop appropriate analysis framework unaided in order to interpret data generated in his/her research.

In this paper, we describe a pilot study performed in a hemodialysis operation room on the behavior of clinical engineers during the set-up of hemodialysis machines. We developed gaze metrics representing clinical engineers' attention allocation during information acquisition, and applied the metrics to the cognitive task analysis of novice engineers' learning in the set-up tasks. The gaze metrics as well as the underlying idea proposed in this paper implies the principle that guides how eye tracking data in some cognitive tasks should be processed.

TASK FOCUSED IN OUR STUDY

To perform hemodialysis, a clinical engineer is firstly required to do a set-up task of the hemodialysis machine. In this set-up task, each engineer has to install the appropriate tubes, open a valve in venous side in order to do priming, and check the current status of the machine before starting hemodialysis (see Figure 1). Set-up task is recognized as one of the most critical tasks in whole hemodialysis operation since the errors committed in this task may cause serious injuries to the patient.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Hemodialysis machine.

GAZE METRICS

For analysis of eye tracking data during hemodialysis operation, we adopted the gaze-based metrics shown in Table 1. As explained in the table, each metric is directly connected to the aspects relating to efficient and safe information acquisition during the set-up task. The detailed explanation about each metric is described below.

| Aspect for evaluation | Corresponding gaze metrics used in this study | | | | | | |
|--|--|--|--|--|--|--|--|
| 1. Information acquisition from all the task- | | | | | | | |
| related information 1-1. How much did the engineer try to acquire the information? 1-2. How long did the engineer spend time for each information? 1-3. How redundantly did the engineer inspect information? | → Number of gazed AOIs → Mean fixation duration per AOI er → Mean number of fixations per gazed AOI | | | | | | |
| Information acquisition from key information that is critical in the current 2-1. Did the engineer inspect all the key information? 2-2. Did the engineer select the key information efficiently? 2-3. Did the engineer inspect key information redundantly? | task → Coverage rate of key AOI → Rate of key AOI included in gazed AOIs → Mean number offixationsperkeyAOI | | | | | | |

Table 1. Gaze metrics.

Gaze Metrics for Information Acquisition Processes from All the Task-Related Information

We firstly focus on engineers' information acquisition from all task-related information. In the hemodialysis machine, we could identify 24 areas of interests (AOIs), all of which seemed to be related to the set-up task. Based on the AOIs, we calculated the following metrics:

Number of gazed AOI (area of interest). This metric is concerned with a question how much information an engineer tries to acquire within a given task. This can be obtained as net number of all of gazed AOIs within the setup task. Since 24 AOIs were identified in the hemodialysis machine, the maximum numerical value of this metric is 24 in this study. Figure 2 shows a schematic example of the scan-paths of two engineers. Two out of five AOIs are identified as key AOIs, which will be explained below, in this example. In this example, the metric for engineers A and B are obtained as 5 and 3, respectively.

Mean fixation duration per gazed AOI (area of interest). This represents how long an engineer spends time in order to process each AOI's information. This metric can be



Figure 2. Schematic example of scan-paths among AOIs.

obtained as an average of a single fixation's duration over all the gazed AOIs.

Mean number of fixations per gazed AOI (area of interest). This metric reflects our expectancy that reliable activity can be ensured by multiple inspections, and not only by a single inspection on required information. This metric indicates how many times a gazed AOI was fixated in average. In the example shown in Figure 2, the metric for engineers A and B are obtained by (1+4+1+1+1)/5=1.60 and (1+2+2)/3=1.67, respectively.

Gaze Metrics for Information Acquisition Processes from Key Information

Secondly, we focus on the information acquisition processes from critical information (in this paper named "key information") that are necessary at the moment. The procedure for identifying such critical information (i.e., "key AOI") will be presented in the following chapter.

Coverage rate of key AOI. This metric is relating to how much critical information is successfully gazed by an engineer during the task. This is calculated as the number of key AOIs that are actually gazed divided by the total number of key AOIs. In the example shown in Figure 2, this metric for both of engineers A and B is 1.0.

Rate of key AOI included in gazed AOIs. This metric derives from a question how efficiently an engineer can select the key AOIs. This metric is calculated as the number of gazed key AOI by a specific engineer divided by the number of gazed key AOI. In Figure 2, this metric for engineers A and B are calculated as 2/5=0.4 and 2/3=0.67, respectively.

Mean number of fixations per key AOI. The idea of this metric corresponds to that of the Mean number of fixations per gazed AOI. This metric indicates how many times a key AOI was fixated in average. In Figure 2, this metric for engineers A and B are calculated as (4+1)/2=2.5 and (2+2)/2=2.0, respectively.

DATA COLLECTION

Three clinical engineers working at a Japanese hospital was observed during this study. Two engineers (E1 and E2) had just started to work just after their graduation from



started to work

Figure 3. Observation procedure.



Figure 4. An engineer participating in our observation.

university when the observations began. Both of their age was 23 with less than 1 month experiences of the hemodialysis operation. The last engineer was an expert who had more than 16 years experiences of the operation. Figure 3 illustrates the observation procedure. As shown in the figure, we performed the observations for five consecutive months (see Figure 3).

We observed the set-up task in real hemodialysis operation. Before the task, we asked engineers to take a SMI eye tracking system and to perform calibration. The SMI's system was determined to have a spatial accuracy of 0.5-1.0°, and the sampling frequency was 60Hz. We repeated the calibration if the calibration was off by approximately more than 2.0°. The calibration procedure took around 5-10 minutes in total for each engineer. Then we started to record their eye movement as well as video recording of their behavior (see Figure 4). As explained before, most of Japanese hospitals, in general, are reluctant to adopt eye tracking recording in their working places. This observation could be carried out by a special permission given from people responsible for the operation room, who has a great understanding of research in Human Factors area. To perform the observation smoothly, we explained about the eve trackers to patients in advance, which allowed us to avoid undesired conflicts with them.

RESULTS

Before analyzing the two engineers' (E1 and E2) data, we firstly identified what AOIs were gazed by the expert engineer. According to previous works, the use of eye movement as expert's skill information is recognized as very effective in training in the field of aircraft inspection [7], assembly line [5], and so forth. In such application, an expert's eye movement is taken as an ideal eye-gaze sequence. Considering this, we determined that the AOIs gazed by the expert engineer are "key AOIs" for each of three subtasks (i.e., install tubes to appropriate areas, to open a valve in venous side, and to check all the current status of the machine).

Transitions of the gaze metrics for information acquisition processes from all task-related information and from key information with days are shown in Figures 5 and 6, respectively. As for number of gazed AOI, show for both engineers (E1 and E2) a decreasing tendency along with their experience increases (see Figure 5 (a)). Especially, E2 seems to succeed in reducing unnecessary gazes within less than 40-day experience, though E1 tends to look at too many AOIs in the identical period. From time needed to process information view, both engineers do not seem to be acquainted with processing information efficiently within the observation period (see Figure 5 (b)). As found in Figure 5 (c), E1 gazes at each AOI more frequently, whereas E2 seems to become stable (approx. 1.0) by 40-day experience. The results mentioned above may indicate that the E2 could adapt well to the set-up task compared with E1. The results may also indicate that some follow-up training for E1 could be necessary, with close inspection of the setup task in order to compensate for E1's lacking skills.

As for metric "Coverage rate of key AOI", both engineers do not show improvement in the observation period. The mean values of the metric for E1 and E2 are 0.399 and 0.343, respectively (see Figure 6 (a)). At the very early learning stage, the efficiency in selecting key AOIs represented by "Rate of key AOI included in gazed AOI" seems to be improved by E1 (see Figure 6 (b)). Along with the 80-120 days experience, both of engineers seem to show stable values (around 0.6-0.8) in this metric. As seen in Figure 6 (c), E1 gazes at key AOIs very frequently at the early period in learning, and shows sudden drop just after







Figure 6. Transitions of gaze metrics (key information).

20-day experience. E2 on the other hand, shows relatively stable values in the same period. Mean values of the metric from 40 to 140 day of experience for E1 and E2 1.79 and 1.17, respectively. Among the results mentioned above, we think that the tendencies found in the metric "Mean number of fixations per key AOI" are relatively problematic. The values of the metric having less than 2.0 indicate that the most of key AOI are likely to be inspected less than twice. This implies that opportunity of revision of the inspection error may be lost when some error occurs at an initial inspection of a key AOI. From this result, we suggest that the novice engineers should be instructed to perform multiple inspections to key AOIs. In addition, the effect of the instruction should be evaluated by applying our metric "Mean number of fixations per key AOI".

CONCLUSION

In the present paper, we propose a six gaze metrics and a calculation procedure for analyzing hemodialysis operations. The metrics were closely connected with questions that arose from cognitive aspects in efficient and safe operations. We think that the gaze metrics as well as the underlying idea proposed can be applied to other tasks classified into "well-defined" cognitive tasks, meaning that the structure of task such as task goal, constraints, working procedure and so forth can be described in advance. The proposed metrics were applied to analysis of the two novice clinical engineers' learning processes in the set-up tasks included in the hemodialysis operation. Examining the novice engineers' performance elicited by the metrics, we could identify differences of tendencies in information acquisition processes with learning, especially the taskrelated information processes were notable.

- 1. Aoki, H. Itoh, K. Tomizawa, Y., Matayoshi, T. and Yozu, R. Application of eye tracking approach to medical domain -Eye-tracking analysis of skilled performance in pump-oxygenator artificial heart-lung machine operation-. Presented at *15th European Conference on Eye Movement*, 2009.
- 2. Krupinski, E. A., and Roehrig, H. Pulmonary nodule detection and visual search: P45 and P104 monochrome versus color monitor displays. *Academic Radiology 9*, 6 (2002), 638-645.
- 3. Law, B., Atkins, M.S., Kirkpatrick, A.E., and Lomax, A.J. Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proc. Eye Tracking Research and Application Symposium 2004*, ACM Press (2004), 41-48.
- 4. Mello-Thoms, C., Dunn, S., Nodine, C.F., Kundel, H.L., and Weinstein, S.P. The perception of breast cancer: What differentiates missed from reported cancers in mammography? *Academic Radiology* 9, 9 (2002), 1004-1012.
- Nagamatsu, T., Kaieda, Y., Kamahara, J., and Shimada, H. Development of a skill acquisition support system using expert's eye movement. In *Proc. HCI International 2007*, Splinger-Verlag (2007), 430-439.
- Reingold, E.M., Charness, N. Pomplun, M., and Stampe, D.M. Visual span in expert chess players: Evidence from eye movements. *Psychological Science* 12, 1 (2001), 48-55.
- Sadasivan, S., Greenstein, J.S., Gramopadhye, A.K., and Duchowski, A.T. Use of eye movements as feedforward training for a synthetic aircraft inspection task. In *Proc. CHI* 2005, ACM Press (2005), 142-149.
- Seagull, F.J., Xiao, Y., MacKenzie, C.F., Jaberi, M., and Dutton, R.P. Monitoring behavior: A pilot study using an ambulatory eye-tracker in surgical operating rooms. In *Proc. 1999 Human Factors and Ergonomics Society Annual Meeting*, Human Factors and Ergonomics Society (1999), 850-854.
- Segall, N. Taekman, J.M., Mark, J.B., Hobbs, G., and Wright, M.C. Coding and visualizing eye tracking data in simulated anesthesia care. In *Proc. 2007 Human Factors and Ergonomics Society Annual Meeting*, Human Factors and Ergonomics Society (2007), 765-769.
- 10. Tchalenko, J., Dempere-Marco, L., Hu, X.P., and Yang, G.Z. Quantitative analysis of eye control in surgical skill assessment. Presented at *11th European Conference on Eye Movements*, 2001.

Automatic Recognition of Lower Facial Action Units

Isabel Gonzalez igonzale@etro.vub.ac.be Hichem Sahli hichem.sahli@etro.vub.ac.be Werner Verhelst wverhels@etro.vub.ac.be

Joint Research Group on Audio Visual Signal Processing (AVSP), Vrije Universiteit Brussel, Department ETRO, Pleinlaan 2, 1050 Brussels

ABSTRACT

The face is an important source of information in multimodal communication. Facial expressions are generated by contractions of facial muscles, which lead to subtle changes in the area of the eyelids, eye brows, nose, lips and skin texture, often revealed by wrinkles and bulges. To measure these subtle changes, Ekman et al. [5] developed the Facial Action Coding System (FACS). FACS is a human-observer-based system designed to detect subtle changes in facial features, and describes facial expressions by action units (AUs). We present a technique to automatically recognize lower facial Action Units, independently from one another. Even though we do not explicitly take into account AU combinations, thereby making the classification process harder, an average F_1 score of 94.83% is achieved.

Author Keywords

Facial Action Units, AdaBoost, OVL, SVM.

INTRODUCTION

The Facial Action Coding System (FACS) [5] is a humanobserver-based system designed to detect subtle changes in facial features, and describes facial expressions by 44 anatomically based Action Units (AUs). AUs can occur individually or in combinations. When AUs occur in combination they may be nonadditive, in which case the combination changes the appearance of the constituents.

FACS coding is very labor-intensive; automating this process has attracted a lot of attention since the early 1990s. Automatic facial analysis requires extraction of features from static images or image sequences, and classification into AUs, or AU combinations. Basically, there are two main approaches for feature extraction: geometry-, and appearance-based. Geometry-based feature extraction consists of detecting and tracking facial feature

points. Appearance-based features concern motion and texture changes. Methods to extract these features are Gabor Wavelets Analysis, optical flow analysis, PCA, ICA.

It has been shown that the combination of geometry- and appearance-based features give better recognition results, [4, 10, 13].

State of the art classification methods are Support Vector Machines (SVM) [1, 10], Neural Networks, Hidden Markov Models, and Discriminant Analysis.

FEATURE EXTRACTION

Our features are geometry, and consist of a shape model of 83 facial feature points [6], see Figure 1. Once we have the shape model for all video frames, we apply an affine transformation by warping each image to a common view.

Our initial set of features is the set of all facial feature points $p_i = (x_i, y_i)$, with i = 1... 83. Next, we calculate for all points p_i , the displacements of the facial feature points from the current frame *t* relative to those in a neutral frame 0:

$$f(p_i)_t = || p_{i,t} - p_{i,0} ||$$

The last set of features consists of several features calculated from the facial feature points. The equations are constructed from the linguistic description of the FACS manual, and from state of the art work [8, 13]. Table 1 shows these equations; Table 2 presents the description of auxiliary points derived from the shape model. $d(p_b p_j)$ denotes the Euclidean distance between points p_i and p_j , while $dv(p_b p_j)$ denotes the vertical distance between the



Figure 1. Shape Model, © J.F. Cohn

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.
| Id | Equation | Ref |
|----|--|------|
| 1 | $(dv(52,15)_t + dv(52,7)_t)/2 - (dv(52,15)_0 + dv(52,7)_0)/2$ | [13] |
| 2 | $d(52,O)_t - d(52,O)_0$ | [13] |
| 3 | $d(52,58)_t - d(52,58)_0$ | [13] |
| 4 | $d(44,41)_t - d(44,41)_0$ | (*) |
| 5 | $(d(A,49)_t + d(A,55)_t) - (d(A,49)_0 + d(A,55)_0)$ | [8] |
| 6 | $ d(55,9)_t - d(55,9)_0 $ | [13] |
| 7 | $ d(49,1)_t - d(49,1)_0 $ | [13] |
| 8 | $d(55,49)_t - d(55,49)_0$ | [13] |
| 9 | $ \begin{array}{rl} (dv(15,29)_t + dv(7,19)_t)/2 - & (dv(15,29)_0 + \\ & dv(7,19)_0)/2 \end{array} $ | [13] |
| 10 | $(d(9,27)_t + d(1,17)_t)/2 - (d(9,27)_0 + d(1,17)_0)/2$ | [13] |
| 11 | $d(J,J')_t - d(J,J')_0$ | [13] |
| 12 | $ \begin{array}{rl} (d(15,11)_t + d(7,3)_t)/2 &- & (d(15,11)_0 + \\ & d(7,3)_0)/2 \end{array} $ | (*) |
| 13 | $d(63,52)_t - d(63,52)_0$ | (*) |
| 14 | $d(62,51)_t - d(62,51)_0$ | (*) |
| 15 | $d(64,53)_t - d(64,53)_0$ | (*) |
| 16 | $d(58,67)_t - d(58,67)_0$ | (*) |
| 17 | $d(57,66)_t - d(57,66)_0$ | (*) |
| 18 | $d(59,68)_t - d(59,68)_0$ | (*) |
| 19 | $d(B,B')_t - d(B,B')_0$ | (*) |
| 20 | $d(58,52)_t - d(55,49)_t$ | (*) |

Table 1. Equations formulated according to AU Description in FACS manual [5] by I. Gonzalez and P. Fan, ETRO, VUB (*); by Zhang [13], and el Kaliouby [8].

two points:

$$dv (p_i, p_j)_t = |y_{i;t} - y_{j;t}|$$

Equations (1) to (19) are estimates on the current frame t relative to the neutral frame 0, and compute the following: (1) the average vertical distance between the upper lip center and the eyes; (2) the distance of the upper lip center to the center point of the nose; (3) the mouth height in the center of the upper and lower lip; (4) the width between the outer nostril wings; (5) the width of the mouth, from left corner to center and from right corner to center; (6) and (7), the distance from the mouth corner to the inner eye corner, right and left respectively; (8) the mouth width, from left to right corner; (9) the average vertical distance between lower eye lid center and upper brow center; (10) the average distance between the inner brows; (11) the distance between the inner brows; (12) the

parts of the lips, in all points; and (19) the mean height of the mouth. Finally, (20) computes the difference between the height and the width of the mouth. All feature sets, i.e. the facial feature points, the displacements, the other geometric features of Table 1, and a combination thereof, are evaluated in the experiments.

CLASSIFICATION

We compared classification by AdaBoost and SVM. AdaBoost (Adaptive Boosting) is a machine learning algorithm [7] that constructs a strong classifier as a linear combination of weak classifiers. AdaBoost was also used as a feature selection technique. For the feature selection, we used decision stumps as weak classifiers. The feature that minimizes the classification error on the weighted samples is chosen in each iteration.

A Support Vector Machine (SVM) is a supervised method used for classification and regression, and has a wellfounded mathematical theory [12]. Through training, a SVM builds a hyper plane, or set of hyper planes, in a high or infinite dimensional feature space which separates two or more classes. The maximum-margin hyper plane is the hyper plane that maximizes the distance from it to the nearest data point on each side. The larger the margin is, the lower the generalization error of the classifier.

EXPERIMENTS

Our system was trained on the Cohn and Kanade's DFAT-504 dataset [9]. This database consists of 486 sequences of facial displays that are produced by 98 University students from 18 to 30 years old, of which 65% is female. All sequences are annotated by certified FACS coders. We selected all sequences where at least one of the analyzed AUs is present. This resulted in 364 sequences from 94 subjects (66% female). Training and evaluation is done on the last frame of each sequence, which contains the apex of the expression.

Our final goal is to develop a system that detects behavioral patterns in communication between two persons, more specifically between mother and infant. We therefore selected a set of AUs from state of the art work in mother-infant communication, e.g. [11]. As lower facial AUs do not involve lots of wrinkles and furrows, which need appearance-based features, we started with the analysis of the lower AUs. Table 3 shows the AUs we analyze.

| Point | Description | | |
|-------|--|--|--|
| А | Midpoint between two mouth corners [8] | | |
| 0 | Center point of nose [13] | | |
| J/J' | Right/Left Inner Brow point [13] | | |
| B/B' | Mean inner upper/lower lip point (*) | | |

Table 2. Auxiliary points used in Table 1.

| AU | Action | Description | |
|----|------------|-------------------|---|
| 12 | Oblique | Lip Corner Puller | V |
| 20 | Horizontal | Lip Stretcher | |
| 23 | Orbital | Lip Tightener | |
| 25 | Up/ Down | Lips Part | |
| 27 | Up/ Down | Mouth Stretch | |

Table 3. Action Units [5] with their description.

A binary classifier is constructed for each AU. Positive examples for a given AU are those samples that contain the specific AU, regardless of the occurrence of other AUs. Negative examples are the samples not containing that AU. Generalization to new subjects was tested using leave-one-subject-out cross validation, in which all sequences of the test subject are excluded from training. Performance was measured by the F_1 score. It considers both the precision p and the recall r to compute the score:

$$p = tp/(tp + fp)$$
$$r = tp/(tp + fn)$$

with tp the number of items correctly labeled as belonging to the positive class (true positives); fp the items incorrectly labeled as belonging to the positive class (false positives); and fn the items labeled as belonging to the negative class but which belong to the positive class (false negatives). The F_1 score can be interpreted as a weighted average of the precision and recall:

$$F_1 = 2*(p*r)/(p+r)$$

We experimented with different feature sets, and classifiers SVM, and AdaBoost. We combined feature selection with SVM classification. Table 4 gives an overview of the AUs with the number of positive and negative samples.

| AU | 12 | 20 | 23 | 25 | 27 |
|----|-----|-----|-----|-----|-----|
| Р | 111 | 68 | 42 | 291 | 75 |
| Ν | 253 | 296 | 322 | 73 | 289 |

 Table 4. Number of positive (P) and negative (P) samples for each AU.

SVM Versus AdaBoost

We first started with two state of the art classification techniques in AU recognition. SVMs were trained and evaluated using LibSVM [2], while AdaBoost is

| AU | F ₁ | Acc | Hit / FA |
|----|----------------|-------|------------|
| 12 | 95.37 | 97.25 | 92.79/0.79 |
| | 95.37 | 97.25 | 92.79/0.79 |
| 20 | 87.30 | 95.60 | 80.88/1.01 |
| | 87.88 | 95.60 | 85.29/2.03 |
| 23 | 92.68 | 98.35 | 90.48/0.62 |
| | 92.50 | 98.35 | 88.10/0.31 |
| 25 | 98.28 | 97.25 | 97.94/5.48 |
| | 98.80 | 98.08 | 98.97/5.48 |
| 27 | 97.99 | 99.18 | 97.33/0.35 |
| | 98.01 | 99.18 | 98.67/0.69 |

Table 5. Results for SVM, features selected by AdaBoost (first row), or OVL (second row). Results given by F_1 score, accuracy (Acc), Hit and False Alarm Rate (FA).

constructed using OpenCV. For SVM, we tested kernels such as Linear, and Radial Basis Functions (RBF). We experimented with different variants of boosting algorithms, such as Discrete, Real, Gentle AdaBoost, and LogitBoost. The results show that SVM outperforms AdaBoost for almost every AU. SVM achieves an average F_1 score of 92.77%, while AdaBoost has an average of 90.73%. We tried all feature sets, and the best results were always the geometric features of Table 1, or a combination of those, and the displacements of facial feature points.

Feature Selection and SVM Classification

The features used in these experiments consist of estimations on the entire shape model. Feature selection is performed by AdaBoost, or according to the overlapping coefficient (OVL). AdaBoost is a well-known technique for feature selection, and is used in state of the art AU recognition. Because we wanted to know how discriminative the features for each AU are, and how well AdaBoost was able to select these discriminative features, we also tried out selecting features according to the OVL coefficient. OVL [3] is defined to be the area intersected by graphing two probability density functions, in our case two normal distributions. So, if $f_1(x)$ and $f_2(x)$ are two probability density functions defined on the *n*-dimensional real numbers R^n , then the OVL is defined as:

$$OVL = \int_{R}^{n} min(f_{1}(x), f_{2}(x))dx$$

In our case, $f_1(x)$ and $f_2(x)$ are the distributions of positive, respectively negative samples for each feature. The lower the OVL coefficient, the more a feature is discriminative. We evaluate a SVM for each AU with the first *k* features selected by AdaBoost, or according to the OVL coefficient. We chose the number of features for which the SVM has the best performance. Table 5 gives an overview of the best results. OVL slightly outperforms AdaBoost. OVL achieves on average a F_1 score of 94.51%, whereas AdaBoost has an average of 94.32%. Taking the best technique for each AU, we achieve an average F_1 score of

| AU\Method | All feats | Low feats |
|------------------|-----------|-----------|
| 12 | 95.37 | 95.41 |
| 20 | 87.88 | 88.89 |
| 23 | 92.68 | 93.83 |
| 25 | 98.80 | 98.63 |
| 27 | 98.01 | 97.37 |

Table 6. F_I score for classification with facial features ofthe entire shape model (all feats) versus the lower facialfeatures (Low Feats).

94.55%. In general, the most discriminative features are also selected by AdaBoost, though not in the order given by the OVL coefficient. AUs that have no real discriminative features, such as AU20 (with lowest OVL of 0.51) benefit the most of doing feature selection according to the OVL coefficient. AUs with low OVL coefficients, such as AU25 and AU27, achieve the best results.

The Effect of Upper Facial Features

In the previous experiments, upper facial features were always included in the feature selection with the best results, so we wanted to know why, and what the effect was of excluding the upper facial features. For this experiment, we restart AdaBoost with only the features of the lower face. The OVL coefficient is independently calculated for each feature; we take only the lower facial features. On average, we get a slightly better performance using only these features, i.e. 94.83%. Using AdaBoost as feature selection technique slightly outperforms OVL, i.e. on average, we obtain 94.59% for AdaBoost and 94.16% for OVL. Removing upper facial features from the set, makes AdaBoost select different (lower facial) features in another order than when all the features are present. Table 6 shows the F_1 score of the best results using all features versus features from the lower part of the face only. For all AUs, when all features are present, features from the eyes and brows are included in the selection process. For AU12, AU20, and AU23, these features are not highly discriminative. AdaBoost does a better selection when these features are not present. For AU25 and AU27, these features are moderately discriminative, thereby a slightly worse performance is obtained when removing them. The moderate discriminativity of these features could be due to the high percentage of sequences that only include AU25 or AU25+AU27, without action in the upper facial region.

CONCLUSION AND FURTHER RESEARCH

We achieve an average F_1 score of 94.83%. Performing feature selection before classification improves the results. AdaBoost also selects the most discriminative features, though not in the order given by the OVL coefficient. The feature selection always includes upper facial features. Using only lower facial features gives on average a better performance. In our further research, we will continue to look for the best feature set for each AU, combining geometry- and appearance-based features. In future work, we will extend the AU list to the entire face, investigate the dynamics of the AUs, and analyze the results and possible differences when the technique is applied on posed and spontaneous expressions.

- Bartlett, M., Littlewort, G., Lainscsek, C., Fasel, I., and Movellan, J. Machine learning methods for fully automatic recognition of facial expressions and facial actions. In IEEE International Conference on Systems, Man & Cybernetics (2004), 592-597.
- Chang, C.-C. and Lin, C.-J. LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
- Clemons, T.E. and Bradley Jr., E.L. A nonparametric measure of the overlapping coefficient. Computational Statistics &. Data Analysis. 34, 1 (2000), 51-61.
- Cohn, J.F., Reed, L., Moriyama, T., Xiao, J., Schmidt, K., and Ambadar, Z. Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. In Proc. FG 2004, IEEE Computer Society (2004), 129-135.
- 5. Ekman, P., Friesen, W., and Hager, J.C. The Facial Action Coding System. Second Edition. Research Nexus eBook (2002).
- Hou, Y., Sahli, H., Ravyse, I., Zhang, Y., and Zhao, R. Robust Shape-Based Head Tracking. In Proc. ACIVS 2007, Springer-Verlag (2007), 340-351.
- Freund, Y., and Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55, 1 (1997), 119-139.
- el Kaliouby, R. Mind-Reading Machines: Automated Inference of Complex Mental States. PhD thesis, University of Cambridge, Computer Laboratory (2005).
- Kanade, T., Cohn, J.F., and Tian, Y. Comprehensive database for facial expression analysis. In Proc. FGR 2000, IEEE Computer Society (2000), 46-53.
- 10. Mahoor, M.H., Cadavid, S., Messinger, D.S., and Cohn, J.F. A framework for automated measurement of the intensity of non-posed facial action units. In Proc. CVPR4HB 2009, 12-18.
- Messinger, D., Mahoor, M., Chow, S., and Cohn, J.F. Automated Measurement of Facial Expression in Infant-Mother Interaction: A Pilot Study, Infancy, 14, 3 (2009), 285-305.
- 12. Vapnik, V.N. The nature of statistical learning theory. Second Edition, Springer-Verlag (1999).
- Zhang, Y.M. and Ji, Q.A. Active and dynamic information fusion for facial expression understanding from image sequences. PAMI, 27, 5 (2005), 699-714.

Facial Expression Classification Based on Local Spatiotemporal Edge and Texture Descriptors

Yulia Gizatdinova

Research Group for Emotions, Sociality and Computing, University of Tampere, Finland yulia.gizatdinova@cs.uta.fi Veikko Surakka Research Group for Emotions, Sociality and Computing, University of Tampere, Finland veikko.surakka@cs.uta.fi

Guoying Zhao

Machine Vision Group, University of Oulu, Finland guoying.zhao@ee.oulu.fi

Erno Mäkinen Multimodal Interaction Research Group,

University of Tampere, Finland erno.makinen@cs.uta.fi

ABSTRACT

Facial expressions are emotionally, socially and otherwise meaningful reflective signals in the face. Facial expressions play a critical role in human life, providing an important channel of nonverbal communication. Automation of the entire process of expression analysis can potentially facilitate human-computer interaction, making it to resemble mechanisms of human-human communication. In this paper, we present an ongoing research that aims at development of a novel spatiotemporal approach to expression classification in video. The novelty comes from a new facial representation that is based on local spatiotemporal feature descriptors. In particular, a combined dynamic edge and texture information is used for reliable description of both appearance and motion of the expression. Support vector machines are utilized to perform a final expression classification. The planned experiments will further systematically evaluate the performance of the developed method with several databases of complex facial expressions.

Author Keywords

Human behaviour understanding, expression classification, spatiotemporal descriptor, local oriented edge, local binary pattern, facial expression, action unit, emotion.

ACM Classification Keywords

I. Computing methodologies: I.4 Image processing and computer vision; I.2 Artificial intelligence: I.2.10 Vision

Roope Raisamo Multimodal Interaction Research Group, University of Tampere, Finland roope.raisamo@cs.uta.fi

and scene understanding.

INTRODUCTION

During the past few decades, automatic classification or recognition of facial expressions has attracted a considerable attention in computer vision, pattern recognition and human-computer interaction (HCI). The main reason for this has been a growing demand for the development of new generation of HCI interfaces that actively perceive a user, detect visual cues of facial behaviours and use this information to initiate interaction, offer help, or assist the user in performing a certain action or task. Generally, facial expressions have been classified either as emotion-associated facial displays or action units (AU). AUs represent momentary changes in facial appearance which are brought about by single or conjoint facial muscle activations [3]. Although the research in the area of automatic expression classification has seen a lot of progress [4,11,14], there are still challenges that need to be solved in order to achieve true applicability of expression classifiers in real-world situations. Thus, there is a need for improvement of the existing classification schemes in terms of their accuracy, speed and robustness to unconstrained environmental conditions. On the other hand, there is a challenge of reliable analysis of between- and within-person variations as facial expressions vary in their appearance both across human population and within facial behaviour of a given individual. Therefore, one of the essential prerequisites for a successful development of expression recognition systems is to find descriptive, robust yet fastand easy-to-compute facial representations.

So far, the majority of the existing research has been dealing with processing of static image data [4,11,14]. However, recent advances [13] have clearly demonstrated that humans recognize facial expressions better when the dynamics of the expression is taken into consideration. Following these findings, several spatiotemporal

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

approaches to expression recognition have been reported. Among others, the most frequently used approaches are hidden Markov models [1], dynamic texture descriptors [15], geometrical displacements [6] and dynamic Bayesian networks [16]. Dynamic texture descriptors combine appearance and motion features which occur at several spatiotemporal scales. Local binary pattern (LBP) texture descriptors [10] have been widely used to describe static textures and recently extended to temporal domain. LBP operator encodes various local primitives such as points, curved edges, spots, flat areas, etc. LBP descriptors have been used for facial representation and reported to provide a number of important advantages like local processing, multi-scale representation, robustness to monotonic greyscale changes and simple computation [15]. LBP descriptors, which are calculated on three orthogonal planes (TOP) of the image sequence, result in LBP-TOP dynamic texture descriptors and have been used to recognize a limited set of posted emotion-associated facial expressions [15]. The reported classification performance has approached the classification performance of a human observer with more than 90% of classification success. Recently, several attempts have been done in order to enhance LBP-TOP operator by applying it to gradient [9] or Gabor-decomposed [7,8] images.

In this paper, we present an ongoing research that aims at development of a novel spatiotemporal approach to facial expression classification in video. The novelty comes from a new facial representation that is based on local spatiotemporal feature descriptors. In particular, a combined dynamic edge and texture information is used for interpretation of the appearance and dynamics of facial behaviour. Differently from previous studies, we aim to enhance LBP-TOP descriptor by combining it with a dynamic edge descriptor based on local oriented edges (LOE) [5]. LOE descriptors encode local edges of multiple orientations at different resolution levels. In the earlier studies [5], LOE descriptors have been successfully applied to the task of facial feature localization from images with complex facial expressions. Therefore, dynamic LOE-TOP descriptors are expected to improve the descriptive property of LBP-TOP operator.

In the proposed method, LOE- and LBP-TOP feature vectors (histograms) are calculated in spatiotemporal domain and concatenated into a single histogram. The received LOE-LBP-TOP spatiotemporal histogram is further used for enhanced yet relatively compact and easy-to-compute representation of the appearance and motion of the expression in video data. This idea differs from previous works in which edge map of the image is constructed first and after that encoded by LBP-TOP operator. In the latter case, the parameters which are used to construct edge map of the image (*e.g.* filter size, degree of image smoothing, *etc.*) ultimately define the amount of information available to LBP operator. If parameters are selected too globally, local features which are important for facial expression

recognition (*e.g.* small wrinkles, protrusions, shadings, *etc.*) [3] may get lost. That is why we see a promising way to calculate LOE- and LBP-TOP histograms both on raw intensity images and concatenate them at the later stages of processing. This preserves the possibility for LOE operator to use somewhat more global information while ensuring that LBP operator captures all local statistics. In the remaining of the paper, we introduce a methodology for classification of facial expressions in video. The classifier is based on the proposed local spatiotemporal edge and texture facial representation. Support vector machines are used to classify an expression into one of the expression categories. At the end of the paper we discuss shortly about the planned testing of the new method.

METHODOLOGY

The entire methodology of the proposed expression classification is shown in Figure 1. The video is considered as a sequence of images denoted by $\{I(x, y, t)\}_i$, where the first two dimensions (x, y) represent spatial domain and the temporal dimension t is defined by the number of images (frames) i in the sequence. As the first step, facial area is detected using Viola-Jones real-time face detector [12] that has proved to work robustly with detection accuracy suitable for a realistic application. After that 3D image grid is constructed that divides the entire image sequence into a number of video volumes.

LOE- and LBP-TOP spatiotemporal descriptors are then calculated on TOP of each video volume. In Figure 1, the orthogonal planes are depicted as central slices of a video volume and are defined as XY (violet), XT (red) and YT



Figure 1. Block-scheme of the proposed method.

(blue). From these planes, LOE-LBP-XY, LOE-LBP-XT and LOE-LBP-YT histograms are calculated. In this process, LBP histograms are added to the end of LOE histograms (or vice versa) for every plane in each video volume. The histograms obtained from three planes compose a feature vector in a form of concatenated LOE-LBP-TOP volume histogram. After this, the concatenated LOE-LBP-TOP volume histograms are combined into a single concatenated histogram of the entire image sequence that represents a spatiotemporal signature of a given expression. Video volumes can have a certain degree of overlap. As noted in [15], the use of volume overlaps leads to the improved performance of the expression classifier.

A non-linear SVM classifier [2] with radial basis function (RBF) kernel is used to classify a given expression into one of the expression categories. In the training phase, short video clips of facial expressions are used to construct expression models for each expression category. Training of the classifier happens with concatenated LOE-LBP-TOP histograms obtained from video data as described above. The best parameters C and γ of the classifier are selected empirically by the 10-fold cross validation procedure in a grid approach. In the testing phase, two-feature classification (one-against-one) scheme is used for multifeature classification. The idea is to create classifiers for every pair of two classes and the aim is to learn more specific and discriminative features from each pair. This way, N-feature classification problem is decomposed to N(N-1)/2 two-feature problems and a voting scheme is used to define the result of the "winning" classifier.

LBP-TOP Spatiotemporal Descriptors

Local binary pattern (LBP) operator [10] is a grey-scale texture measure that is derived from the image by thresholding pixel values in a local neighbourhood of arbitrary circular shape. LBP(P,R) operator produces 2^{P} different binary codes that are formed by P pixels in the local neighbourhood of radius R. The derived binary numbers encode local texture primitives such as points, curved edges, spots, flat areas, etc. After computing LBP codes for the whole image, a histogram is constructed that describes occurrences of binary codes in the image. This way, LBP histogram represents distribution of local texture patterns over the whole image. For computation of spatially enhanced LBP histogram, binary codes are calculated in separate blocks of 2D image grid of size $N \times M$. The resulted $N \cdot M$ block histograms are further concatenated into a single spatially enhanced LBP histogram that holds information about occurrences of local texture patterns in different parts of the image.

For computation of LBP-TOP histogram [15], statistics on three planes (XY, XT and YT) are computed and then concatenated into a single histogram. Correspondently, the resulting feature vector (histogram) is of $3 \cdot 2^{P}$ length. Figure 2 illustrates the construction of LBP-TOP histogram. In such a scheme, LBP descriptors encode the appearance and



Figure 2. Spatially-enhanced LBP-TOP histogram.

motion of the expression in three directions, incorporating spatial information in LBP-XY histogram and spatiotemporal co-occurrence statistics in LBP-XT and LBP-YT histograms. In our implementation, LBP-TOP histograms are computed in each volume of the input video, resulting into spatially enhanced LBP-TOP histogram of the image sequence as shown in Figure 2. LBP-TOP operator is expressed as:

$$LBP - TOP(P_{XY}, P_{XT}, P_{YT}, R_X, R_Y, R_T)$$

$$(1)$$

where the notation (PXY, PXT, PYT, RX, RY, RT) denotes a neighbourhood of *P* points equally sampled on a circle of radius *R* on XY, XT and YT planes respectively. The best results for emotion-associated expression classification have been obtained [15] with uniform LBP-TOP (8,8,8,3,3,3) operator on 9×8 grid with 70% of volume overlap.

LOE-TOP Spatiotemporal Descriptors

Local oriented edge (LOE) operator [5] is used to detect local edges by convolving pixel values in a local neighbourhood with a set of convolution kernels. Convolution kernels result from differences of two Gaussians with shifted centres and encode the orientation of a local edge in the central pixel of the neighbourhood. Figure 3 shows the orientation template that defines 16 different edge orientations with a step of 22.5°. Before computing LOE descriptors, image is smoothed with a Gaussian filter in order to eliminate noise. LOE operator



Figure 3. Orientation template, $\varphi_k = k \cdot 22.5^\circ$, $k = 0 \div 15$.

outputs a histogram that describes occurrences of local edges of certain orientation in the image.

The idea of LOE histogram calculation in the spatiotemporal domain is similar to the spatial case described above. A reduction of dimensionality of LOE-TOP histogram is done similarly to LBP-TOP. Thus, LOE descriptors are computed on the orthogonal planes XY, XT and YT of each video volume, concatenated first into volume histograms and, finally, into spatially enhanced LOE-TOP histogram of the entire image sequence. LOE-TOP operator is denoted as:

$$LOE - TOP(\varphi_k, \sigma, N, l, P_{XY}, P_{XT}, P_{YT}, R_X, R_Y, R_T)$$
(2)

where φ_k is angle of the Gaussian rotation, $\varphi_k = k \cdot 22.5^\circ$, $k = 0 \div 16$; σ is a root mean square deviation of the Gaussian distribution; *N* denote a size of the filter; *l* defines a level of image smoothing (resolution). The parameter values for facial feature detection from expressive images have been reported [5] as $\sigma = 1.2$, N = 7 and l = 2.

DISCUSSION AND FUTURE WORK

A novel spatiotemporal approach to classification of facial expressions in image sequences is presented. The method is based the combined LOE-LBP-TOP local on spatiotemporal representation. The main advantage of the new representation is its capacity of modelling local dynamic statistics that originates from LOE and LPB descriptors. The proposed LOE-LBP-TOP representation reflects more local variations as compared to original LBP-TOP approach and, therefore, is expected to achieve better performance. Based on the theoretical methodology presented, it is reasonable to expect the combined LOE-LBP-TOP descriptors to constitute a promising facial representation for expression classification purposes.

In the future, we plan to apply the developed methodology to recognize an extensive set of emotion-associated and AU-coded expressions. The results of a systematical testing of the new method will be compared against existing studies that applied LBP-TOP on, for example, gradient and Gabor-decomposed images.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland (projects 129354 and 129502) and the University of Tampere.

- 1. Aleksic, P.S. and Katsaggelos, A.K. Automatic facial expression recognition using facial animation parameters and multi-stream HMMs. *IEEE Trans. Information Forensics and Security 1*, 1 (2005), 3.
- Chang, C.-C. and Lin, C.-J. LIBSVM: A library for support vector machines (2001). http://www.csie.ntu.edu.tw/~cjlin/libsvm.

- 3. Ekman, P. and Friesen, W. Facial Action Coding System (FACS): A technique for the measurement of facial action. Consulting Psychologists Press, Palo Alto, California, 1978.
- 4. Fasel, B. and Luettin, J. Automatic facial expression analysis: A survey. *J. Pattern Recognition 36*, 1 (2003), 259-275.
- 5. Gizatdinova, Y. and Surakka, V. Feature-based detection of facial landmarks from neutral and expressive facial images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 28, 1 (2006), 135-139.
- 6. Kotsia, I. and Pitas, I. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans. Image Processing 16*, 1 (2007), 172-187.
- Lei, Z., Liao, S., He, R., Pietikäinen, M. and Li, SZ. Gabor volume based local binary pattern for face representation and recognition. In *Proc. Int. Conf. Automatic Face and Gesture Recognition* (2008), 1-6.
- 8. Liao, S., Fan, W., Chung, A.C.S., and Yeung, D.-Y. Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In *Proc. IEEE Int. Conf. Image Processing* (2006), 665-668.
- Mattivi, R. and Shao, L. Human action recognition using LBP-TOP as sparse spatio-temporal feature descriptor. *Lecture Notes in Computer Science 5702*. In Proc. Int. Conf. Computer Analysis of Images and Patterns (2009), 740-747.
- 10. Ojala, T., Pietikäinen, M. and Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence 24*, 7 (2002), 971–987.
- 11. Pantic, M., Rothkrantz, J. Automatic analysis of facial expressions: The state of the art. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22, 12 (2000), 1424–1445.
- 12. Viola, P. and Jones, M. Robust real-time face detection. *Int. J. Computer Vision* 57, 2 (2004), 137–154.
- 13. Wehrle, Th., Kaiser, S., Schmidt, S., Scherer Kl.R. Studying the dynamics of emotional expression using synthesized facial muscle movements. *J. Personality and Social Psychology* 78, 1 (2000), 105-119.
- 14. Zeng, Z., Pantic, M., Roisman, G.I. and Huang T.S. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence 31*, 1 (2009), 39–58.
- 15. Zhao, G. and Pietikäinen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence* 29, 6 (2007), 915–928.
- 16. Zhang, Y. and Ji, Q. Active and dynamic information fusion for facial expression understanding from image sequence. *IEEE Trans. Pattern Analysis and Machine Intelligence 27*, 5 (2005), 699–714.

Comfort and Acceptability Study of 3D Devices

Jean-Marc Diverrez

Telecom Bretagne, Département LUSSI jm.diverrez@telecom-bretagne.eu Alice Naylor UBO, Master Image et Son projet 3D-Fovéa lal2ly@hotmail.com

Dusan Iorgovan

Telecom Bretagne, Département LUSSI, dusan.iorgovan@telecom-bretagne.eu

Céline Nicolas

Telecom Bretagne, Département LUSSI, celine.correnicolas@telecom-bretagne.eu

Gilles Coppin Telecom Bretagne, Département LUSSI, gilles.coppin@telecom-bretagne.eu

ABSTRACT

This paper describes the implementation of an experimental procedure to study human behavior when confronted with virtual 3D content (e.g. games and movies). The study focus's mainly on fatigue measurement. The study includes the description of measurement tools, and a definition of survey questionnaires and tests. The aim of the procedure is to standardize rules and recommendations for 3D viewing.

Author Keywords

3D vision, fatigue, eye tracker, convergenceaccommodation, pupil size.

INTRODUCTION

Stereoscopy is created by the eyes which, being horizontally separated by several centimetres, observe two distinct images of the same visual scene, thereby producing two different viewpoints. Binocular disparity or binocular parallax represents the relative difference in position of the projection of the same object on the retina of the eye. Disparities or differences between the retinal image of the right eye versus the left eye, located throughout the visual field outside the point set, will allow the brain to accurately reconstruct three-dimensional space.

Perception 3D Real / Virtual 3D:

In the real world, when we focus on an object, all other objects in our field of vision appear blurred. object all other objects in our field of vision will appear clear. In a real scene when we look at an object, our eyes converge and adjust on it. The object is therefore located at our focal distance (see Figure 1).

In a 3D scene, focus is made naturally on the object (here, the projection screen) and our eyes converge on virtual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org. objects which are projected in the space in front of us [1].

But because of our reflexes, when an object is projected into virtual space, our eyes want to adapt their shape as required and change focal distance. However, if this happened, we would focus on the 3D effect, and the film (screen) would seem blurred – and we wouldn't be able to follow the movie. We make the first eye movement but not the second. The illusion forces our eyes to converge without adjusting. This inevitable "decoupling", spread out over 90 minutes at the cinema, has beenprove to be out of the major cause of asthenopia (eye fatigue) related to 3D[2].

The wide diffusion of James Cameron's Avatar, considered as the first major movie in 3D, allowed a real life test of 3D virtual acceptability.

Following the release, ophthalmologists have now begun to receive their first patients, mainly for migraines during or after the showing[3].

The purpose of this study is to establish a measure of 3D acceptability, in order to record and quantify the inconvenience caused, and implement standardization rules and recommendations for 3D viewing. The purpose of this paper is to present the measurement chain, for measuring



Figure 1. Convergence/Accommodation disparity.



Figure 2. Organization chart of 3D-Fovéa project: Test platform -CHU-contents producer.

the impact of 3D visualisation for users. This approach is based principally on fatigue measurement.

Literature

Studies have already been made on the topic of fatigue and 3D vision. In 2008 Hoffman et al. conducted a study on convergence- accommodation[1] conflicts. Earlier, in 1997, Tetsuri and Hitoshi made the link between the measure of accommodation and convergence, and explored the possible interpretations of pupil size. They showed that when the virtual image and the screen were close enough, convergence followed accommodation. On the other hand, in the case of hyper-stereoscopy there is a conflict between these two criteria, which can generate fatigue. They then showed that the pupil size was smaller when viewing a 3D image, but failed to link the size of the pupil to fatigue[2].

Harle et al. have been able to show a relationship between the size of the pupil and migraines[4]. In 2006 Kazuhiko Ukai worked on fatigue and accommodation-convergence, associated with a 3D images[5]. These studies imply that accommodation and convergence appear to be relevant in fatigue studies, and permit us to consider specific changes to test conditions and to include other parameters.

MATERIAL & METHODS

Materials

We use several 3D display devices, various types of active glasses (Eyes triple shut, Nvidia, Expand) different means of visualization (3D projector, 3D screen) and a variety of content (images, games, movies). Our first tests are on games, using the 3D vision[6] from Nvidia.

Steps

The measurement chain is based on the eye tracker (Seeing Machines, FaceLab). We developed software Fatigue-Monitor (see Figure 3) for this project and allows us to evaluate the raw data of the eye tracker. This application works as a plugin for The Observer^M. We developed this application in Java language. Essentially, its main task is to recover, display, and store the data logged from the eye tracker.



Figure 3. Fatigue-Monitor Software screen shot.

Measure Criteria

The study aims to measure several physiological criteria, related to fatigue and eye discomfort, and then to establish a link between these results and compare them with the results obtained by questionnaire from the subject (see below).

FaceLab directly computes the PERCLOS¹, a factor representing the fatigue. Our work commences utilizing this parameter to which we will then add additional factors. We aim to study the dynamics of convergence and more specifically, the decoupling of vergence-accommodation. We hope to observe the evolution of visual scanning across the screen and to monitor the total distance travelled by the eye. Finally, we will consider pupil dynamics. The Fatigue-Monitor software has been designed in a flexible manner and allows us to integrate new factors, as they arise (Figure 4).

Protocol

For more accurate results, in each test, a comparison between 2D and 3D will be made. In order to avoid any bias in the results the running order will be continuously reversed. The comparison of 2D-3D measures will allow to us to "calibrate" the results based on the subject.

Survey

We have developed questionnaires based on the SSQ of Kennedy et al.[7] and the questionnaire of Witmer and Singer[8]. A questionnaire will be completed at the end of the experiment and a second, the day after, in order to identify any potential impact of 3D vision in the delayed term. These questionnaires have been used in the past and adapted, notably to cover augmented reality, by Sophie Côté and Stéphane Bouchard[9].

PRELIMINARY RESULTS

The first test results were used to calibrate the measurement chain, using controlled stimuli. Further tests were

¹ A metric for fatigue based on the percentage the eyelid that is closed over the pupil in a given time window.



Figure 4. Fatigue monitor environment.

undertaken with 3D. After completely calibrating the measurement chain the platform was rendered operational. The first test was operated on two persons with different inter ocular distance (I.O.D.).

The curve of the first tester with small I.OD, is increasing faster, and stronger than the second tester with normal I.O.D. (see Figures 5 and 6).

CONCLUSION

The objective of the project is to establish a testing procedure based on tools such as an eye tracker and questionnaires. In order to implement standardisation rules we will be conducting tests on 50-100 people so as to verify our methods and conclusions. The project will involve the CHU, for two specific reasons: to define new relevant criteria for the test procedures and the characterisation of subjects as a function of their binocular vision characteristics. The project must also involve content producers who will be able to develop appropriate stimuli for our tests and subjects (see Figure 2). In the short term, we plan to incorporate our measurement tools over a range of visualisation equipment (screen, video projector..).

We have not yet considered the auto stereoscopy concept which has similar issues. It will be part of future



Figure 5. Perclos small I.O.D. : 57.7mm.



Figure 6. Perclos Normal IOD : 65mm.

developments. In the near future it is planned to link in new measurement tools to the test platform, for example thermal cameras, and new physiological sensors (EEG, ECG in particular).

- Hoffman, D.M., Girshick, A.R., Akeley, K., & Banks, M. S. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8, 3 (2008), 1-30.
- Inoue, T. and Ohzu, H. Accommodative responses to stereoscopic three-dimensional display. *Applied optics*, 36 (1997), 19.
- 3. Granet D. 3D headache may be 'avatar' of strabismus or other binocular vision condition. *European Society of Cataract and Refractive Surgeons*, avril 2010.
- 4. Harle, D.E. , Wolffsohn, J.S., Evans, B. J.W. The pupillary light reflex in migraine, In *Ophthalmic & physiological optics*, 25, 3, (2005), 240-245.
- Ukai, K. Human factor for stereoscopic images. In IEEE Int. Conf. on Multimedia and Expo, (2006), 1697-1700.
- 6. Nvidia 3D Vision <u>http://www.nvidia.fr/object/GeForce_3D_Vision_Main</u> fr.html
- Kennedy, R. S., Lane, N. E., Berbaum, K. S. and Lilienthal, M. G. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3, 3 (1993), 203-220.
- 8. Witmer, B.G. and Singer, M.J. Measuring Presence in Virtual environments.In *Teleoperators and Virtual Environments*, 7, 3 (1998), 225–240.
- Côté, S. and Bouchard, S. Documenting the Efficacy of Virtual Reality Exposure with Psycho physiological and Information Processing Measures. *Applied Psychophysiology and Biofeedback, 30*, 3 (2005), 217-232.

Eye Tracking Analysis: Application in a Case Study of a Fast Moving Consumer Goods Product

Emma Lógó emma@erg.bme.hu Eszter Józsa eszter.jozsa@gmail.com Balázs Péter Hámornik hamornik@erg.bme.hu

Budapest University of Technology and Economics, Department of Ergonomics and Psychology Egry J. utca 1, Budapest, H-1111 Hungary

ABSTRACT

The usage of eye gaze tracking is a new and fast developing field of marketing research. Our case study is considering the effectiveness of point of sale outplacements, and the shelf appliances of a specific men's cosmetic product. We invited a total of 50 male Hungarian subjects ranging in age from 18 to 24 to participate in the study. Areas of interest (AOI) and fixation duration have been recorded with Tobii X 120 eye tracker. For the purpose of the analysis iMotions' Attention Tool 3.0 software was applied. The eye tracking data have been combined pre- and post interviews in order to analyze consumer behavior and identify outplacement and shelf appliance design.

Author Keywords

Eye tracking, fast moving consumer goods (FMCG), consumer behavior, shelf display, point of sale.

ACM Classification Keywords

H.5.1. Evaluation/ Methodology, H.5.m Miscellaneous

INTRODUCTION

'Our sense of vision allows perceiving the world in images, motion and color. We use information from the visual sense, in order to move around and interact with objects and environment.'[2]

Objective studies of human eye movements date from around the turn of the twentieth century, although methods involving the use of after-images (direct observation) go back to the 18^{th} century.

In 1898, Delaberre made the first eye movement recordings, using a mechanical lever attached to the eye. Dodge and Cline in 1901 introduced a method for photographing movements of reflection of a light source from the cornea, which remained a standard method of recordings for the next 50 years [6]. The method required the head to be kept as still as possible, which meant that studies concerning eye movements made during active tasks (of everyday life) were excluded.

The major innovation by Mackworth and Thomas in 1962 made it possible to record eye movements during relatively unconstrained activity [6]. They used a camera mounted on the head, which simultaneously filmed the view ahead and the corneal reflection. Their device was successfully used to study diving and flying.

For a time, another method seemed promising, the electrooculography tracking technique. It is based on the fact that an electrostatic field exists when eyes rotate. By recording small differences in the skin potential around the eye, the position of the eye can be estimated. Since it requires the close contact of electrodes to the user, the technique is rather troublesome and not well-suited for everyday use.

By the end of the 1980s when video cameras had become much smaller and lighter, a number of commercial eye trackers began to become available. They were usually based on pupil position, made it visible by illuminating the eye with infrared light to produce a 'white' pupil (also called 'Bright Pupil'), which can be tracked electronically. It gives a pictorial display of foveal gaze direction, or 'point of regard' [1,3]. This technique is still used by a number of eye trackers. Additionally, technological advancements, such as increased processor speed and digital video processing have both lowered the cost and dramatically increased the efficiency of eye tracker equipments. The computer eased the data collection and analysis; brought a whole new era for eye movement research [4,6].

In the first years of the decade, the increasing sophistication and accessibility of eye tracking technologies have generated a great deal of interest in the commercial sector. Generally, commercial eye tracking studies function by presenting a target stimulus to a sample of consumers while an eye tracker is used to record the activity of the eye. Examples of target stimulus can include commercials,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

package designs, shelf displays, television programs, movie films, sport events and most recently web-based advertisements. The resulting data can be statistically analyzed and graphically rendered to provide evidence of specific visual patterns. By examining fixations, fixation durations, saccades and eye blink also the effectiveness of a given medium or product can be determined.

Current research assigns eye tracking as a small window what the consumer is thinking [5].

METHOD

The aim of our research project is to apply the eye tracking measurement as a tool in point of sale marketing (POS), and product development. We analyze the eye movement data in order to evaluate the store outplacement and the shelf appliances of a men's cosmetic product (an FMCG product). Our study included 54 male participants from an age group of 18-24 years (university students). The product's target group is the 18-25 years old male population so the research aimed to test this specific population by the recruited sample. Subjects participated individually and they received a product specimen after the measurement. The session lasted approximately from 6 to 8 minutes including the calibration of the eye tracker device.

We used Tobii X120 eye tracker with Imotions' Attention Tool 3.0 software. Due to the hardware's feature we could use projected picture stimuli, what is larger and more lifelike then a computer display view (this may contribute to the ecological validity of the test setting).

The session started with introduction, and with general information, than continued with a demographic data questionnaire (including questions about possible eye related diseases, consummation of coffee). After successful calibration we started the testing.



Figure 1. The schematic arrangement of the store outplacement presented of the target product and other competitor products as a stimulus in different variations of shelf appliance, and without shelf appliance.



Figure 2. An example of a shelf appliance presented as a stimulus.

We have shown two types of test stimuli for the participants. (1) The test stimuli contained store outplacements with three variations of shelf appliances of the tested FMCG product; and for baseline comparison, a store outplacement picture without any shelf appliance. Figure 1 shows the schematic view of the shelf outplacement of the target product and the other products surrounding it. (2) Than the other part of stimulus pictures were three variations of shelf appliances of the same product, without store scene, in a larger size, and with a black background (see Figure 2). Each stimuli picture was visible for the participants for six seconds. We have shown the two types of stimuli slides randomly mixed, and we placed interstimuli slides between the test pictures: these were noise pictures, consisted of black and white points mixed random, and they covered the same size as the following test stimulus. These interstimuli were visible for only two seconds.

After finishing the eye tracking measurement participants completed a questionnaire about their own consumer behavior and brand preferences. Finally an interview was conducted about the presented store outplacement designs, and shelf appliance variations.

Our results suggest the most effective outplacement design and shelf appliance of an FMCG product along with an analysis of recognition and brand preference among men's deodorant product.

Ethical Statement

All the participants agreed personally to participate in the research. The information provided, and the test design is fitting the ethical regulations (Ethical Codex) of the Hungarian Psychologists' Association. The producer of the tested men's cosmetic product approved to publish the pictures tested in the studies, including the product's name.

CONCLUSION

On the grounds of our post-questionnaire we have formed different consumer groups by their brand preference. We expect that we can explore numerous relationships concerning these consumer groups and their eye movement pattern. Our results also can orient display designers to choose the most attractive Point of Sale (POS) marketing tool [5].

This is a presentation of a consumer behavior measurement method by eye tracking. In our case study, we give an example for the application of this method for the store outplacement, and the shelf appliance of a specific men's cosmetic product.

REFERENCES

1. Andreassi, J. L. Pupillary response and behavior. In J. Andreassi: *Psychophysiology: human behavior and physiological response* (4th Edition). Mahwah, NJ. Lawrence Erlbaum Associates. (2000), 218.

- Clarkson, J. Human capability and product design. In N. J. Schifferstein, *Product Experience*. Oxford: Elsevier (2008), 165-198.
- 3. Duchowski, A. Eye *Tracking Methodology: Theory and Practice*. Berlin: Springer, 2003.
- 4. Land, F. M. Eye movements and the control of actions in everyday life. Progress. In *Retinal and Eye Research*, 25 (2006), 296-324.
- Langeman, M. A Review of Eye Movement Tracking Research. (April 25, 2005). <u>http://www.langeman.net/SYDE_740_paper.html</u>.
- Mohamed, A. O., Da Silva, M. P., Courboulay, V. A history of eye gaze tracking. (2007. December 18) <u>http://hal.archives-</u> ouvertes.fr/docs/00/21/59/67/PDF/Rapport_interne_1.pdf.

Physiological Measures, Eye Tracking and Task Analysis to Track User Reactions in User Generated Content

Aude Dufresne

Department of Communication, University of Montreal C.P. Succ Centre-Ville, Montréal, Qc, Canada, H3C 3J7 Aude.Dufresne@umontreal.ca

Sylvain Senechal

Hautes Études Commerciales Montréal, 3000, chemin de la Côte-Sainte-Catherine Montréal (Québec), Canada H3T 2A7 sylvain.senecal@hec.ca

ABSTRACT

In order to improve human-computer interaction, eyetracking, physiological measures and task recognition are used to assess the content and quality of interaction in an ecommerce application with User Generated Content. Results are analyzed in conjunction with users' verbalization and their subjective assessment of attitudes toward the product and their wish to contribute.

Authors Keywords

physiological measures, eye tracking, task analysis, e-commerce, user generated content.

INTRODUCTION

Understanding affective reactions is essential to improve human-computer interaction and the design of systems. This assessment of user motivational state can be used to evaluate interface design, to define adaptation in systems or to study how interfaces influence the use of systems. We will present our research on the integration of different measures and AI techniques to analyze interactions in the context of Web 2.0 applications where e-commerce integrates user generated content.

CONTEXT

The research focuses on using physiological and eyetracking measures to assess online consumer behaviors in the context of Web 2.0 sites. Described as a second-stage

Sandrine Prom Tep

Hautes Études Commerciales Montréal 3000, chemin de la Côte-Sainte-Catherine Montréal (Québec), Canada H3T 2A7 sandrine.prom-tep@hec.ca

François Courtemanche

Department d'Informatique et Recherche Opérationnelle, University of Montreal francois.courtemanche@umontreal.ca

evolution of the Web, this emerging context is characterized by increasing consumer participation which rely on bidirectional communication tools and technologies, enabling "customers to share their opinions and experiences on goods and services with a multitude of other consumers" [32]. Web 2.0 participative technologies do not equate participation though, and site managers struggle to find new Web 2.0 business models that would lead to profitability [15].

Any form of web site contribution from consumers has shown to be primarily motivated by the desire for social interaction [32] and it is important to assess how the interfaces support the social and confidence dimensions of the user toward the content but also to encourage him to contribute.

Eye-Tracking and Measures of Affective Reactions

Eye-tracking measures are being used regularly to analyze what are the points of fixation, the duration of fixations and the scan paths, in order to follow what are the zones of interest and at the opposite, which information or navigation elements may have been missed by users. These measures are used along with usability methodologies to evaluate applications before they are made accessible. In most cases a few subjects are used in a design context, but few precise methodologies or studies are made to define principles that can be used to design or evaluate applications using those measures.

Most eye-tracking applications in human-computer interactions can be separated in mainly two categories: 1) describing the user's state and 2) describing the interactions' state. The former consist of analyzing data brought by an eye-tracker in order to infer different types of information on the user psychological state. Pupil diameter variations and blink rate can be related to the user cognitive

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

load [4, 29]. Information on the user current amount of mental effort can be used for many different purposes in the context of HCI. For example, Bailey & al. [3] use eyetracking to identify moments where users can be interrupted at lower cost. These moments usually occur at task boundaries, which are identifiable by a decrease in pupil size (cognitive load). Conati & Merten [11] used eyetracking for online modeling of students meta-cognitive behaviors (self-explanation and effective exploration) during interaction with a learning environment. Eyemovement and pupil dilation are also used in real context simulations in order to evaluate awareness and vigilance [16]. Pupil size variation can also be used to infer information on the subject affective state. Partala & al. [24], who concluded that pupil enlarge during negative and positive stimuli.

The second category of application of eye-tracking to HCI – describing the interaction's state – has been paid less attention by researchers. Salvucci & Anderson [28] developed different techniques to analyses eye movement protocols in order to infer the user' current position in a cognitive task model. The scan paths have been used to describe more frequent exploration sequences or even to recognize users' activities.

Baccino & Manuta [1] have developed interesting measures to recognize the user's cognitive processes, using eyetracking measures, depending on the saccade angles and the distance between fixations.

Physiological Measures of Affective Reactions

In our study we used in addition to eye-tracking data two physiological measures: Galvanic Skin Response (GSR) and blood volume pressure (BVP) to identify user's emotional state. The GSR measures can inform about stress and relaxation [25] and is indicative of emotionally significant HCI events and situations [33]. The heart rate can reflect emotional activity and can help differentiate between negative and positive emotions. It also reflects the amount of mental effort [6, 27], since heart rate variability correlate with the visual display levels of complexity.

The Need for Task Analysis to Integrate Context in Affect Recognition

From an evolutional point of view, primary emotions such as joy, fear or sadness are hardcoded responses to inner or outer environmental changes. More complex emotions such as pride or shame result from the conscious association between primary emotions and the current situation in which they occur [23]. Therefore, taking into account the contextual interaction factors in which physiological and eye-tracking measures are collected is a key step in attaining efficient affect recognition.

Physiological measures, such as skin conductance, are correlated to arousal but cannot give information on emotional valence (positive of negative aspect of emotions). Contextual information can help fill this lack of information. For example, negative affective reactions like stress are associated to understanding difficulties, uncertainty in choices, unexpected or dead end situations. Positive reactions like surprise or interest may be linked to task success or to a positive orientation reaction toward a stimulus. All these contextual factors can be inferred using appropriate and effective task tracking techniques in order to support affect recognition. In addition to supporting recognition, contextual factors can also be used to identify the cause of affects. For example, the precise icons or interface elements, which are inefficient and cause, in the context of a given task, impatience and task abandon. This information is highly important for HCI diagnostic and quality evaluation.

Within human-computer interactions, task recognition usually aims to analyze the interaction flow coming from the user, using a filter corresponding to the model of the task being executed [7]. The goal is to identify which *part of the task* the user was executing when he generated a particular *sequence of interactions*. In our work, we developed a task recognition approach based on machine learning algorithms – Layered Hidden Markov Models [22], where data (*observed interactions*) are associated to targets (the *task model components*) in a hierarchical structure.

The ever growing complexity of today's interfaces makes difficult the use of task recognition techniques based only on standard user's interactions (mouse clicks and keyboard). Our approach uses eye tracking to overcome this problem. In most interfaces, gaze position can be separated and viewed in terms of areas of interest (AOI). Therefore, the *data* used in our machine learning technique are sequences of interactions, including mouse clicks, keyboard events and gaze position events (AOI). The targets of our machine learning technique are each sub-tasks defined in hierarchical task model (Card, Moran and Newell, 1983). A task model links the way a user mentally structure a task (what to) to the different possible interactions to achieve the task (how to) [3]. It is represented as a hierarchical structure of goals, sub goals.

The task recognition system is developed in two phases. First, the *training stage*, the Layered Hidden Markov Models (LHMM) task recognition algorithm is trained using data collected while subjects execute the proposed task. In order to train the system, the tasks given for training are more specific or the recordings may be tagged to categorize what the subject is doing at a given time in the interaction. During the second phase, the *experimental phase*, different subjects are given the same tasks but they are left free to find a way to do them. The trained LHMM model is then used to analyze their interactions with the system.

The Assessment of Affect in E-Commerce Interaction

With e-commerce and the multiplication of retail Web sites, taking emotions into consideration becomes crucial as it has been clearly established that consumers are not always rational in their choices, nor utilitarian in their motivations [10]. Whereas there is much more than cognitive information processing going on while consumers are online [12] measuring behavior for interactive systems evaluation has almost exclusively limited its focus to cognitive activities. Models combining emotions with human information processing variables are scarce and when available, they rely on reported data [9]. Research has shown though, that emotions are experienced both at the conscious and preconscious levels [3].

So far, studies on consumers' online participative behavior have mainly focused on the "why" and "how" of participation, identifying:

1) *drivers of contribution* such as motivations to create and maintain a blog [5], or to articulate themselves on consumer-opinion platforms [2] as well as

2) factors encouraging *participation* in virtual communities, such as offline interactions and perceived usefulness [17] or

3) the modes of *inter-personal influence* which can take place in these online communities for the adoption and use of products and services usefulness [5].

Reactions to interface design are very often measured in passive situations [26], but in the context of e-commerce applications, it is important to develop researches when there is more interaction. For example in a web site offering online transactions, it is necessary to follow the specific steps according to a task model, where actions and relevant areas of visualization are defined. It is therefore necessary to follow the user activities and to study changes in his physiological and eye-tracking measures accordingly.

EXPERIMENTATION

We were interested to assess how users are influenced by "user generated content" in the context of e-commerce. We experimented using the Amazon web site, where books have comments and assessment (number of stars) from other users. Using a classical experimental design, we tested the differences in consumers' attitude toward contribution, online intention to contribute and actual contribution with and without social functionalities, within Amazon Web site.

To assess the usability and the UGC social perspective we also used think-aloud protocols [9] and qualitative questionnaires. We complemented both sources of data with eye-tracking and physiological measures, so as to obtain a richer perspective on the phenomenon.

User testing was conducted in the Bell Solutions Web Laboratory. We used the Internet Explorer® Web browser with a TOBII (X120 eye-tracker). We recorded comments from the user and the output of the control monitor, with an

overlay of the gaze movement. Physiological measures (GSR, BVP) were monitored using a Biograph system. Each system was running on a different computer, for which the clocks had been synchronized. The Tobii was calibrated and then started with the Biograph both having a measure at each 1/60 sec. Data were synchronized using the time stamp in each data file. Though raw data were collected, they were reduced to means for each fixation in a zone. Screen interactions and navigation activity are also recorded in order to integrate the task analysis dimensions. The task model was described as the structure of possible actions in relation to exploring information on a series of books in the web site. For example they could read descriptions, look at assessment and contributions from other users, add an evaluation, etc. In the training phase, five users were asked to accomplish the tasks. For each user data were collected from physiological and eye-tracking measures (pupil size, AOI), along with navigation information. These observations were used to train the recognition of the structure of actions, so it was possible to recognize what a user was doing, when given more general instructions.

In phase two, the system was tested with 27 subjects. The users were given general instructions to explore the books: descriptions, comments, and evaluations. These instructions were given at the beginning and subjects were left free to explore as long as they wished. After the exploration, they were asked to comment on the interaction and the components of the interface. They were also ask questions about social and credibility dimensions which may have influenced their appreciation of the user generated content. For the experiment, we asked subjects to explore a set of books of general interest, who were saved locally to insure that each subject would have access to the same content. For those pages, different areas of interest (AOI) were defined corresponding to the structure of presentation: description, detailed description, keywords, price, statistics of evaluation (number of stars), comments positive or negative, recent comments, ordering zones, etc.

The data collected gives for each fixation: what was the page and the zone being explored, physiological measures (GSR, BVP) and Pupil size (mean for both eyes), the task structure was also recognized (goal, subgoal). Results were analyzed in order to see how patterns of interactions could be extracted and how they could be related to emotional reactions. GSR and BVP measures were normalized using the baseline for each subject, using data collected at the beginning of the session. The baseline refers to the average measures of GSR during a rest period before a session. We used Lisetti &al.(2004) equation to correct observations :

$GSR_{normalized} = (GSR-GSR_{relaxation})/GSR_{relaxation}$

Pupil size and BVP were accordingly corrected for each subject. Even though they don't vary as much between individuals, we wanted to insure that observed differences were only due to the context of navigation and not to individual differences or individual mood in general.

ANALYSIS AND PRELIMINARY RESULTS

Eye fixations on areas of interest defined around the social functionalities elements of interface, as well as GSR and BVP measures are analyzed along with reported measures of contribution intent and buying intent. As far as evetracking data is concerned: the more eye fixations on a social functionality element of interface design related to one of the book's content, the more useful is the social functionality for decision-making. Also the physiological reactions are compared to eye-tracking measures and impact on attitudes toward the book. The hypothesis is that the longer the eye fixations on a social functionality element of the interface related to one of the book's content and the stronger the reaction to it, the stronger should the participant's intent to contribute online and to buy the book. Pretests have shown that comments were attracting a lot of fixation time, so user were spending time trying to assess the comments. However there was more GSR reactions to description than to comments.

When looking at the types of activity, it appears clear that reading the table of statistics of evaluation was an important part of the interaction. For judging the book, reading the description appears important, and looking at the keywords also. This may look surprising considering the keyword zone is very small. But keywords content is dense and it takes long for the user to read and reflect on implications. We also compared how the task analysis could be compared to results on AOI statistics. In general both measures agree, but in some cases, especially when the zones are small task recognition appears to give better results (for example: reading keywords vs. reading description, reading comments vs. looking at the statistics of evaluation (which is a small zone at the beginning of description). In the experimentation with more observations for training the task recognition algorithm and also with more subjects, the agreement will certainly be improved. But still first results suggest that the structures of activity might be more precise than AOI, since in some cases AOI associated might be small and eye-tracking only might be imprecise. Another interesting result is the analysis of the scan paths and how they are linked to the types of zone.

- 1. Baccino, T., & Manunta, Y. (2005). *Eye-Fixation-Related Potentials: Insight into Parafoveal Processing.* Paper presented at the Journal of Psychophysiology.
- 2. Backs, R.W., Boucsein, W. (Eds.), 2000. Engineering Psychophysiology. Issues and Applications. Erlbaum, Mahwah, NJ.
- 3. Bailey, B. P., Adamczyk, P. D., Chang, T. Y., & Chilson, N. A. (2006). A framework for specifying and monitoring user tasks. Computers in Human Behavior, 22(4), 709-732.

- Beatty, J., Lucero-Wagoner, B. (2000). Pupillary System. Chapter 6 in CACIOPPO, J.T., TASSINARY, L.G., AND BERNTSON, G. Handbook of Psychophysiology. Cambridge University Press.
- 5. Bonie Nardi, Diane J. Schiano, Michelle Gumbrecht, Luke Swartz (2004), "Why we blog?", *Communications of the ACM*, vol 47, no12, pp 41-46.
- 6. Boucsein, W. 2000. The Use of Psychophysiology for Evaluating Stress-Strain Processes in Human-Computer Interaction. In Backs and Boucsein 2000,289–309.
- Carberry, S. (2001). Techniques for Plan Recognition. User Modeling and User-Adapted Interaction, 11(1-2), 31-48.
- 8. Card, S., Moran, T., & Newell, A. (1983). The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale: NJ.
- 9. Chamberlain, L. and Broderick, A.J. (2007). *Qualitative Market Research: An International Journal*, **10** (2): 199-216.
- 10. Childers, T. L., Carr, C.L., Peckc, J., and Carson, S. (2001). Hedonic and utilitarian motivations for online retail shopping behavior. Journal of Retailing, 77(4): 511-535.
- 11. Conati C. and Merten C. (2007). Eye-Tracking for User Modeling in Exploratory Learning Environments: an Empirical Evaluation. Knowledge Based Systems, Volume 20, Issue 6, Elsevier Science Publishers B. V. Amsterdam, The Netherlands
- Ethier, J., Hadaya, P., Talbot, J., Cadieux, J. (2006).
 B2C web site quality and emotions during online episodes: An empirical study, *Information & Management*, 43: 627–639.
- Hirschman, E. C., & Holbrook, M. B. (1982). Hedonic Consumption: Emerging Concepts, Methods and Propositions. *Journal of Marketing*, 46: 92-101.
- Holbrook, M.B., Batra, R. (1987). Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising, *Journal of Consumer Research*, 14(3): 404-419.
- 15. UGC: More Popular than Profitable, eMarketer, 2009
- 16. Ji, Q. and X. Yang, Real-time eye, gaze, and face pose tracking for monitoring driver vigilance, *Real-Time Imaging 3*, (2002) (8), pp. 1077–2014.
- 17. Joon Koh, Young-Gul Kim, Brian Butler, Gee-Woo Bock (2007). Encouraging participation in virtual communities, *Communications of the ACM*, vol 50, no2, pp 68-73.
- 18. Kirwan, B., & Ainsworth, L. K. (1992). A Guide to Task Analysis. Taylor & Francis, Ltd.
- 19. Lee S., McQuiggan S. W., Lester J. C. (2007), Inducing User Affect Recognition Models for Task-Oriented Environments. User Modeling 380-384.

- 20. Lisetti C. L., Nasoz F. (2004), Using Non-invasive Wearable Computers to Recognize Human Emotions from Physiological Signals, EURASIP Journal on Applied Signal Processing - Special Issue on Multimedia Human-Computer Interface, vol. 2004, n 11.
- 21. Nasoz F, Lisetti CL, Alvarez K, Finkelstein N (2003). Emotion Recognition from Physiological Signals for User Modeling of Affect. International Journal of Cognition, Technology and Work - Special Issue on Presence.
- 22. Oliver, N., Garg, A., & Horvitz, E. (2004). Layered representations for learning and inferring office activity from multiple sensory channels. Computer Vision and Image Understanding, 96(2), 163-180.
- 23. Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge [England] ; New York: Cambridge University Press.
- Partala T., Surakka V. (2003). Pupil size variation as an indication of affective processing. International journal of human-computer studies. vol. 59, no1-2, pp. 185-198.
- 25. Picard R. W. (1997) Affective computing. MIT Press.
- 26. Prom Tep, S., Dufresne, A., Saulnier, J., & Archambault, M. (2008). Web site quality evaluation combining eyetracking and physiological measures to self-reported emotions: an exploratory research. In Measuring Behavior (pp. 224-225). Maastricht, Netherland.
- 27. Rowe D.W, Sibert S., Irwin D. (1998). Heart Rate Variability: Indicator of User State as an Aid to Human-Computer Interaction, Proceedings of ACM CHI 98.

- 28. Salvucci, D. D. and J. R. Anderson (2001). Automated eye-movement protocol analysis. *Hum.-Comput. Interact.*, 16(1): 39-86.
- 29. Siegle, G. J., N. Ichikawa, et al. (2008). "Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses." <u>Psychophysiology</u> 45: 679-687.
- 30. Subramani and Rajagopalan (2003), "Knowledge-Sharing and Influence in Online Social Networks via Viral Marketing" *Communications of the ACM*, vol 46, no12, pp 300-307.
- 31. Tao L., Akinobu M., Shigeo M. (2008). Using subjective and physiological measures to evaluate audience-participating movie experience. ACM Press
- 32. Thorsten Hennig-Thurau,. Kevin P. Gwinner; Gianfranco Walsh; Dwayne D. Gremler (2004), "eWOM via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet?", *Journal of Interactive Marketing*, vol 18, no1, p.38.
- 33. Ward, R. D., P. H. Marsden, B. Cahill, and C. Johnson. 2001. Using skin conductivity to detect emotionally significant events in human-computer interaction. In Proceedings of BCS IHM-HCI 2001, 10-14 September, 2001, Lille, France, 2: 20-25.

ACKNOWLEDGEMENTS

This research was made possible by a NSERC grant for research and development in collaboration with Bell Canada.

Measuring Reading Characteristics Using Eye Tracking and a Context Adapting Classification System

Jakob de Lemos jakob.de.lemos@imotionsglobal.com Íris Ólafsdóttir iris.olafsdottir@imotionsglobal.com

iMotions – Emotion Technology A/S Lindevangs Allé 6, st Copenhagen, Denmark

ABSTRACT

This abstract introduces a new automated method of measuring reading patterns by analyzing eye properties via an eye tracking platform and context analysis. The method utilizes eye tracking data and stimulus text properties in a way that makes it possible to identify reading fixations and reading patterns with a high accuracy compared to looking at fixations only. This helps doing a fast initial analysis on how the text is cognitively perceived and thereby provides easier access to valuable high level information on how the text is comprehended.

Author Keywords

Eye Tracking, reading analysis, gaze analysis, text recognition.

ACM Classification Keywords

I. Computing Methodologies, I.5 Pattern Recognition, I.6 Simulation and Modeling.

INTRODUCTION

The eye tracking and context based method for quantifying reading is implemented in Attention Tool® a software for measureing eye tracking, human emotional activation and reading patterns on visual stimuli within consumer research. The reading metrics are obtained by using the output from a non-intrusive (remote) eye tracking hardware running at 60 Hz and an image processing system identifying the stimulus characteristics. Each gaze pattern is extracted from the respondents' eye via the eye tracking device while the respondent is exposed to the stimulus. Utilizing the knowledge about the stimulus characteristics, the reading results are classified and delivered directly as a quantified output of the reading patterns. Via the eye tracker the software furthermore collects and analyzes several subtle changes within the respondent's eye gaze characteristics, blink characteristics, and pupil change characteristics in order to determine the emotional activation (Arousal) and traditional eye tracking metrics.

READING

Reading is a process in which the eyes are fixated on successive locations in a text while information is brought into the processing system of the brain. Reading Meter in Attention Tool® detects the reading pattern of the eyes by combining different methods.

Fixations, Saccades and Reading Patterns

During reading the eyes move across the text in a sequence of fixations where the eyes are steady, separated by fast movements called saccades. The eyes do neither fixate on each letter in a word nor sequentially from word to word. Skilled readers skip words and make regressions to material already read about 15% of the time. The fixation duration is on the range from 100 to over 500 milliseconds the average is 200-250 msec. The distance which the gaze travels during each saccade is between 1 and 20 characters the average being around 7-9 characters. Thus there is a considerable variability in fixations and saccades both between readers and for the same person reading a single passage of text [1]. Figure 1 below shows an example of saccade length and how letters around a fixation are perceived.

Fixation and saccade length, the probability of fixating a word and the number of fixations on individual words can be related to cognitive processing as is done in the E-Z Reader model [3,4] Reading Meter focuses on whether a text block is read or not, without taking other cognitive processes into account and displays the resulting statistics.

Gaze Behavior and Text

Fixation alone cannot reveal if the subject is reading but the

Figure 1. Line one represents a normal line of text. Line two represents a 17 character window (fixation point indicated by the dot) and the third line presents a window for the subject's next fixation. In line two and three, original letters are replaced by x's. [2]

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

saccade pattern can give a good indication. The longer the text, the more obvious is the reading pattern and the system can predict reading with more certainty. Gaze behavior for some objects is similar to that observed for reading a small block of text. For example when a face is observed, the subject usually gazes from one eye to the other repeatedly and down to the mouth. A similar pattern is often observed for smaller text blocks. Figure 2 shows an example of a face and a small text block with the corresponding gaze patterns that are similar in character.

As a line of text becomes longer, the number of fixations along the text increases and therefore more reliable results can be expected from Reading Meter. For a text block of only three words or less it is almost impossible to detect a reading pattern, as the number of fixations is most likely too low. In addition to looking at the fixations and saccade pattern the underlying system also analyses the properties of the text in the stimulus, this image processing based approach provides additional information about the underlying text. Based on this information the system can conclude that fixations across the text are due to reading actual text and not looking at e.g. a face, the robustness and certainty of the reading detection improves the longer the text blocks. Furthermore information about text size helps the system to detect the pattern as the saccade varies proportionally with the text characteristics. Therefore the system analyses the stimulus to detect underlying text through image processing in addition to the gaze analysis. This is explained in the final section.



Figure 2. Gaze patterns while looking at a face and a short text block.

How Many Words Are Needed for a Good Reading Analysis?

It is difficult to say how many words are needed for perfect reading analysis, as the number of fixations is the crucial factor. It is not possible to predict how many fixations are expected for a word of a certain length as the average number of fixations within a word is not proportional to the number of letters in the word [1] But with a line of five words or more, experience with Reading Meter has given a good outcome where all reading patterns were detected. In a recent benchmark study the accuracy of the reading system showed that less than 15% error could be expected in 97% of the cases.

RESULT METRICS

How Many Readers and How Much Read

The yellow sticky note in figure 3 shows two results: *Readers*, which is the number of respondents who read something within the region and Read(%) which is the average amount of text read by the respondents. The Pink label shows the order of the area with most readers - assigning the lowest number to the region with the highest number of readers in the stimulus.

Using simple gaze analysis from traditional eye tracking it is possible to select a text area using an Area Of Interest (AOI) tool to determine how many subjects looked into the selected area. The subjects that looked into the area might have been reading all the text, a few words or simply just looked somewhere within the area without reading a word. If you are interested in knowing if the text was actually read, by how many respondents and how much of it was read on average, you need to apply reading analysis, looking closer at what type of "looking" was going on, on what type of content. The figure below shows the information provided by the Reading Meter analysis. For this image 13 out of 30 subjects read some part of the text (Readers: 13/30) and on average they read around 40% of text (Read(%): 40%).

Reading Intensity Map

The Reading Intensity Map in Reading Meter is a measure of how much time respondents spent reading a particular part of the text. It is normalized to the total number of respondents. Figure 4 shows statistical output from Reading Meter for four different text areas along with the



Figure 3. Reading analysis result example.



Figure 4. Reading Intensity Map.

corresponding intensity map which has three colors: red, yellow and green. The colors interpret the following:

- *Green:* Low reading intensity areas have attracted less than 30% of the total reading time. This can be because the readers read fast or because few readers read this text.
- *Yellow:* Medium reading intensity marks an area where 30-60% of the time was spent. These areas have usually been read by around half the respondents.
- *Red:* High reading intensity area is where more than 60% of the reading time was spent. It is often seen at the start of a text clause because many respondents read only the first line. It can be observed on text that is attention grabbing due to different font type, size or color. High intensity area can also be observed in semantically or syntactically ambiguous sentences due to regressive saccades [3].

In the example in figure 4 most of the text was read by someone in area one, two and three which are green. As only one reader was detected in area four there is no Intensity Map. The yellow color in area one shows us that most readers has reading fixations on the first six lines of text. The red area indicates that most of the respondents read these lines or that many of the respondents had many reading fixations on particular words such as "SloopyBucks" as illustrated in the high intensity reading fixation area in text box 1.

READING METER PROCESS

Reading Meter in Attention Tool® is based on several processes as illustrated by the flow diagram in figure 5. In short, the user selects a text area and if the Image Processing System detects sufficient amount of text within the area, a reading analysis is performed, using information about the underlying text and gaze behavior in the area. Reading Meter then provides statistics about reading for the selected text area. The Image Processing is used for automatic text detection. The text features extracted include information about the text location and characteristics such as text size and length. The user is asked to approve the number of lines detected by the system. The Gaze Analysis system extracts the fixations from the gaze data for every respondent. It calculates fixation location and duration, saccade directional speed and acceleration.

The Reading Pattern Analysis uses the gaze features to detect reading patterns. This method is based on a classical



Figure 5. Flow diagram for reading analysis in Attention Tool. User selects a text area and if text is detected within the area a reading analysis is performed. The reading pattern is analyzed using information about the text characteristics in combination with the gaze data. At last statistics for the text area are provided.

approach to reading pattern analysis, where the probabilities of the saccades being part of a reading pattern or not, is based on its speed and direction. These probabilities are added together throughout time, meaning that if many reading-saccades appear in a row the probability that this gaze behavior is due to reading increases. If the subject looks into another area the probability sum is initialized. To improve this reading detection algorithm, Reading Meter also takes gaze acceleration and the information about the underlying text into account to provide a good estimate on reading behavior.

CONCLUSION

The automatic approach to reading analysis proves to be quite robust on standard product material used in consumer tests. The precision of the system depends on two main factors; 1) the quality of the eye tracking data and 2) the success of the image processing system identifying the text. Further evaluation of the performance is currently ongoing to improve the accuracy of the system.

- 1. Rayner, K., & McConkie, G. W. What guides a reader's eye movements? *Vision Research*, *16* (1976), 829-837.
- 2. Keith Rayner, Eye Movements, Perceptual Span and Reading Disabilities. *Annals of Dyslexia*, 33 (1983).
- Reichle, Pollatsek, Fisher, & Rayner. A-Z Reader model, 1998.
- Reichle E., Rayner K., and Pollatsek A. The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26 (2003), 445-476.

DogCam: a Way to Measure Visual Attention in Dogs

Alejandra Rossi Indiana University 1101 E. 10th street, Bloomington, IN. 47405, USA alrossi@indiana.edu Francisco J. Parada Indiana University 1101 E. 10th street. Bloomington, IN. 47405, USA fjparada@indiana.edu

Colin Allen Indiana University 1011 East Third St, Bloomington IN. 47405, USA colallen@indiana.edu

ABSTRACT

Modern technology, including lightweight cameras and eye tracking equipment now seem to afford the possibility of tracking where an animal's visual attention is directed in real time and with greater precision than can be had by fixed cameras or third-party observers. The current hypothesis is that a dog-mounted camera can provide a reliable source of precise information about the dog's moment-to-moment visual attention at a resolution that exceeds other techniques for assessing gaze direction. In particular, we are interested in gaining precision at scales of 10cm or less, and 500ms or less, which will allow us to say where dogs look at when are engaged in communicative signaling, such as during a stereotypical play bow or petting situation.

With this new technology we will put some of the techniques that have been effective in generating knowledge about human cognition to work in studying dog-dog and dog-human interaction.

Author Keywords

Social attention, dogs, eye-tracking, animal cognition methods.

INTRODUCTION

Social cognition is one of the most studied areas in current cognitive-neuro-endocrinological research; its relevance and applicability are of great interest to the scientific community. Research on social cognition has importantly increased our understanding of relevant aspects of non-human and human social behaviors, e.g. empathy, theory of mind and its relevant malfunctioning in humans, such as autism spectrum disorder (ASD) and anxiety disorders, among others [4].

The current project is evolution-oriented and takes the intriguing model of the dog as its focus. The study of dogs could shed light on interesting aspects of evolution in general and social cognition in particular. Research on dogs grew out of criticisms of research conducted on non-human primates. Non-human primates are generally considered to provide the best model for interspecies comparisons that would improve our understanding of human cognitive abilities and disorders. However, there are ethical, methodological and financial stumbling blocks that hinder non-human primate research. These issues have led researchers to adopt two solutions. Some researchers have pursued the strategy of running experiments in locations where they can count on the presence of a high population of certain primates. An example is this approach is the study of the rhesus monkeys at El Cayo Santiago. However, because this species was highly habituated to human presence, questions have been raised about generalizability of the results.

Another solution researchers on animal cognition have adopted has been to find new models. The study of domestic dogs (Canis lupus familiaris) is ideal for two reasons: for their cognitive abilities they possess and for the fact that they live among humans, so that their natural environment is one in which humans play a central role [8]. Dogs are a unique case in nature, as humans domesticated them 15,000 years ago [3]. The specifics of the domestication processes are still not clear, but what is clear is that the domestication was successful. An apparent coevolution with humans is one of the factors that make this inter-species relationship unique. Dogs are extremely social animals; notably, dogs are able to understand and to work with social cues such as pointing, head nodding and gazing, among others. These sophisticated social skills may be specialized as a result of domestication and possibly coevolution. Finally, dogs are also sensitive to attentional states in humans and they also have been shown to be sensitive to the visibility of human's eyes [5, 6].

This last point was a major motivating factor to develop this project: if dogs are attentive to faces and look at the eye region searching for information (as some research has suggested) that would make them even more unique in nature. It would not be unreasonable to think that they have evolved a specialized system that allows this social behavior. For example, apes do not look at human's eyes without previous training [7]. This does not mean that they are not capable of this particular behavior; rather, because

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

apes, such as chimpanzees, have a strict social hierarchy, eye contact can be read as a sign of aggression.

Interestingly, the dog's closest relative – the wolf - does not look spontaneously at faces, does not follow gazes, and, in general, performs worse in all of the aforementioned tasks than dogs do.

Domestication processes could explain why dogs can and do look at faces. And even though dogs are not 'visual animals', the simple possibility of measuring the visual field or measuring the dog's gaze is a huge step in animal cognition research, where third-person view recordings and behavioral description has been the leading approach. In the next section we will explain the methods a little bit further.

METHODOLOGICAL APPROACH

Unfortunately, the evidence for social cognitive abilities in non-human animals, specifically of gaze-related behaviors, is still weak. Researchers in animal cognition rely primarily on third-person behavioral measures. Studies typically focus on where the animal looks and use video cameras to record the experimental sessions. Due to several practical constraints, data gathered in this way does not tell much about the exact location of where the animal's attention is directed.

These methodological issues have led us to develop new methods to investigate with higher precision and accuracy where the dog's focus of attention is directed in sociocognitive tasks; specifically, we have developed the use of a head-mounted camera since it can provide a reliable source of precise information about the dog's moment-to-moment visual attention at a resolution that exceeds other techniques for assessing behaviors.

In particular, we are interested in gaining precision at scales of 10cm or less, and 1000ms or less, which will allow us to roughly estimate gaze direction (and indirectly attentional focusing). Hence, we are able to determine, for example,



Figure 1. Whiteboard with calibration dots.

precisely which portions of a play partner's body are relevant when the partner is engaged in communicative signaling, such as during a stereotypical play bow or petting situation. In order to relate behaviors between human and dog, the human participants wear a portable human eyetracking system and the dog participant wear a DogCam. These video streams are synchronized in time. We have designed and built two DogCam models and we will briefly present them here.

DogCam-v1. This model consists in a head-mounted camera using a specially designed harness that is comfortably adjusted to the dog's head. The participant dog is trained to wear the harness (without the camera) for a period of a week. During this accommodation period the dog has to play with the owner during 25 minutes or so in a daily basis. After this period we can carry on with the experiments. Even though, DogCam-v1 records the dog's visual field, in order to analyze the data we need to make a rudimentary calibration.

Calibrating DogCam-v1. The owner commands the dog to sit and we present a whiteboard in front of the dog. We position the whiteboard in order to fill the whole visual field that the camera can record (fig 1). In the whiteboard there are 13 calibration dots and we used a treat to attract the dog's attention to a dot (fig 2). After the dog looks at the treat (and therefore the calibration dot) we give the treat to the dog. After completing all 13 dots the experimental procedure can begin.

Analyzing DogCam-v1 data. The video stream is converted into frames. Our cameras record 30 frames per second; so in a normal 20-minute recording, 36000 images are obtained. The images are loaded into Matlab in order to be further analyzed. We use the calibration frames to define an 'interest area', this procedure is just a simple pixel cropping using the degrees of freedom of the dog's head movement plus the degrees of freedom of the dog's eye movement when attending to the treat as expressed in the recorded visual field (fig 3). Now within the new reduced resolution region-of-interest (ROI) we can count frames of important events or run a saliency map algorithm (or any other machine vision technique such as Image Segmentation, Event Detection or Object Recognition algorithms) to define important features. Preliminary data



Figure 2. Two frames showing when the dog directs its attention to the treat. Each frame will contribute with 1 point that will help to define a ROI.



Figure 3. Same frame as fig 1. ROI plotted on top defined with 1950 points extracted from the calibration frames. Shape irregularity due to different dog's movement when looking at each treat/dot.

gathered using DogCam-v1 [5] have shown that when dogs are in a normal playing interaction with their owner, the owner's face stays within the ROI more time than any other part of the body. Hands and arms are highly salient as well. Presumably so as to gather information related to what it is going on at a specific moment in a given context. On the other hand when analyzing the same frames on the human eye-tracking data, we could see that humans look at the dog's eye and face during the interaction.

One may argue that dogs may benefit from having eye contact with humans and reading other biological movement produced by the owners. The explanation may be simple: domestic dogs live among humans and most of the time they are also reared among us. So they might have been pushed, evolutionarily speaking, to look where humans are most expressive, namely in the face. This might be an indicator of the dog's cognitive skills. If they actually developed the ability to make eye contact with humans and to react to this contact, several of the new data about doghuman interactions would thereby be explained.

After obtaining the Bloomington Indiana Animal Care and Use Committee (BIACUC, protocol number 09-002) approval, we ran several experiments aimed both at obtaining data and at learning what adjustments would have to be made for a successful head-mounted camera experiment with dogs.

DogCam-v2. Due to the questions raised by the DogCamv1 data and the technical restrictions it has, we recently developed the next version. This model resembles a human eye-tracking system in its totality. The harness of the original DogCam was replaced by a set of Doggles, goggles for dogs, a popular pet accessory [9]. DogCam-v2 consists of two cameras, one of them is pointing to the world and the other one is pointing to the dog's eye. The accommodation period is done with the new model as well, in this case the dog has to get used to wearing the doggles, which takes almost no time (most owners have reported that their dogs seem to wear the doggles comfortably after just a couple of days).

Calibrating DogCam-v2. The calibration procedure is exactly the same that we explained in the DogCam-v1 calibration section.

Analyzing DogCam-v2 data. Now that DogCam is an actual eyetracker. We use ExpertEyes open source eyetracking application to pre-process the data [1]. Now that human and dog participants are wearing eye-tracking systems we can apply the very same analysis techniques.

Right now we are gathering data with our new DogCam model in a simple playing situation and we expect to successfully combine the data with the human-eyetracker one. In the future we may want to explore interactions between conspecifics or create more sophisticated experiments in order to explore an isolated behavior or investigate some other interactions between humans and dogs.

CONCLUSION

Research on animal cognition gives some light to how cognitive processes work in different species and therefore may allow us to better understand human evolution through comparative analysis. Thus, it is not surprising that the literature in animal cognition has increased in number of papers submitted, methodologies and species used. Nevertheless new improvements in research methods are still needed to address complex questions about complex behaviors in animals.

The goal of this project is to develop equipment and techniques that can be deployed experimentally in future studies.

- 1. Akavipat, R., Parada, F.J. Busey, T.A. ExpertEyes: An OpenSource Eyetracking Application. [Online] Available http://code.google.com/p/experteyes/, June 30, 2009.
- 2. Babcock, J.S. & Pelz, J. (2004). Building a lightweight eyetracking headgear, ETRA 2004: Eye Tracking Research and Applications Symposium, 109-113.
- Hare, B., Brown, M., Williamson, C., & Tomasello, M. The domestication of social cognition in dogs. *Science*, 298, 5598 (2002), 1634-1636.
- Lieberman, M. D. (2010). Social cognitive neuroscience. S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds). *Handbook of Social Psychology* (5th ed.), New York, NY: McGraw-Hill, 143-193.
- 5. Miklósi, Ádám. (2007). *Dog behavior, evolution and cognition*. Oxford: University Press.
- 6. Rossi, A. Parada, FJ. & Allen, C. Dogcam: A way to measure visual patterns in social interactions in dogs. In

Proceedings of the 31st Annual Conference of the Cognitive Science Society, 2009.

- Tomasello, M., Call, J. & Hare, B. Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, 55 (1998). 1063-1069.
- Topál József., Miklósi Á., Gácsi M., Dóka A., Pongrácz P., Kubinyi E., Virányi Zs., & Csányi V. The Dog As a Model for Understanding Human Social Behavior. Advances in the Study of Animal Behaviour, 39 (2009). 71–116
- 9. http://doggles.com/.

A Dyadic Approach for Measuring and Testing Agreement in Interpersonal Perception

Antonio Solanas antonio.solanas@ub.edu +34 933125076 David Leiva dleivaur@ub.edu +34 934031137

Rumen Manolov rrumenov13@ub.edu +34 934031137

Faculty of Psychology. University of Barcelona Pg. Vall d'Hebron, 171, 08035-Barcelona, Spain

ABSTRACT

The present research deals with developing a new index to measure interpersonal perception agreement at group level. This index is decomposed into components to obtain dyadic and individual effects. Unlike other indexes and statistical methods founded on correlation coefficients to quantify interpersonal perception relationships and agreement, the procedure proposed is based on interpersonal perceptions discrepancies among individuals. Specifically, this study deals with interpersonal perception measurements obtained by means of p-point rating scales where p is the number of response options, for instance, Likert response scales. Apart from descriptive statistical analysis, the procedure enables to obtain statistical significance by means of Monte Carlo sampling. The statistical method proposed can be useful for social researchers interested in studying group and teams, as this procedure allows them to quantify interpersonal perception agreement at several levels, to obtain statistical significance, and to use these measures as new characteristics for understanding social processes.

Author Keywords

Interpersonal perception agreement; dyadic analysis; group and team processes.

INTRODUCTION

Social interaction investigations have been often founded on individualistic research methods. That is, social researchers commonly measure people one at a time and then apply statistical methods that assume people's actions, beliefs, thoughts, feelings, and perceptions are independent of the others. Nevertheless, it has been pointed out that a psychology of interpersonal relations should take into account the perceptual underpinnings since people's actions towards others are shaped by their perceptions of them [4, 5]. Additionally, individualistic research methods are not suitable for finding patterns of mutual influence and interdependence [1]. Hence, the main drawback of the individualistic approach is ignoring that social phenomena involve more than one person and that is why dyadic methods have been increasingly developed and proposed to analyze social data.

The Social Relations Model [SRM; 6] has been proposed to study interpersonal perception as a two-sided process since each group member assesses the others. Data are decomposed into four components (constant, actor, partner, and relationship effects) and dyadic and generalized reciprocity are estimated. Dyadic reciprocity in interpersonal perception refers to how all pairs of individuals in dyads perceive each other, while generalized reciprocity corresponds to how individuals are perceived by the others and how each individual perceives the others. It should be noted that dyadic and generalized reciprocity, respectively, correspond to dyadic and individual levels of analysis. In the SRM, correlation values are obtained to quantify dyadic and generalized reciprocity in interpersonal perception. The SRM undoubtedly shows psychologists' interest for measuring agreement, although there are other illustrative examples in psychological literature. For instance, a coefficient of judges' agreement for unordered scales has been proposed to measure to which extent nonchance factors are operating in the direction of agreement [2]. More recently, an index has been developed to assess the degree of interrater agreement for ratings of a single item and multi-item scales [3].

According to Kenny [5], dyadic and generalized reciprocity in interpersonal perception are defined and consequently measured as correlations among people's responses. Nevertheless, agreement in interpersonal perception could also be measured as the discrepancy between individuals' interpersonal perceptions, not only as correlation coefficient values. Interestingly, if such a discrepancy measure were developed, social researchers would have two procedures to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

study interpersonal perception. Both measures would enable them to obtain different information about interpersonal perception processes.

The present study is concerned with research designs in which observations have been made on all dyads of a group. Particularly, this work deals with interpersonal perception measurements obtained by means of p-point rating scales where p is the number of response options, for instance, a Likert response scale from 1 to 6. The research was specifically intended to provide a new statistic, based on dyadic discrepancies between rates assigned mutually by individuals, for measuring agreement in interpersonal perception at global level. Additionally, the statistic should be decomposed into parts to estimate dyadic and individual effects, and exact analytical results for the expected value and standard error should be also derived in order to make proper comparisons between theoretical and empirical values. Moreover, as statistics' exact distribution has not been derived yet, a statistical test founded on Monte Carlo sampling is proposed.

MEASURING AGREEMENT IN INTERPERSONAL PERCEPTION

The statistic proposed for measuring agreement in interpersonal perception is as follows:

$$\eta = \frac{2\sum_{i=1}^{n}\sum_{j=i+1}^{n} (x_{ij} - x_{ji})^{2}}{n(n-1)(p-1)^{2}}, \quad n \ge 2, p \ge 2,$$

where x_{ij} , n, and p respectively denote how individual i perceives individual j, group size, and the maximum value for *Likert* scales. This statistic ranges from 0 to 1. If η equals 0, it means that all individuals give others the same score they receive from them, that is, there is agreement in interpersonal perception. On the other hand, the maximum value of discrepancy among participants corresponds to $\eta = 1$. In order to decompose dyadic effects, note that

$$\eta = \frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} 2(x_{ij} - x_{ji})^{2}}{n(n-1)(p-1)^{2}} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \eta_{ij}$$

where η_{ij} denotes dyads' contributions to overall interpersonal perception disagreement, being $\eta_{ij} = \eta_{ji}$. Similarly, individual effects can also be decomposed as follows:

$$\eta = \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \ j \neq i}}^{n} (x_{ij} - x_{ji})^{2}}{n(n-1)(p-1)^{2}} = \sum_{i=1}^{n} \left(\frac{\sum_{\substack{j=1 \ j \neq i}}^{n} (x_{ij} - x_{ji})^{2}}{n(n-1)(p-1)^{2}} \right) = \sum_{i=1}^{n} \eta_{i}$$

where η_i denotes individuals' contributions.

EXPECTED VALUES

In order to obtain expected values and statistical significance is needed to suppose a probability distribution for individuals' ratings. Discrete uniform distribution is here proposed as it corresponds to the lack of preference. Thus, it can be proved the expected values for the global index are:

$$E[\eta] = \frac{p+1}{6(p-1)}$$
$$SE[\eta] = \sqrt{\frac{7p^4 - 20p^2 + 13}{90n(n-1)(p-1)^4}}$$

These expected values allow researchers to carry out proper compassion since the bias and variability of the statistic are known.

STATISTICAL INFERENCE

For statistical testing of agreement in interpersonal perception at group level, the null hypothesis states that each group member randomly assigns ordered values to the others. It has been often assumed in related statistical techniques that random variables x_{ij} follow a discrete uniform distribution. Nevertheless, there is no theoretical proof to discard other discrete mass probability functions for modeling group members' responses (e.g., triangular and U-shaped distributions). Hence, exact sampling distribution depends on mass probability functions, group size, and the range of the *Likert* scale. Table 1, whose values have been analytically obtained, shows the exact sampling distribution of the global statistic for n = 3 and p = 4, assuming a discrete uniform distribution.

For practical purposes, obtaining the exact distribution for the statistic has no sense since it depends on the mass probability function that random variables x_{ij} follow. For this reason, a more useful statistical method is required. Monte Carlo methods are a way of solving the problem and estimating probabilities, that is, p values. We have developed software with R language, which is delivered upon request, for testing agreement in interpersonal perception. Monte Carlo sampling allows researchers to obtain statistical significance for each index, that is, at group, dyadic, and individual levels. It is only required to specify group size, maximum values for *Likert* scale, and mass probability functions for each x_{ij} . Finally, users are also required to establish the number of simulated groups or, if preferred, samples.

| Statistic value | <i>p</i> -value |
|-----------------|-----------------|
| 1.0000 | .001953 |
| .8148 | .013672 |
| .7037 | .031250 |
| .6667 | .042969 |
| .6296 | .066406 |
| .5185 | .136718 |
| .4815 | .183593 |
| .4444 | .199218 |
| .4074 | .251952 |
| .3704 | .322264 |
| .3333 | .416014 |
| .2963 | .462889 |
| .2222 | .568358 |
| .1852 | .708983 |
| .1481 | .755858 |
| .1111 | .808592 |
| .0741 | .914061 |
| .0370 | .984373 |
| .0000 | 1.000000 |

Table 1. p-values associated to global agreement statistic values.

EXAMPLE

Suppose a group of n = 3 individuals has been formed to develop a new product. After the task the participants were asked to score each other, by a rating scale from 1 to 4 (i.e., p = 4) in the following question: 'She/He profited the time available to solve the task.' The following matrix shows the fictitious data:

$$\mathbf{X} = \begin{pmatrix} 0 & 4 & 4 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$$

The global statistic can be computed as follows:

$$\eta = \frac{2(9+9+0)}{54} \simeq .6667$$

As the mathematical expectancy for the global index equals .2778, the value obtained clearly exceeds it. Furthermore, the exact *p*-value associated to the statistic is approximately equal to .0117. These results suggest that group members' interpersonal perceptions agree less than expected by chance.

Dyadic contributions to interpersonal perception disagreement are $\eta_{12} = 1/3$, $\eta_{13} = 1/3$, and $\eta_{23} = 0$. Hence, dyads that rate each other more dissimilarly are 1-2 and 1-3. As regards individual' effects, $\eta_1 = 1/3$, $\eta_2 = 1/6$, and $\eta_3 = 1/6$. The individual who shows the highest discrepancy between directed and received scores is denoted by *1*.

CONCLUSIONS

We have developed a statistical method that enables social researchers to obtain interpersonal perception agreement measurements at global, dyadic, and individual levels in groups. This statistical method allows social researchers to carry out descriptive and inferential statistical analyses and therefore they can quantify and make decisions concerning agreement in interpersonal perception. Statistical decisions can be taken for all levels of analysis, that is, global, dyadic, and individual levels. An R package has been written to make feasible obtaining statistical significance for all statistics.

- Bond, C.F., Jr., Horn, E.M., & Kenny, D.A. A model for triadic relations. *Psychological Methods*, 2 (1997), 79-94.
- 2. Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20 (1960), 37-46.
- 3. James, L.R., Demaree, R. G., & Wolfe, G. Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69 (1984), 85-98.
- 4. Jones, E.E. *Interpersonal perception*. New York, NJ: W. H. Freeman and Company, 1990.
- Kenny, D.A. Interpersonal perception. A social relations analysis. New York, NJ: The Guilford Press, 1994.
- Kenny, D.A., & La Voie, L.J. The social relations model. In L. Berkovitz (ed.), *Advances in experimental psychology*. San Diego, CA: Academic Press (1984), 142-182.

Acquisition, Translativity and Transfer of a Matching to Sample Task Under Different Language Modes

Daniel Gómez Fuentes d1031443576@aol.com T. 52-288-8158619 Emilio Ribes Iñesta eribes@uv.mx +52-228-8143498

Melgar y Juan Escutia s/n, Colonia Revolución C.P. 91100, Xalapa, Veracruz, México

The linguistic modes during its acquisition are identified as complementary pairs (King & Quigley, 1985; Marschark, 1993; Rondal, 1980). These pairs are named active and reactive based depending on its functional relationship. The acquisition of the reactive modes precedes the active one and it is necessary to give feedback to the precision and efficiency of the active modes. The pairs gesticulating/ pointing-observing, writing-reading and speaking-listening can therefore, be identified. The reactive modes represent functions of the individual such as reader listener and observer. The active modes work as mediators of other modes. They represent the individual's actions such as: pointer/gesticulator, speaker and writer (Ribes, 1990; Gómez Fuentes, & Ribes, 2008).

The purpose of this study was to evaluate the acquisition, translativity and transference of the performance of the language modes in two different arrangements related to conditional discrimination, matching to sample of first and second orders, using direct and arbitrary relationships among stimuli and techniques to evaluate the effect of the intrinsic feedback provided by the reactive modes on the active ones.

METHODS AND TECHNIQUES USED

A pre-test and post-test design are used, five training sessions and three transference tests for each language mode. The language modes during the training, and the transference conditions were presented in three different sequences. The order for presenting the transference tests in each mode remained constant in all the experiments. The participants were 96 children experimentally unaware, aged 8-12, from fourth grade of elementary school of Veracruz, México. The program, tasks and data gathering were designed by using the Toolbook Instructor of Windows. The independent variable used was the intrinsic feedback: the presence of the intrinsic feedback performed by the reactive mode before the active one or its absence when the active modes are performed without the active ones. The dependent variables used were: 1) speed and completion level of the acquisition of behaviors during the first stage in the training sequence; 2) translativity of a same kind of performance from one mode to another during the training sequence; and 3) transference of behaviors as a result of the previous training with stimuli, modalities or diverse criteria.

A first package of four experiments was carried out, then a second package of four experiments. In the two packages, experimental arrangements were used: matching to sample of first and second orders, under the active modes with or without reactive feedback. The only difference was the kind of relationship between the stimuli used and the tasks. In the first package, criteria of direct matching among stimuli were used. In the second one, an arbitrary matching only.

The tasks of the first order were designed using a sample stimulus and three for comparison. One of the comparing stimuli was identical to the sample, others were similar in shape, color or size, and the third was different in shape, color or size. In a direct matching, geometrical shapes were used. In the arbitrary matching, the geometrical shape stimulus was substituted by a rectangle that included numerical stimuli, roman and arabic numbers. The first ones were able to identify the shape and the second ones, the color or the size of the figure.

In the matching to sample of the second order, two stimuli were presented in the upper part of the screen, indicating the matching criteria. In the middle part of the screen, a sample stimulus was used, and in the lower part, four comparing stimuli. One of the comparing stimuli was identical to the sample; two were similar in shape, color or size. In direct matching, some geometrical stimuli were used. In the upper part of the screen, in arbitrary matching, one of the geometrical stimuli was substituted by a rectangle that included roman and arabic numbers. In the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

middle part, the sample geometrical stimulus was substituted by roman and arabic numbers.

The language modes in conditions of training were presented in three possible sequences, each one of them integrated by the three modes. In addition to the pretest and posttest, each experiment included transference tests. In each experiment 36 trials by session appeared, five consecutive sessions for each phase of training and a session for each test. In each session, half of trials were of similarity and other half by matching the different ones, except in the tests of extra relational transference that they used inclusion and exclusion matching. In the matching to sample of first order, the sessions were divided in two parts, each one of 18 trials; in one of them the trials of matching by similarity in color and forms and the other were distributed at random, those of difference.

In each session, when finalizing half of the trials, the computer system presented a screen that informed to the subject that the matching criterion would change. Under matching to sample of second order, the presentation of the trials during the training and trials of transference were at random. In the training, the subject received feedback of trial to trial and the number of successes at the end of the session. In the transference tests, at the end of the session, the subject received information on the number of successes obtained. Each session was preceded by appropriate instructions.

EXPERIMENTAL METHOD

Pre-test and Post-test. These tests were presented in the first and last sessions of each experiment. In the pre-test and post-test the stimuli were similar to the ones used during the training phase. The subject used the mouse of the computer placing the arrow that appeared in the screen on the chosen figure and pressed the left button of the mouse. In these tests no information was provided to the children in relation to what was correct or incorrect about their answers.

Training phase: Writing/Reading. The subject chose a stimulus of comparison based on the criterion of matching established by the task. The features of the chosen figure were written on rectangles placed in the lower part of the screen. In the first rectangle, the shape was written and in the second, the color. In the active mode without reactive feedback, the writing response was carried out without reading what was written due to the fact that the rectangle where it was written was black. When the reactive feedback was used before the active mode, the subject read what he/she wrote simultaneously on a white rectangle.

Training phase: Speaking/Listening. In this phase, the subject was trained to speak/listen to the features of the figures. When the active mode was carried out without the reactive feedback, the subject listened to a tune with sounds of nature through the earphones with functions of "white

noise" while he/she pronounced the shape of the selected figure. After that, using the same procedure pronounced the color of the figure. When the reactive feedback was used, the tune with noise functions was eliminated. The speaking answer was achieved pronouncing in a loud voice and listening at the same time the shape or the color of the selected figure.

Training phase: Pointing/Observing. When the active modes were achieved without the reactive feedback, the task consisted of observing the figures in a first screen briefly (of the first or second orders). Soon after that, in a second screen, the task consisted of pointing with the mouse at the blank space that corresponded to the selected figure. When the reactive feedback was incorporated, the subject observed the task and responded by pointing out in the same screen.

Transference tests. The transference tests were applied in three different subsequent sessions, one session for each test, after each training phase. In the first one, the subject responded at new geometrical figures with a different color. In the second one, the same stimuli of the training process were used. The shape of the figures remained and the size was substituted as modality. In the third one, geometrical and numerical figures were included with or without stripes and shade. Inclusion and exclusion relationships were used. When each test was concluded, the subjects received information about the obtained score of their right answers.

CONCLUSIONS

The methods and techniques used have demonstrated the effect of the intrinsic feedback provided by the reactive modes on the active and confirmed the prediction that the active and reactive modes are complementary. The analysis of these processes can provide information about the development of the linguistic modes and its application to human behavior analysis, as well as to the development of methods and techniques in behavioral research to promote learning and the functional use of the linguistic modes.

- 1. Gómez Fuentes, A.D., & Ribes E. Adquisición y transferencia de una discriminación condicional del primer orden en distintos modos del lenguaje. *Acta Comportamentalia*, *16*, 2 (2008), 183-209.
- 2. King, C.M., & Quigley, S.P. *Reading and deafness*. San Diego, CA: College Hill Press, 1985.
- 3. Marschark, M. *Psychological development of deaf children*. New York: Oxford University Press, 1993.
- 4. Ribes, E. El lenguaje como conducta. Mediación funcional versus descripción morfológica. *Psicología General.* México: Trillas (1990), 147-176.
- 5. Rondal, J.P. Father's and mother's speech in early language development. *Journal of Child Language*, 7 (1980), 353-369.

Early Social Skills as a Function of Gender: An Observational Study of Children Between 3-8 Years in Andhra Pradesh, India

Mayuri Kandalla

Angr Agricultural University, College of Home Science, Hyderabad 500 004, India pramaynn@yahoo.com

Pavitra Bhat

Angr Agricultural University, College of Home Science, Hyderabad 500 004, India

ABSTRACT

Children's Social skills were observed through video filming of their interactions in school and apartment play grounds and summer camps. In all 72 video clippings of 10 minutes duration each were analyzed using Observer Behaviour Software XT 7.0., by developing a wide range of coding scheme for conversations, body language, gender related behaviours, and for acceptance, rejection, and resistance strategies and unsociable actions. Data in frequencies and duration was analyzed using two way ANOVA.

Author Keywords

Early social Skills, gender differences, conversation, body language, pro social behaviours, unsocial behaviours, acceptance –rejection strategies, resistance strategies.

INTRODUCTION

Social experiences with peers constitute an important developmental context for children. In these contexts, children acquire a wide range of behaviours, skills, attitudes and experiences that influence their adaptations during the life span. One of the major tasks of the early childhood years is to learn positive and socially acceptable ways of interacting with others. As much of this learning occurs within the context of the peer group, positive peer interactions make a substantial contribution to children's socioemotional and cognitive development beyond the influences of family, school and neighbourhood.

Relationships with peers have significant importance in the lives of even very young children by allowing them to experiment with roles and relationships and develop social cognitive and behavioural skills (Asher, 1990; Rubin & Asendorpf, 1993).

Children's interactions with one another do endure over a

long period of time and are vital for normal social development. With increasing age, play partners become better able to agree with each other about the roles, rules and themes of their pretence. They are also better able to maintain their play interactions by adding new dimensions to their expressed ideas.

"... the single best childhood predictor of adult adaptation is not IQ, not school grades, and not classroom behaviour but, rather the adequacy with which the child gets along with other children. Children who are generally disliked, who are aggressive and disruptive, who are unable to sustain close relationships with other children and who cannot establish a place for themselves in the peer culture are seriously "at risk" (Hartup, 1997).

Observation of Early Social Skills Is Important:

- Studying children's behaviour in naturalistic settings helps understand their control over their own peer culture, coping capabilities when there is very little adult help.
- Interaction with peers can create internal individual disequilibrium implying that cognitive conflict helps in the development of social understanding and learning. Piaget noted that, 'one may conceive of co-operation as constituting the ideal form of equilibrium towards which society tends when compulsory conformity comes to break down'.
- Verba (1994) concedes that the child needs to be an active participant in a situation where social interaction takes place. She emphasised the value of peers' social interactions on cognitive competence as the children are at relatively equal status and competency in comparison to those of adults.
- Bronfenbrenner and Crouter (1983) observe, "In the light of the increasing evidence for the influence of peer group on the behaviour and psychological development of children... it is questionable whether any society... can afford to leave largely to chance the directions of this influence and realization of its high potential for fostering constructive development both for the child and his/her society".

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Early social skills need studying to assist children in improving their social skills such as communication, interpersonal interaction and social analysis abilities to fit with the set patterns of their culture.

Naturalistic Observation social interactions among children has been a rare area of study in the Indian context. This study was undertaken to examine naturally occurring social interactions among children 3-8 years in their school settings and informal groups in apartment building parks.

Objectives:

To observe and analyze social interactions of children in the age group of 3- 8 years in their natural/ informal settings for gender differences.

REVIEW OF LITERATURE

Though children grow in similar kind of ecological systems, and are almost exposed to similar kind of experiences; differences in parenting styles, stereotypic gender conceptions held by family and society give rise to gender differences in social interactions with peers. Individual temperament as well as biological pruning may also be the reason for gender differences that are evident in the ways the two sexes interact with their peers. Early in childhood children show preference for same sex peer's company. Although girls tend to initiate same sex play more than boys at age 3, this gender difference is reversed and even heightened by age 5 (Pitcher & Schultz, 1983). Children's preference for same-sex play mates is a universal aspect of growing up (Whiting & Edwards, 1988). This relies on similarity based account of identification. Children see themselves as similar to other children, or because girls (or boys) see themselves as similar to other girls (or boys), they identify with, and are influenced by, these similar others or those who share category membership.

Several studies have indicated that girls spend more time in small group social activities, in cooperative and turn-taking games, engage in more person fantasy and are more sensitive to the requirements of collaboration. Boys, on the other hand, prefer to engage in larger group physically active games and rough and tumble play (Walker, Sue 2004, Dorsch & Keane, 1994; Fabes, 1994; Fagot,1985; Jones & Glenn, 1991; Lewis & Phillipsen, 1998; Maccoby, 1988; Mollor, Hymel & Rubin, 1992).

Gender segregation is most likely to occur when children are not with adults but with others of a similar age. Gender segregation involves not only the preference of same-sex playmates but also the avoidance of playmates of the other sex. Serbin, Powlishta and Gulko (1993) found that there was increase in preference for same-sex peers among kindergarten children and children in the early school years.

The nature of relationships also differs between the sexes. Whereas girls' friendships are characterized by emotional and physical closeness, the friendships of boys are founded on shared activities and interests (Maccoby,1988). Boys and girls not only differ in their conflict resolution skills, but also differ in their use of antisocial trajectories with peers. As boys are usually found to use aggressive strategies more often than girls, they may be expected to stand high on antisocial behaviour. The influence of friends and poor peer relations play a crucial but different role in the development of antisocial behaviour among boys and girls.

Culture has its own role to play in the gender typed behaviours. As Putallaz et al. (1995) suggest socialisation and cultural processes may act to discourage overt conflict behaviour by girls while encouraging the use of affiliative conflict strategies designed to minimise disruption. Similarly, overt conflict involving aggression may be not only acceptable for boys but positively valued as a means of establishing their social position if it is used as means for standing up for oneself.

Do these various findings hold true in the Indian context too, or do cultural differences bring out differences in early social interactions among children? This study tried to explore the issue.

METHODS

The city of Hyderabad, Andhra Pradesh, India was selected for the study as there were a number of schools with cosmopolitan environments wherein children from various backgrounds, cultures, and wide socioeconomic groups could be observed interacting with each other.

Sampling Procedure

Purposive sampling technique was used to gather data in order to video film social interactions of 3-8 year olds. The video recordings were taken in the following settings from five different schools, two apartment play grounds and two summer camps.

School settings: Outdoor play, Indoor play: block play, pretend play and snack time.

Around 85 video clippings were shot, but only 72 clippings were selected for the study in order to

- Systematically distribute the samples in various play settings according to objectives of the study.
- Select children interacting normally without being self conscious about being videotaped.

Coding Schemes Developed and Used

- Coding scheme for conversations both sociable and unsociable conversations, content
- Coding scheme for **body language social**, **unsocial**, **physical proximity**, **facial expressions**
- Coding scheme for gender related behaviours sex appropriate, neutral and cross gender behaviours
- Coding scheme for acceptance, rejection, and resistance strategies and unsociable actions.

Coding Scheme for Conversations

Type of Conversations

- a. **Social conversations** Initiation, Following, Direction, directing peers, Pleasantries, Supporting Encouragement
- b. Unsociable conversations Quiet, Interruption, Arguments, Gossiping, Teasing, Domination

Content of Conversations

- a. **Social content-**Social speech, Common interest topics, Fantasy, Words of gratitude, Daily events, Secrets, Requests, Compliments, Approvals
- b. Unsociable content-About self, Abusive words, Words of rejection, Disapprovals

Coding Scheme for Body Language

Type of Activity

- a. Social activity Active, into groups
- b. Unsociable activity Lethargic, Outside group, Dispirited
- c. **Physical proximity-** Holding hands, Hugging, Patting etc
- d. **Facial expressions- Social facial expressions** Pleasant/happy, smiling Excited, Surprised, Laugh
- e. Unsociable facial expressions–Sad, angry, threatening etc

Coding Scheme for Gender Related Behaviours - Sex appropriate behaviour Encouraging sex appropriate behaviour, Condemning sex in appropriate behaviour, Cross gender roles, Encouraging cross gender behaviour

Coding Scheme for Acceptance, Rejection and Resistance Strategies and Unsociable Actions

- a. Acceptance and co-operation strategies Showing interest/gaining entry into play, Imitating play – trying to behave similarly as that of peers, Striving for attention, Teasing in a friendly manner Recognition of leader – selecting a leader with common opinion, Competing with leader, Using friendship to bargain or negotiate, Team work – working or playing together in a group towards a common goal, Sharing, Explaining rules of play
- b. **Rejection strategies -** Group rejection- preventing entry, Indirect rejection, Denying friendship, Claims of ownership
- c. **Resistance strategies** -Physical resistance, yelling resistance, Disputes over toy or roles, Threatening, Disturbing others play

The coding also included 19 modifiers – modifiers were persons, objects or behaviours that formed the reason for exhibiting the particular behaviour with the peers during their interactions. E.g.: children expressed anger because of other's rejection. Here, "anger" is behaviour whereas "others rejection" is the modifier of the behaviour "angry".

📅 H:\3-4 years (nursery)\outdoor clippings\3-4 yrs 0... 🔀



Figure 1. Data visualization window.

Observation

The video tapes were loaded into the computer and observations made with The Observer XT 7.0, which gives data on frequency and duration of behaviour occurrences. The data was analyzed using two-way ANOVA to find out differences in gender differences in different settings.

RESULTS

Some of the Results Emerged as Follows

- Sex appropriate behaviours increased with age whereas, gender neutral behaviours decreased with age during pretend play.
- The social behaviours of girls were very high in frequency and duration as compared to that of boys during pretend play.
- Children of both genders used social conversations, social body language and acceptance strategies for almost same duration during block play.
- The major contents of conversation among boys were pleasantries and self talk whereas girls used pleasantries and discussions on daily events.
- Physical proximity towards peers was high among boys of 3-5 years which was over taken by girls during 5-8 years.
- Social skills refined with age; with boys using humorous body language and girls using touch as support.
- Boys used humour to gain attention, as an important acceptance strategy whereas sharing and playing together as a team formed an important acceptance strategy among girls.
- Overall unsociable behaviours, such as self centered speech were present more frequently and for longer

duration among boys as compared to girls. Girls used dominations and arguments.

- Boys used aggressive body language more frequently whereas girls used rejections more often
- Girls were found to be more flexible in the use of gender related behaviours whereas boys were more sex stereotypic.
- Girls usually picked up adult roles and pretended daily events more whereas boys picked up fantasized roles such as super hero.
- Both boys and girls used acceptance strategies more frequently with same gender peers in all the play settings.
- Both boys and girls remained out of the group when rejected by opposite gender peers whereas they expressed anger and rejection towards same gender peers.
- Boys were more provocative towards opposite gender peers during all the play settings.
- In general children used indirect rejection strategies more frequently with same gender peers and physical resistance and unsociable actions frequently with opposite gender peers.

CONCLUSION

The present observational study serves as an important basic research in the area of peer interactions, which brings out the general trends in social and unsociable behaviours of children, during their interactions with peers in natural play settings. The study focuses on the social skills of young children who are trying to expand their horizon of social network.

- 1. Asher, S.R. Recent advances in the study of peer rejection. Social Behaviour and Peer Acceptance. In Asher, S.R. & Coie, J.D. (Eds.). *Peer rejection in early childhood* (1990), 3-14.
- Bronfenbrenner, U., & Crouter, A.C. The evolution of environmental models in developmental research. In P. H. Mussen (Series Ed.) & W. Kessen (Vol. Ed). Handbook of Child Psychology: Vol. 1. History, theory and methods (4th ed.). New York: Wiley (1983), 357-414.
- 3. Dorsch, A. & Keane, S.P. Contextual factors in children's social information processing. Developmental Psychology, *30*, 5 (1994), 611-616.

- 4. Fabes, R.A. Physiological, emotional and behavioural correlates of gender segregation. *New Directions for Child Development*, 65 (1994), 19-34.
- Fagot, B.I. Beyond the reinforcement principle: Another step towards understanding sex roles. *Developmental Psychology*, 21 (1985), 1097-1104.
- Hartup, W.W. Peer Relations. In E.M. Hetherington (Ed.), Handbook of Child Psychology. Socialization, personality, and social development, New York: Wiley & Sons, 1983.
- 7. Jones, A. & Glenn, S.M. Gender differences in pretend play in primary school group. *Early Child Development and Care*, 77 (1991), 127-135.
- 8. Lewis, T.E. & Phillipsen, L.C. Interactions on an elementary school playground: Variations by age, gender, race, group size and playground area. *Child Study Journal*, 28, 4, (1998), 309-321.
- 9. Maccoby, E.E. Gender as a social category. Developmental Psychology, 24, 6 (1988), 755-765.
- Mollor, L.C., Hymel, S. & Rubin, K.H. Sex typing in play and popularity in middle childhood. *Sex Roles*, 27, 7/8 (1992), 331-353.
- 11. Pitcher, E.G., & Schultz, L.H. *Boys and girls at play: The development of sex roles.* South Handley, MA: Bergin and Garvey, 1983.
- Putallaz, M., Hellstern, L., Sheppard, B. L., Grimes, C. L. & Glodis, K. A. Conflict, social competence, and gender: Maternal and peer contexts. *Early Education and Development*, 6, 4 (1995), 433-447.
- Serbin, L., Powlishta, K., & Gulko, J. The development of sex- typing in middle childhood. *Monographs of the Society for Research in Child Development*, 58 1993, 955-963.
- 14. Verba, M. The beginnings of collaboration in peer interaction. *Human Development*, 37 (1994), 125-139.
- 15. Walker, S. Teacher reports of social behaviour and peer acceptance in early childhood: Sex and social status differences. *Child Study Journal*, *34*, 1 (2004), 13-28.
- Whiting, B., & Edwards, C. Children of different worlds: The formation of social behaviour. Cambridge, MA: Harvard University Press, 1988.

Measuring Potential Cues for Depression in Adolescents

Martijn Balsters M.J.H.Balsters@uvt.nl **Emiel Krahmer** E.J.Krahmer@uvt.nl Marc Swerts M.G.J.Swerts@uvt.nl

Department of Communication and Information Sciences, University of Tilburg, PO Box 90153, 5000 LE Tilburg The Netherlands Ad Vingerhoets Department of Clinical Psychology, University of Tilburg PO Box 90153, 5000 LE Tilburg, The Netherlands Vingerhoets@uvt.nl

ABSTRACT

So far, depression research has mainly relied on adult data and self-reports. For better prevention and treatment, it could be of great importance to diagnose additional behavioral cues in earlier stages of life. We therefore analysed video recordings of adolescents (low versus high risk groups with regard to developing depression) performing a free speech task using linguistic, nonverbal and acoustic methods of measurement. High risk adolescents used fewer words and talked more about home and family than their low risk counterparts, suggesting possible relations with introvert and attachment-related behavior. Nonverbally, females showed more affiliation and flight behavior than males, possibly reflecting genderspecific coping strategies. Moreover, interactions show gender differences between risk groups; high risk females use less words and more flight behavior than low risk females, whereas males show the exact opposite pattern. These findings might contribute to better prediction and treatment of depression in the future.

Author Keywords

Depression, behavior, verbal, nonverbal, acoustic, adolescents, prediction.

INTRODUCTION

Mood disorders currently cover a substantial part of mental health problems of the population, with major depression disorder (MDD) being the most significant one. Worldwide, 8 to 12 percent of all people experience at least one episode of depression during their lifetime [8].

First depressive episodes typically develop during adolescence [11]. Early diagnosis of depression could be of great importance for prevention and improvement of

clinical treatment. In this study, we therefore analyzed video recordings of adolescents (belonging to low or highrisk groups with regard to depression) on linguistic, nonverbal and acoustic behaviors in order to identify behavioral cues with potentially predictive value with regard to depression.

The current way of diagnosing depression is on the basis of the well-known DSM–IV criteria, which use a great majority of diagnostic tools (interviews and questionnaires). The diagnosis is mainly based on self-reports of the patient on how he or she feels, behaves, and thinks. The assumption hereby is that the respondent is both honest and able to adequately verbally report on these issues. An additional assumption is that these aspects of behavior are the only or most important for clinical decision making.

However, there are also additional behavioral features, which distinct between depressed individuals and healthy people, which, however, do not have the official status of symptoms. Research showed that, for example, depressed individuals use more self-references and negatively valenced words [12]. On a nonverbal level, they showed a global restriction of expression, reflected by decreased motor activity, body movements and reaction times [13,16]. On an acoustic level, voice analyses have shown that depressive patients display less pitch variability, speak faster and use more pause time when speaking [1,10]. We expected these or similar findings to manifest themselves in adolescents belonging to the high-risk group, thereby distinguishing themselves as possible predictors of depression.

METHODS

Sample

The sample was taken from the TRacking Adolescents' Individual Lives Survey (TRAILS), a large prospective cohort study in which Dutch adolescents are measured biennially until at least the age of 25.

The data used for the current analyses were collected during the third assessment wave, conducted between September 2006 and December 2007. At this wave, approximately 750 adolescents (age 15-17 years) were measured while

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

performing extended laboratory tasks. Based on demographic data (temperament scores, parental psychopathology and living in a single-parent family situation), each individual adolescent was assigned to one of eight risk-groups varying from score 0 (low-risk for developing depression) to 7 (high-risk for developing depression).

Our analyses were done using video recordings in which the adolescents had to perform the so-called Groningen Social Stress Test (GSST); a standardized protocol based on the Trier Social Stress Task [9]. Its goal is to induce moderate performance related social stress by the inclusion of a public speaking task and mental arithmetics.

Test Procedure

We analysed video recordings of 20 "low-risk" and 20 "high-risk" adolescents (50% female) while they performed the GSST. Data collection took place in sound-proof rooms with blinded windows at the participants' town of residence. The social stress test was always the last part of a three hour test session covering various other psychological and medical tests.

Participants were instructed to prepare a six-minute nonstop speech about themselves and their lives and to deliver it in front of a video camera. Preceding their performance, they were told they would be judged on several aspects, such as content of speech as well as the use of their voice and posture. In order to induce a threat of social rejection, it was also suggested that they would be rank-ordered by a panel of peers afterwards. The test leader watched the whole performance carefully and took care not to show any sign of empathy or encouragement.

Analyses

All analyses were conducted blind; the experimenter and analyst never knew what group the adolescent belonged to.

Linguistic Analysis

All words uttered during the given six-minute speech time were transcribed using ELAN software, to analyse the data for possible underlying linguistic dimensions. These annotations were then subjected to the Linguistic Inquiry and Word Count (LIWC) program, which analyses texts on a probabilistic word-by-word basis to a default dictionary of words and word stems [12]. This dictionary comprises over 70 language categories, ranging from linguistic dimensions (such as words per sentences, articles), relativity (e.g. in time and space), psychological processes (emotional and cognitive) and personal concerns. The LIWC analysis produces an output in which the analysed text is given in percentages of total words found along the aforementioned language categories. Since all participants spoke Dutch, we used the 2004 Dutch LIWC-version, which has a dictionary of 6568 words divided over 66 predefined word dimensions. These Dutch dimensions show a high correlation with their corresponding English equivalents [18]. In addition to the analysis of the aforementioned

words and language categories, we also calculated the speaking rate by counting the amount of spoken words per minute.

Nonverbal Behavior Analysis

For the analysis of nonverbal behavior, we used the Ethological Coding System for Interviews (ECSI), developed by Alfonso Troisi [15]. This system was specifically designed to measure nonverbal behavior during clinical interviews using ethological principles.

The version of ECSI we used includes 37 different behavior patterns (varying from hand and body movements to facial expressions) divided over eight behavioral categories. ECSI categories have been shown to discriminate between control and psychiatric patients in clinical interviews, as well as between different psychiatric subgroups [15]. Specifically with regard to depression related research, it has been shown that the behavioral categories used in our study derive from the combination of behavior patterns reflecting a distinctive set of antecedent situations, or of consequent responses or events [3,15]. Based on these findings, we scored the video recordings on 25 behavioral patterns divided over four categories: Affiliation (associated with ECSI behaviors 2-6: e.g. head tilt, smile, raising eyebrows), Flight (behaviors 10-15: e.g. to look away/down, eyes shut, chin to chest), Displacement Behavior (24-32: e.g. handface touching, yawning) and Relaxation (33-37: e.g. settle, fold arms, laugh).

Before the actual analysis, it was made sure that the observer reached an inter-observer reliability (κ) of at least 0.70 for each behavior pattern. This evaluation was based on a sample of 40 different video recordings also rated by another observer.

For the current analysis, we cut out the first minute of the free speech task (from the onset of the first spoken word) and removed the sound. While watching the video, the observer recorded whether each behavioral pattern occurred during these 60 seconds or not. The occurrence of each pattern was scored as "1", no occurrence as "0". Following this, the scores for each individual were then summed up per behavioral category in order to obtain "amounts" of behavior.

Acoustic Analysis

From the video recordings used in the nonverbal behavior analysis, the first minute of audio was isolated and normalized at a maximum amplitude of -3 dB using Audacity software, resulting in 40 audio clips in total. It was made sure only the participant was speaking, without any interruption of the test leader.

The first measure we analysed was average fundamental frequency (F0). In order to obtain this, a CSL pitch contour analysis was run from each individual sound file using PRAAT software. Second, we also calculated speech onset time for each participant. This measure was defined by the
time between offset of the last spoken word by the test leader and the onset of the very first spoken word by the participant.

Statistical Analysis

All statistical analyses were performed using a 2 (low versus high risk for developing depression) x 2 (male versus female) ANOVA.

RESULTS

Linguistic Analysis

Contrary to our expectations, no effects with regard to first person singular words were found between the risk groups (p>0.05). We did find a group effect for word rate (F(1,37)= 4,514, p=.040); low-risk adolescents talk more than high-risk ones. Moreover, we did find group effects for family (F(1,36) = 7,052, p=.012) and home related words (F(1,36) = 5,223, p=.028); adolescents belonging to the high-risk group used these words significantly more than those belonging to the low-risk group.

Figure 1 shows an interaction between group and gender (F(1,36) = 5,802, p=.021) with regard to the number of words used during the whole measuring period. In the low-risk group, female adolescents speak more than males, whereas the high-risk group shows the opposite pattern.

Nonverbal Analysis

No group effects were found on any of the four behavioral categories (all p>.05). However, we did find significant gender effects for affiliation (F(1,36)=4,714, p=.037) and flight behaviour (F(1,36)=13,226, p=.001); overall, female adolescents show more affiliation and flight behaviour than males. Moreover, an interaction between group and gender was found for flight behaviour (F(1,36)=6,339, p=.016),



Figure 1. Mean word use (number of words spoken in 6 minutes) compared between a) males and females and b) low versus high risk of developing depression (n=40).

(Figure 2). Men show less flight oriented behavior in the high-risk compared to the low-risk condition, whereas women display exactly the opposite behavior.

Acoustic Analysis

In addition to the expected difference in fundamental frequency between males and females (F(1,37)=209,261, p=,001), no effects were found for average fundamental frequency and speech onset time for both condition and gender (all p>.05).

DISCUSSION

Our analyses yielded promising results with regard to linguistic and nonverbal features. Adolescents who have a low risk of developing depression talk more during the first minute than their high-risk counterparts. These results are in line with earlier findings, showing that introverted individuals have a higher probability of being depressed [2,6].

Moreover, family and home related words were used more in the high-risk condition, showing possible attachmentrelated behavior. Previously, Irons and Gilbert [5] showed that insecure attachment in adolescents is positively correlated with submissive behaviour, depression and anxiety symptoms. Since the high risk-adolescents in our study comply with insecure attachment characteristics (for instance, growing up in a single-parent family), our findings are likely to correspond with these earlier data.

The interactions found on both linguistic and nonverbal analyses show both decreased word use and increased flight behaviour for females in the high-risk condition. Contrary to this, high-risk male adolescents talk more in and display significantly less flight behavior. In line with this, earlier research showed that depressed or homesick boys tend to



Figure 2. Average amount of nonverbal behavior in the ECSI category Flight compared with a) males and females and b) low versus high risk of developing depression (n=40).

display more aggressive and extravert behavior compared to girls [14], possibly indicating gender-specific coping strategies. The main gender effects on a nonverbal level for increased affiliation and flight behavior in women compared to men, might reflect this as well.

In conclusion, these multilevel measurements show interesting results which might be valuable to early detection or prediction of depression. Whether these findings convey predecessors or actual symptoms of depression remains open for further exploration.

ACKNOWLEDGMENTS

We would like to thank prof. dr. Johan Ormel for providing us the video recordings. Furthermore, we thank Hanna Zijlstra en prof. dr. Rinie Geenen for providing us the Dutch LIWC-dictionary.

REFERENCES

- 1. Alpert, M., Pouget, E.R., & Silva, R.R. Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders*, *66*, 1 (2001), 59-69.
- 2. Barnett, P,A., Gotlib, I, H. Psychosocial functioning and depression: Distinguishing among antecedents, concomitants, and consequences. *Psychological Bulletin, 104*, 1 (1988), 97-126.
- Branningan, C.R., and Humphries, D.A. Human nonverbal behaviour: a means of communication. In: Blurton Jones N, editor. *Ethological studies of child behaviour*. Cambridge, Cambridge University Press (1974), 37-64.
- Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., and Snyder, P.J. Voice acoustical measurement of the severity of major depression. *Brain and Cognition*, 56, 1 (2004), 30-35.
- 5. Irons, C., and Gilbert, P. Evolved mechanisms in adolescent anxiety and depression symptoms: the role of the attachment and social rank systems. *Journal of Adolescence*, *28*, 3 (2005), 325-341.
- 6. Jasper, H.H. The measurement of depression-elation and its relation to a measure of extraversion-introversion. *Journal of Abnormal and Social Psychology*, 25, 3 (1930), 307-318.
- Kendler, K.S., Kuhn, J., and Prescott, C.A. The Interrelationship of Neuroticism, Sex, and Stressful Life Events in the Prediction of Episodes of Major Depression. *American Journal of Psychiatry*, 161, 4 (2004), 631-636.
- 8. Kessler, R.C., Berglund, P. Demler, O. The Epidemiology of Major Depressive Disorder: Results

From the National Comorbidity Survey Replication (NCS-R). *Journal of American Medical Association*, 289, 23 (2003), 3095-3105.

- 9. Kirschbaum, C., Strasburger, C.J., and Langkrär, J. Attenuated cortisol response to psychological stress but not to CRH or ergometry in young habitual smokers. *Pharmacology Biochemistry and Behavior, 44*, 3 (1993), 527-531.
- 10. Mundt, J.C., Snyder, P.J., Cannizzaro, M.S., Chappie, K., and Geralts, D.S. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics*, 20, 1 (2007), 50-64.
- 11. Nolen-Hoeksema, S. Gender Differences in Depression. *Current Directions in Psychological Science*, 10, 5 (2001), 173-176.
- 12. Rude, S.S., Gortner, E.M., and Pennebaker, J.W. Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, *18* (2004), 1121-1133.
- 13. Sobin, C., and Sackeim, H.A. Psychomotor Symptoms of Depression. *American Journal of Psychiatry*, 154 (1997), 4-17.
- 14. Thurber, C.A. The Phenomenology of Homesickness in Boys. *Journal of Abnormal Child Psychology*, 27, 2 (1999), 125-139.
- 15. Troisi, A. Ethological research in clinical psychiatry: the study of nonverbal behavior during interviews. *Neuroscience & Biobehavioral Reviews, 23*, 7 (1999), 905-913.
- Troisi, A., and Moles, A. Gender differences in depression: an ethological study of nonverbal behavior during interviews. *Journal of Psychiatric Research*, 33, 3 (1999), 243-250.
- 17. Troisi, A., Pompili, E., Binello, L., and Sterpone, A. Facial expressivity during the clinical interview as a predictor functional disability in schizophrenia. A pilot study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *31*, 2 (2007), 475-481.
- 18.Zijlstra, H., Van Meerveld, T., Van Middendorp, H., Pennebaker, J.W., and Geenen, R. De Nederlandse versie van de 'Linguistic Inquiry and Word Count' (LIWC), een gecomputeriseerd tekstanalyseprogramma [Dutch version of the Linguistic Inquiry and Word Count (LIWC), a computerized text analysis program], 2004.

Observing the Paraverbal Communication of Coaches in Competitive Match Situations

Marta Castañer mcastaner@inefc.es Catarina Miguel mcastaner@inefc.es

INEFC-Lleida, University of Lleida, Partida Caparrella s/n. 25192, Lleida, Spain

M^a. Teresa Anguera

Faculty of Psychology, University of Barcelona, Campus Mundet, Vall d'Hebrón, Barcelona, Spain tanguera@ub.edu

Gudberg K. Jonsson

Human Behaviour Laboratory, University of Iceland, Reykjavik, Iceland, & Department of Psychology, University of Aberdeen, UK gjonsson@hi.is

ABSTRACT

The present study examines the relationship between the verbal communication skills and paraverbal communication (both kinesic and proxemic) of indoor soccer coaches in competitive match situations, senior women's competitive level. T-patterns detection analysis provides a sequential analysis of data, illustrating the communicative style and flow of each coach.

Author Keywords

Paraverbal communication, competitive match, T-patterns, observational methodology, indoor soccer, coaches' behaviour.

INTRODUCTION

Coaches' behaviour in competitive matches is an interesting subject for observational methodology in sports. The current approach has undergone significant developments in recent decades and is now well integrated within the scientific context. The flexibility and rigour of this methodology makes it fully consistent with the characteristics of the study and it has become a standard approach to observational research [1] especially in the field of motor behaviour [7] and sport [6,8]. Of particular relevance is its multi-dimensional nature, which enables it to be adapted to the successive events of paraverbal behaviour, as well as to each of its components.

In sum, observational methodology can be applied to many different facets of human communication [3, 4, 10, 11], and

the wide range of possibilities it offers enables us to optimise the demarcation of units or the development of *ad hoc* instruments such as SOCOP_Coach [6], by adapting the SOCOP (System for Observing Paraverbal Communication) [5].

AIMS

The aim is to study how coaches convey the whole range of verbal and paraverbal communication during the course of a competitive match. The main purpose of the paper is to illustrate the structure of the SOCOP_Coach observational instrument, which can be applied to any type of coach in competitive situations.

METHODS

The study was consistent with the basic tenets of observational methodology in that the coaches' behaviour was analysed without influencing it (spontaneity of the behaviour), it was studied in a competitive situation (naturalistic context), and the design was idiographic, point and multidimensional (I/ P / M) [2]: idiographic because it was centred on the analysis of different subjects, point because three matches were considered without any one being pre-established, and multidimensional because the *ad hoc* system codes consisted of seven criteria and 23 codes.

The data were derived from the study of two coaches and three competitions, giving a total of 240 minutes of empirical material.

Instrument

The instrument used is SOCOP_Coach. By starting from an already validated observational system (SOCOP) (System for Observing Paraverbal Communication) [5] (Table 1) the present study was able to focus on the empirical component, enabling us to validate the communicative specificity of indoor soccer coaches in competitive situations. Coding was carried out using ThemeCoder [12], while data were analysed using Theme v. 5 [9].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

| CRITERIA | CODES | CRITERIA | CODES |
|---|--|---|--|
| Typology : The information concerns the sort of action to be performed by the athlete. | Instruction (I): The information is supplied with the aim of encouraging future actions. | (continued from the left column) | Observation (VB): The coach watches without working or illustrating the result of the competition. |
| Verbal Communication of | Feedback (F): The reciprocal action is referred as a valuable judgement in accordance with the performance of the athlete. Positive evaluation (EP) : The | ParaverbalCommunicationofMorphology:Theinformation is conveyed viakinesic gestures that aremorphologically defined for | Emblem (EMB): The information is supplied via kinesic gestures which are iconically defined and agreed upon, where verbal language is not necessary. |
| Function : The information given by the coach plays a mediating role with respect to the athlete's performance. | coach makes a favourable judgement of the athlete's performance. | the athlete. | Deictic (DEI) : The information is supplied via kinesic gestures that indicate the location of people and/or objects. |
| | coach makes an unfavourable judgement of the athlete's performance. | | Kinetographic (KIN): Gesture that draws actions or movements in space. |
| | Description (D): The coach describes the way in which the athlete performs or performed their actions. | | Beats (BEA): The information is supplied via kinesic gestures that are iconically defined, in accordance with the |
| | Prescription (P): The coach communicates with and directs an athlete as to how he/she must carry out future actions. | Paraverbal Communication of Posture: The information is supplied | Biped (BI) : The coach remains standing but without moving around. |
| Verbal Communication of Morphology: The information given by the coach is of a given form designed for the athlete. | Interrogative (IRG) : The coach questions the athlete as to his/her performance with the aim of raising his/her awareness of the mistakes made or the correct way | from a given postural position. | Sitting (SEA) : The coach is seated. |
| | to perform the action. Imperative (IMP): The coach tells | | Locomotion (LCM): The coach moves to the technical area. |
| | the athlete firmly what to do or what should have been done in order to draw his/her attention to this aspect | | Alteration of level (ALT): The coach alters his posture with respect to the height of his body. |
| | Exclamatory (EXC) : The coach expresses a strong emotion in response to the athlete's performance. | CommunicationofAdaptation:Theinformation is supplied viakinesic gestures but withoutthe aim of control orillustrationThese gestures | Self-adaptor (SE) : When the teacher maintains contact with other parts of his/her body but without any communicative purpose. |
| ParaverbalCommunicationofFunction:The informationgiven by the coach involvesa kinesic gesture with acommunicative intention. | Regulator (RE): The information is given via kinesic gestures that control and link together the moments of interaction between people. It requires an immediate response from the athlete. | physical contact with other people and/or objects. | Hetero-adaptor (HE): When the teacher maintains bodily contact with other people but without any communicative purpose |
| | Illustrator (IL): The information is supplied via kinesic gestures with the aim of reinforcing the verbal language that is used by the coach and does not require an | | Multi-adaptor (MUL): When several of these adaptor gestures are combined. |
| | immediate response from the athlete. | Table 1 SOCOP-Coa | ch Observation Instrument |

DATA ANALYSIS

The observation of a natural context requires the use of the above-mentioned observational instrument. In-depth analysis is then possible with the detection and analysis of temporal patterns (T-patterns) in the transcribed actions.

How to read the pattern tree graph: The tree graph shows the events occurring within the pattern, listed in the order in which they occur within the pattern. The first event in the pattern appears at the top and the last at the bottom. The pattern diagram (the lines connecting the dots) shows the connection between events.

Both pattern tree graphs / dendograms show three levels of concurrence of paraverbal communicative behaviours, Figure 1 shows an example of a T-patterns from the paraverbal communicative behaviour of a coach during alongside three competitive matches.



Figure 1. Example of coach communicative T-pattern obtained from three matches.

RESULTS

(1) Communicative situations involving regulation are those in which the coach requires an immediate response from players (for example, orders, questions, etc.). In this kind of situation, regulatory gestures (RE) are morphologically coded predominantly by means of non specific football Emblems (EMB). Deictic forms (DEI) of gestures have a special meaning since, anthropologically speaking, they are perhaps the first communicative gesture whose function was to indicate or point at something. Therefore, above and beyond the individual style of each coach in using one deictic gesture or another, the important aspect is which one is used and how. The T-patterns conducted here show that such gestures are usually associated with regulatory behaviours, although they may also appear when the coach illustrates as part of an explanation.

(2) Communicative situations involving *illustration* are those in which the coach does not require an immediate response from players. As such, most explanations made by

a coach regarding situations, or the feedback provided about a situation already performed, are examples of illustrative behaviour. In this kind of situation, illustrative gestures (IL) are coded through Beats (BEA), which are gestures without any specific iconic definition. Also Pictographs (PIC) and Kinetographs (KIN) are of interest in relation to the effectiveness and discursive clarity of coaches involving illustration. Coaches use them in a way that is more adequately tailored to their own communicative style.

(3) Many adapters were observed (for example, object adaptor, multi-adaptor, hetero-adaptor and, especially, self-adaptor), although these gestures have no communicative purpose expert coaches use such gestures in a way that avoids any interference with the quality of their communication, since the gestures are made when they are not communicating directly with players.

CONCLUSION

Observation of the different matches revealed some logical sequences in the coaches' behaviour over time, since they adopt the same behaviour when facing certain situations already encountered in a previous game. We now aim to recruit a larger sample and analyse the data using the Theme software [9], which yields behavioural patterns in a recurring (log) and sequential (lag) way. These are known as T-patterns and will provide valuable information about the communication profile of coaches.

With respect to the criteria of the observation instrument SOCOP-Coach the relevant T-patterns obtained and described in the results section invite a more detailed discussion in the communicative styles of specific and expert coaches.

- 1. Anguera, M.T. Observational Methods (General). In R. Fernández-Ballesteros (Ed.), *Encyclopedia of Psychological Assessment*, Vol. 2. London: Sage (2003), 632-637).
- Anguera, M.T., Blanco, A. and Losada, J.L. Diseños Observacionales, cuestión clave en el proceso de la metodología observacional. *Metodología de las Ciencias del Comportamiento*, 3, 2 (2001), 135-161.
- 3. Baesler, E.J. and Burgoon, J.K. Measurement and reliability of nonverbal behavior. *Journal of Nonverbal Behavior*, *11*, 4 (1987), 205-233.
- 4. Boice, R., Hanley, C.P., Gansler, D., Shaughnessy, P. and Dudek, B.C. Generality of observational skill across verbal and nonverbal modes: Literature review and experimental test. *Journal of Nonverbal Behavior*, 8, 3 (1984), 172-186.
- Castañer, M. SOCOP, sistema de observación para la optimización de la comunicación paraverbal del docente. *Temps d'Educació 36* (2009a), 231-246.

- Castañer, M., Miguel, C. and Anguera, M.T. Socop_Coach: An Instrument to Observe the Paraverbal Communication of Coaches in Competitive Match Situations. *Redaf Journal Physical Activity and Sport.* 2, 2 (2009b), 2-10.
- Castañer, Torrents, Anguera, Dinušová, and Jonsson. Identifying and analyzing motor skill responses in body movement and dance. *Behavior Research Methods* 41, 3, (2009c), 857-867.
- Jonsson, G.K., Anguera, M.T., Blanco-Villaseñor, A., Losada, J.L., Hernández-Mendo, A., Ardá, T., Camerino, O. and Castellano, J. Hidden patterns of play interaction in soccer using SOF-CODER. *Behavior Research Methods, Instruments & Computers, 38, 3* (2006), 372-381.
- 9. Magnusson, M.S. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments & Computers, 32*, 1 (2000), 93-110.
- 10. Murphy, N.A. Using thin slices for behavioural coding. *Journal of Nonverbal Behavior*, 29, 4 (2005), 235-246.
- 11. Woolfolk, A.E. The eye of the beholder: Methodological considerations when observers assess nonverbal communication. *Journal of Nonverbal Behavior*, 5, 3 (1981), 199-204.
- 12. Pattern Vision, *THEME Coder* (software), 2001. Retrieved January 15, 2002, (from http://www.patternvision.com.).

Temporal Dynamics of Multimodal Multiparty Interactions: A Microgenesis of Early Social Interaction

Kaya de Barbaro kaya@cogsci.ucsd.edu

Christine M. Johnson johnson@cogsci.ucsd.edu

Department of Cognitive Science University of California, San Diego 9500 Gilman Drive, La Jolla, CA 92093-0515, USA T. +1 630 788 4147

Deborah Forster forstermobu@gmail.com MOBU Research Inc. 3757 1/2 7th Ave San Diego CA 92103, USA Gedeon O. Deák

Department of Cognitive Science, University of California, San Diego 9500 Gilman Drive, La Jolla, CA 92093-0515, USA T. +1 (858) 822-3352

ABSTRACT

This project characterizes the development of early social interactions in a longitudinal sample of mothers and their infants. To do this we track the interaction in as a temporal configuration of multimodal, multiparty components, coding a total of 14 dimensions at each frame. Our final goal is to develop methods which we can use to systematically document regularities in the way attention gets organized longitudinally in a large sample of mother-infant dyads. Coding systematically at this scale allows us to ground a developmental account of early social interaction. For this project, we developed 1) ways to reduce the labor intensity of coding frame-by-frame multidimensional data over the entire sample and 2) ways of quantifying complex configural regularities across our sample.

Author Keywords

Infant multimodal development process, attention, triadic, dyadic social interaction, "event-based analyses".

INTRODUCTION

This project characterizes the development of early social interactions in a longitudinal sample of mothers and their infants. In particular it focuses on infants' rapidly developing abilities to coordinate attention to objects and people. The period 9-12 months is thought to bring a new ability to coordinate with and share the objects of another persons' attention—like imitating the actions of a parent or playing a ball tossing game. This has been termed "triadic" or "you-me-it" attention.

Previous research has documented the development of early mother infant interactions by longitudinally tracking the high level construct of "focus of attention". From these accounts we have a series of binary performance scores: early on, the infant can attend to mom, but not toys, then at six months onwards, to toys, but not mom, and finally, at 12 months, they can fluidly attend to both mom and toys [1]. Inherently discontinuous, these results have puzzled researchers eager for a coherent developmental account of the processes that bridge to the triadic "shift" at 12 months [2, 3, p.34]. Previous accounts attribute these novel abilities to a sudden "cognitive shift": the emergence of the ability to "read" others' intentions [4]. We argue that the apparently sudden appearance of these high level activities at 12 months may be an artifact of such discontinuous, unidimensional assessments.

Instead, in our work we track such interactions as temporal configurations of multimodal, multiparty components, and consider triadic attention a special configuration of these dimensions, wherein the infant is able to coordinate its attentional resources (e.g. gaze, hands, mouth) in conjunction with the adult's towards a set of toys. In our analysis, we focus on achievements of social interaction that successfully occur at all ages. For example, in dyads of all ages, infants can direct their gaze, hands, and mouth to nearby toys. However, the scaffolding necessary for this achievement is very different across developmental time. For example, young infants may need maternal manual elaboration of a toy to maintain gaze to it, where older

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

infants may be able to do so in a wider variety of contexts of maternal activity.

In contrast to previous methods, our methods allow for a direct comparison of the moment-to-moment processes by which younger and older dyads coordinate their attentional resources to toys and one another. Specifically, rather than focusing on changing performance, we document the changing participation dynamics: the organization and timing of the micro level components as the dyad gets organized to attend to one another and toys at each age [for examples of shifting emphasis from measuring outcome to tracking participation dynamics, see 5, 6]. This will require tracking the interaction using a much higher number of dimensions than has previously been coded.

Our final goal is to develop methods which we can use to systematically document regularities in the way attention gets organized longitudinally in a large sample of motherinfant dyads. Our video corpus is a sample of 40 motherinfant dyads each in 4 longitudinal sessions in which the infants were four, six, nine and twelve months of age. Cameras recorded three angles during a 7-min "free play" interaction with a set of toys. Coding systematically at this scale allows us to ground a developmental account of early social interaction. For this project we developed 1) ways to reduce the labor intensity of coding frame-by-frame multidimensional data over the entire sample, and 2) ways of quantifying complex configural regularities across our sample. We describe our approach below.

OVERVIEW

We used the qualitative analyses to 1) identify regularities in the way the dyad gets organized to attend to toys and one another at each age and 2) to identify a regularly occurring event that will allow us to easily and quickly identify important segments of the interaction. These segments are important in that they would be able to capture the regularities observed. Initial systematic analyses are a test of the regularities observed as well as of the validity of the events for capturing those regularities. When scaling up to coding and analysis of the 40 dyad subsample, an event based analysis allows us to selectively employ our high dimensional coding scheme: specifically, in only those segments of the interaction that have been pre-identified as most relevant for our questions.

QUALITATIVE ANALYSES

We have completed an extensive qualitative analysis of a subsample of 5 interacting dyads each in 4 longitudinal sessions in which the infants were four, six, nine and twelve months of age. The results of the qualitative analysis are a description of typical dyadic profiles at each session surrounding two key segments of interaction 1) How does the infant get organized to attend to the toys at each age? And 2) How does the infant respond to the actions of the mom? These were our successful achievements upon which we could ground our developmental account of triadic

attention: they occurred at each age, but the way in which they occurred were very different across ages. Broadly, the order of which modalities are engaged, the infant's capacity to divide its attention and recover from interruptions, whether the same interventions by the mom were supportive or disruptive, etc. appeared to vary systematically with ages. The regularities that we identified reveal an increasing autonomy and complexity of the infants' abilities leading up to the 12 month "shift" to triadic attention. Based on these observations, we identified the moment that a mother introduced a novel toy (MINT) and the following unfolding dyadic organization as an event that would capture the developmental changes we wished to capture.

Systematic Coding

We systematically coded three minutes of each session for the purposes of a quantitative testing of our predictions from the qualitative analysis and testing of the MINT event. Coding was done using human coders using commercially available software. In systematic analyses, we coded 14 dimensions, including: Maternal Gaze, Right hand, Left hand, Affect; Infant Gaze, Right hand, Left hand, Affect and Mouth, the location of the three toys, as well as the vocalizations of mom and infant. Each dimension was coded frame by frame at 10 frames per second (FPS), although not all dimensions were coded with mutually exclusive and exhaustive codes. Each dimension had a number of distinctions. For example, for hands we have differentiated between reach, grasp, contact, passive tactile contact, hover, actions on the toys (specified for type of action, e.g. bang, hit, squeeze), as well as certain types of toy motion.

During the creation of each coding scheme, we used commercially available timeline making software in order to represent multiple coders' annotations of the same video segments in a colorful and visible way. This allowed us to very easily visually assess where mismatches occurred and thus easily pinpoint where the coding schemes needed further specification or alternatively where individual coders needed to be more careful about the coding criteria.

PATTERN IDENTIFICATION AND QUANTITATIVE ANALYSES

We also used the timeline software to visualize the components of the interaction in order to further identify and specify regularities in the multimodal multiparty configurations.

Precise specification of the regularities is important for creating quantifiable features that can be used to test observational results. We have created a set of features that correspond to the various regularities observed in the qualitative analyses. These configurations capture complex multidimensional configurations that occur in real time. Based on our qualitative analyses, it is precisely these sorts of features that we believe will differentiate older and younger dyads. For example, we observed a slow, rigid sequence of modalities in typical four month olds, where gaze to toy preceded and co-occurred with manual contact, followed by mouthing of toy. Additionally, infant gaze remained locked to toy for majority of contact. We can quantify this complex configuration of multidimensional components by calculating features such as the following 1) the percent likelihood that infant gaze contact with toy occurs before infant left or right hand, and 2) the ratio of the duration of infant gaze contact with toy relative to infant manual contact with toy.

Features are based on observations from individual sessions, but all features will be calculated at each session, such that there is a single numerical value associated with each feature for each MINT event. These features can be averaged across infants of the same ages in order to then make comparisons across the ages. These features thus provide a test of our proposed trajectory of changes between the four, six, nine and twelve month dyads. Systematic verification of these qualitatively observed regularities will thus ground theoretical claims of increasing complexity.

This feature-based analysis is event-based in that the features are calculated within the boundaries given by the event. Specifically, the unfolding interaction following each MINT provides the dyad a chance to organize the infants' attention to the novel toy being introduced by the mom. It is the particulars of the interaction following the MINT that we wish to quantify in order to test our qualitative observations. However it is an empirical question how much of the interaction following the MINT will be needed to best capture the predicted regularities. This is critical for scaling our analyses up to the 40-dyad sample, as the labor intensive nature of coding will only allow us to code those segments of the interaction which are necessary for calculating features. Our systematically coded subsample will thus function as 1) a preliminary test of our qualitative observations and 2) a testing ground for precisely defining the start and end boundaries of the MINT events for high dimensional coding and feature calculation for the 40-dyad sample.

- 1. Bakeman, R. and Adamson, L. Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*, 4 (1984), 1278-1289.
- Fogel, A. and DeKoeyer-Laros, I. The Developmental Transition to Secondary Intersubjectivity in the Second Half Year: A Microgenetic Case Study. *Journal of Developmental Processes*, 2 (2007), 63-90.
- 3. Adamson, L. and Bakeman, R. The development of shared attention during infancy. Annals of child development, 8 (1991), 1–41.
- 4. Tomasello, M., et al., Understanding and sharing intentions: The origins of cultural cognition. Behavioral and Brain Sciences, 28, 5 (2005), 675-691.
- 5. Johnson, C.M. Distributed primate cognition: A review. *Animal Cognition*, *3*, 4 (2001), 167-183.
- Forster, D. and P.F. Rodriguez, Social Complexity and Distributed Cognition in Olive Baboons (Papio anubis): Adding System Dynamics to Analysis of Interaction Data. *Aquatic Mammals*, 32, 4 (2006), 528-543.

Measuring Behavioral Indices of Cognitive Processing in Children

Rosemary Meza

Linda Phan

San Diego State University, Laboratory for Language and Cognition, 6330 Alvarado Crt Suite 208, San Diego Ca, 92102 LindaP.Phan@gmail.com

San Diego State University Laboratory for Language and Cognition, 6330 Alvarado Crt Suite 208, San Diego Ca, 92102 RosemaryDMeza@gmail.com

Gwen Littlewort

University of San Diego California Machine Perception Lab 9500, Gilman Dr. La Jolla Ca, 92093 Gwen@mplab.ucsd.edu

Marni Bartlett

University of San Diego California, Machine Perception Lab, 9500 Gilman Dr. La Jolla Ca, 92093 Marni@salk.edu

Judy Reilly

San Diego State University, Laboratory for Language and Cognition, 6330 Alvarado Crt Suite 208, San Diego Ca, 92102 Reilly1@mail.sdsu.edu

ABSTRACT

The integration of verbal and nonverbal behavior is a complex interaction. Previous studies have mainly focused on understanding the development of single behaviors, and in doing so the interaction is often not apparent. By using a multi-modal coding system, the interaction among verbal and non-verbal behaviors will give us better insight into how they develop and change over time.

Author Keywords

Multi-modal, integration, verbal, non-verbal behavior, development.

INTRODUCTION

How children plan, organize their thoughts, and resolve problems has long been of interest to psychologists and educators. However, little is known about how children recruit and integrate verbal and non-verbal behavior. By using the video coding interface ELAN [1,2], we have developed a multimodal coding system that allows us to map temporal and developmental changes in patterns of verbal and non-verbal behavior, as well as how children integrate verbal and non-verbal behavior (e.g., speech and facial expression).

Our coding system will permit us to identify behavioral indices of cognitive processing in children in dyadic problem-solving contexts. Which behaviors characterize the thinking processes in children? How do such patterns of behavior change over time?

METHODS

As part of a larger study preschool (N = 12; ages 3-4yrs), school-age children (N = 12; ages 7-8yrs) and college-age adults (ages 19-35yrs, N = 12) took part in two problem-solving contexts:

1) Biographical interview—a naturalistic conversation between the participant and experimenter which provides a naturalistic sample of the participant's language abilities and use of language to address questions. From the biographical interview three question-answer sequences with varying levels of required cognitive effort were of particular interest (see Table 1).

2) Mystery box—a problem solving task in which participants are asked to identify five objects (wooden cube, triangle, cylinder, toy dinosaur, rubber duck) using only haptic input.

Both types of interactions are recorded from three angles using HD video cameras (see Figure 1).

| Question | Cognitive effort |
|---|---------------------|
| What is your name? | Low |
| How old are you? | Low |
| For children: What do you like to do at school? | High |
| For adults: What is your best subject and why? | High |

Table 1. Input requiring varying levels of cognitive effort.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Top-down view of experiment set up. The three camera angles capture multiple perspectives of the interaction; multiple camera angles also allow for more accurate analysis (e.g., detecting direction of participant's eye gaze).

CODING AND ANALYSIS

In each problem-solving context, the interaction between experimenter and participant is schematized into four phases:

Interaction Schema

- Input: participant receives question or a mystery object
- Latency: period in which the participant is thinking
- <u>Answer</u>: participant gives a response
- <u>Turn surrender</u>: participants behavior after completing a response

The duration of each phase is measured and annotated in ELAN. In each interaction phase the participant's <u>Eye gaze</u>, <u>Facial Expression</u>, and <u>Speech</u> is measured as follows:

Eye gaze

Eye contact categories (+K = makes eye contact with experimenter; -K = no eye contact with experimenter) are adapted from Reilly and colleagues [5]. Confirmation of eye contact with experimenter was determined from two camera angles. Frequency of eye contact and gaze aversion is calculated per phase.

Facial expression

Facial expressions are coded during the <u>Latency</u> and <u>Answer</u> phases using the Facial Action Coding System (FACS) [3], a taxonomy based on adult facial muscle movements. The upper and lower parts of the face were coded separately to account for muscle movements in the lower face influencing muscles in the upper face and vice versa.

Language

Linguistic data is transcribed using the CHILDES system [4]. Speech is coded for hedges (i.e., linguistic markers of



Figure 2. Example of coded verbal and non-verbal behavior in ELAN.

uncertainty). Both linguistic and non-linguistic data are annotated in the ELAN grid (see Figure 2).

RESULTS

Results suggest that there are significant developmental differences in patterns of behavior in problem-solving contexts. Such that, patterns of eye gaze in older children are similar to those of the adults; however, it seems that younger children are not using gaze patterns in the same manner (see Figure 3). Furthermore, the degree of cognitive effort that is required to solve a problem also influences patterns of eve gaze in children. During high cognitive demand, adults and older children showed similar gaze patterns by averting their gaze from the experimenter; however, younger children continued to make eye contact with the experimenter suggesting that they are using the adult in the situation as a social reference. The integration of language and speech suggests that when discussing a response that has positive affective content, both younger and older children will display a positive facial expression. However, when recounting events of negative affective content, older children are more likely to show positive facial expression suggesting that they are masking negative affect (see Figure 4 & 5). Moreover, in a problem-solving context, older children and adults are able to use language (i.e., linguistic markers that reflect varying levels of certainty) to reflect the cognitive ability of understanding dual perspectives; whereas younger children's answers were short and concise (see Table 2).

| Object | Younger child | Older child | Adult |
|----------|------------------|--|---|
| Cylinder | A circle. | This one is still a building block except it's like I forgot what the name iscylinder. Cylinder? | It's like a round stick it's shaped like a chapstick but it's a block. |

Table 2. Example of an older child and adult using linguistic hedges to indicate uncertainty.



Figure 3. Proportion of eye contact made by each age group during the mystery box task.



Figure 4. The younger group is more likely to show positive affect that is consistent with linguistic content then they are to show negative affect

DISCUSSION

By characterizing patterns of behaviors and seeing how these patterns temporally integrate in problem-solving contexts we are able to better understand the child as a problem-solving agent. The developmental picture that emerges is a complex interaction of multiple modalities. This would not have been so evident if we had only focused on a single behavior or modality. Using a multi-modal coding system allows us to see patterns of behavior and the complex interactions that arise from their integration.

Older Group



Figure 5. The older group is more likely to show positive affect that is consistent with linguistic content than negative. Also, they are more likely than the younger group to mask negative linguistic content by displaying positive affect.

- Brugman, H., & Russel, A. Annotating Multimedia/ Multi-modal resources with ELAN. In: *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*, (2004), 2065-2068.
- Crasborn, O., & Sloetjes, H. Enhanced ELAN functionality for sign language corpora. In: *Proceedings* of *LREC 2008*, Sixth International Conference on Language Resources and Evaluation, (2008), 39-43.
- 3. Ekman, P., & Friesen, W. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1* (1969), 49-98.
- McWhinney, B. & Snow, C.E. The Child Language Data Exchange System. *Journal of Child Language*, 12 (1985), 271-296.
- Reilly, J.S., M.L. Mcintire, and H. Seago. Affective prosody in American Sign Language. *Sign Language Studies*, 75 (1992), 113-128.
- Wertsch, J., McName, G., & McLane, J. The adult-child dyad as a problem-solving system. *Child Development*, 51 (1980), 125-1221.

A Method for Remote and Semi-Automatic Usability Evaluation of Web-based Applications Through Users Behavior Analysis

Ariel Vargas

Institute of Computing University of Campinas P.O. Box 6176, 13084-971, Campinas, SP, Brazil vargas@ic.unicamp.br Harold Weffers Laboratory for Ouality Software

Eindhoven University of Technology P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands h.t.g.weffers@tue.nl

Heloísa Vieira da Rocha

Institute of Computing University of Campinas P.O. Box 6176, 13084-971, Campinas, SP, Brazil heloisa@ic.unicamp.br

ABSTRACT

In this paper we describe a method for evaluating the usability of web-based applications. Our method is based on remote and automatic capture and semi-automatic analysis of users behavior, in order to find usability problems in the applications' interfaces. The goal of our method is to allow an analysis of the way users actually interact with the evaluated interface. Through the analysis of users behavior is possible to find patterns of interaction. Analyzing the patterns found and comparing it to the expected behavior for the tasks performed by users, we can detect usability problems. In this paper we also briefly describe a first experiment with our method and some initial results that point to the potential of the method in performing remote and automatic usability evaluations.

Author Keywords

Usability evaluation, remote usability evaluation, user activity tracking, log file analysis, semi-automatic analysis, web usage mining, pattern mining, sequence mining, user behavior analysis, user experience.

ACM Classification Keywords

H.5.2 User Interfaces: Evaluation/methodology

General Terms

Ergonomics, Human Factors, User Issues, Graphical User Interfaces (GUI), Web-based interaction, Pattern analysis, Experimentation.

INTRODUCTION

In the last years the web has become a common environment for computer applications. This has aroused a growing interest about the usability of these applications among researchers and developers. In order to develop a good web-based application interface it is important to consider its usability since the beginning of the development process and throughout the lifetime of the application. Regardless of the development methodology applied, it is important to evaluate the application's usability during and also after the end of the development [6]. There are many methods to evaluate graphical user interfaces of computer applications. The most popular method is the user test, in which an evaluator observes the user behavior during his interaction with the interface, in order to detect usability problems [11]. Observing users behavior is an efficient method to find usability problems in the interface, however it is an expensive method due to the costs of finding users to test, moving them to the test laboratory, preparing the infra-structure, carrying out the test, collecting and analyzing the results [10,4,9]. Due to these costs it is common just to analyze the behavior of a few users in user tests. Furthermore, analyzing the behavior of a few users hinders a quantitative analysis giving to the evaluation just a qualitative feature. Some problems could just be highlighted in a quantitative analysis and also the impact of them could just be evaluated analyzing a large number of users [9,12]. Another important factor when the usability of an interface is evaluated is its context of usage, which is difficult to simulate in a user test made in a laboratory. To deal with the difficulties mentioned above we propose a method for usability evaluation based on remote and automatic capture and semi-automatic analysis of users behavior in order to find usability problems in webbased application's interfaces.

RELATED WORK

According to [2], in a remote and automatic usability evaluation users and evaluators are separated in space and time, i.e., users perform their interaction with the application in their usual work environment without moving to a test laboratory. Evaluators do not watch users during their interaction with the application. They just analyze their behavior afterwards. To perform this kind of analysis, the interaction of all users, performed in the application's

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

interface, is captured in a log file. Software tools analyze these captured data afterwards. There are some tools already developed to perform this kind of usability evaluation as: WebVip [16], WET [8], TEA [12] and UsaProxy [3], which just perform the capture of users data but do not analyze it. WebVip and TEA logs are similar to web server logs, with no detailed information about actions performed by users such as, mouse clicks and mouse movements. WET performs more detailed capture than WebVip. However, WebVip and WET need manual insertion of code in each web page to make the capture. TEA needs to be installed in the client's machine. UsaProxy, differently from the previous approaches, makes detailed capture of users actions and has the advantage of working as a proxy between server and client. It does not need any manual insertion of code in web pages or installations in client or server. Besides the capture tools mentioned, there are tools which perform the capture and some analysis of the data captured as: WebRemUsine [13]. AWUSA [17], WAUTER [5], WELFIT [15] and WebQuilt WebRemUsine and WELFIT capture detailed [9]. information about users actions, however they need to insert some code in each web page on the server in order to make the capture. WebRemUsine analysis compares the expected sequence of actions for a task to the sequence of actions performed by a user. As a result, the tool shows the differences between both sequences. WELFIT performs an analysis of all users actions on a single page but does not care about the path followed between pages. AWUSA works with server logs and due to this it has no precision in capturing users actions. Its analysis is the same as WebRemUsine, comparing two sequences of actions and showing the differences. WAUTER captures users interaction in the client side. Its capturing is more precise, getting all actions performed by users in the web application, but it needs to install software on the client's computer. WAUTER makes its analysis just like AWUSA and WebRemUsine, comparing expected and performed sequence of actions for a task. WebQuilt works the same way as WAUTER capturing actions in the client side, its capture is however not so accurate, just getting information about pages requested to the web server. Nevertheless, its analysis is smarter than other tools referred before. It gets all actions performed by all users and groups them to find the most common navigation path for the pages of the application.

PROPOSED METHOD

Our proposed method for performing usability evaluation based on remote and automatic capture and semi-automatic analysis is called WebHint. It is composed of 3 steps as described below and shown in Figure 1.

Step 1 - Task Definition

The first step in the WebHint method is the definition of the tasks to be analyzed in the evaluation. A task is a sequence of actions performed by users in the application's interface with a specific goal. Ex: The sequence of actions performed



Figure 1. The WebHint proposed Method.

to read an e-mail message, in a webmail application, constitutes a task. In Step 1, an evaluator defines the tasks performing the expected sequence of actions for each task in the interface of the evaluated application, as was planned by the application designer. The sequence of actions performed is captured by software and saved in a log file.

Step 2 – Users Interaction Capture

In the step 2 the users interaction with the application interface is monitored. All actions performed by all users in the web application's interface are captured: mouse movements, keystrokes, links accessed, pages loaded, etc.

Step 3 – Data Analysis

This is the most important step in the evaluation. Here all users actions captured in the step 2 are analyzed, as shown in Figure 2 and described below.

The first activity in the analysis step is the extraction of the tasks performed by users from the log files. For this purpose some algorithms were implemented. They are explained below: 1) Split log - the algorithm splits the log with all interaction of all users in several log files containing the interaction of one single user per file. 2) Determine sequence intervals - the algorithm finds in each log file the intervals in which there is a sequence of actions that represents the execution of a certain task. This is made looking for representative actions in the task as the "begin" and "end" points. The algorithm also deals with intervals without a "begin" or "end", i.e., possible incomplete tasks. The intervals found are extracted from the log. 3) Extract executed tasks - For each interval found, the algorithm applies a LCS (Longest Common Subsequence) [14] function to measure how similar the extracted interval is related to the expected sequence for the referred task. If the interval has a certain similarity rate, the sequence is extracted from the log.



Figure 2. Data analysis.

The second activity in the analysis is the clustering of the extracted sequences and the searching for the most common patterns for the task, i.e. the most common way users performed the referred task.

In the third activity, the expected sequence for the referred task is compared with the most common patterns of execution for the task. Looking for the differences between the "expected" execution and the "actual" patterns is possible to identify if the task is really performed as planned by designers or not. Moreover, it is possible to find problems in the execution of the task and usability problems in the interface of the application.

In the fourth activity heuristics are used to detect problems. The heuristics are sequences of actions that represent known usability problems. If in one pattern found has a match of actions with any heuristic, a possible problem can be detected.

In the end of the analysis, the evaluator obtains the results composed by: the most common patterns of execution for each task; differences between the "expected" execution and the "actual" patterns found; and hints of usability problems detected.

Considerations About the Proposed Method

Our proposed method has some advantages over the tools presented in the related work section of this paper. In Step 1 of WebHint, we tried to simplify the task definition avoiding the use of notations. Opposite to WebRemUsine and WAUTER using notations, our method just requires a simple execution of the task in the application's interface to define it.

In Step 2, we monitor users behavior capturing all actions performed in the interface of the web application evaluated. We intend to find the most common patterns of actions executed by users performing the tasks as in WebQuilt's approach. However, our analysis is deeper than in WebQuilt and AWUSA due to our captured data being more accurate, including mouse movements, clicks and all actions performed in a webpage, not only the sequence of pages requested. Using a proxy approach as UsaProxy we do not have the workload of manually editing each webpage in order to insert code to capture users interaction as in WebVip, WET, WebRemUsine and WELFIT. It is also not necessary to install software on a client's computer as in WAUTER or WebQuilt.

In Step 3 of WebHint, we analyze all actions from all users interacting with the application. It allows us to make a quantitative analysis of the usability of the application. Differently than WebRemUsine and WAUTER, just performing the comparison between two single sequences of actions, our analysis uses the most common patterns found to compare to the expected sequence. It gives us a comprehensive analysis of the users behavior. Finally, in our method we intend to use heuristics in order to automate the usability problems detection.

METHOD APPLICATION

In order to validate our method, a first experiment was carried out. In the experiment, 52 users had their interaction monitored for a period of 2 months in TelEduc¹ (beta version 4.1.1). TelEduc is an environment for on-line courses where users have tools to interact with, as mailbox, file repository, wall, etc. A simulated course was prepared in TelEduc for the experiment and the users were invited to perform some tasks as participants of the course. The experiment had the goal to be a pilot test for the method.

In the experiment we used the UsaProxy tool for capturing the users interaction in Step 2 and for capturing the task definition in Step 1. In Step 3 we used some implemented algorithms for extracting the tasks from the log, as mentioned previously in the Proposed Method section. The process-mining tool ProM [1] [7] was also used for clustering and detecting the execution patterns of the tasks.

RESULTS

In this pilot experiment WebHint showed good potential in finding usability problems, as described in the example below. The task analyzed in the example consists in: "reply an email message received", using the tool Mail of TelEduc. One of the interaction patterns found in the users behavior for this task is illustrated in Figure 3.

¹ http://www.teleduc.org.br



Figure 3. Example of usability problem detected in the experiment.

The loop in the diagram shows that users perform the following steps: "press the button Send", "select the receiver" and finally "press the button Send" again to finish the task. This behavior differs from the expected sequence of actions for the task. The expected sequence consists of "pressing the button Send" once to finish the task. There is no selecting of the receiver. Analyzing this pattern found it was possible to figure out that the interface of the application does not set the receiver of the message as default, when the button "Reply" is pressed. So, it is always necessary to set the receiver of the message, even if the user is just performing a reply message action. This is a usability problem because the application breaks an interface standard for e-mail tools, which consists in setting the previous sender as the receiver for the reply message.

In the analysis of the data from the experiment we did not use heuristics to automatic detection of usability problems. The heuristics for problems detection are still in development. The usability problems, as the one illustrated in the example above, were detected analyzing the results obtained from the sequences clustering, the patterns of usage found, and the patterns comparison performed using the ProM tool.

The pilot experiment developed with WebHint, even executed with a small number of users, achieved its goal, being useful to validate the method. The results of the experiment point to the potential of our method in performing remote and semi-automatic usability evaluations based on users behavior analysis.

REFERENCES

- 1. Van der Aalst, W.M.P., Weijters, A.J.M.M. 2004. Process mining: a research agenda. Computers in Industry 53, 231–244.
- 2. Andreasen, M. S., Nielsen, H. V., Schrøder, S. O., and Stage, J. 2007. What happened to remote usability

testing?: an empirical study of three methods. In Proc. of the SIGCHI - CHI '07. ACM, New York, NY, 1405-1414.

- Atterer, R. 2006. Logging Usage of AJAX Applications With the "UsaProxy" HTTP Proxy. In Proc. of the WWW 2006 Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, Edinburgh, Scotland, May 2006.
- Baker, S.; Au, F.; Dobbie, G.; Warren, I. 2008. Automated Usability Testing Using HUI Analyzer. In ASWEC 2008, vol., no., pp.579-588, 26-28 March 2008
- Balbo, S, Goschnick, S, Tong, D, Paris, C. 2005. Leading Web Usability Evaluations to WAUTER. In The Eleventh Australasian World Wide Web Conference, Gold Coast, 2005.
- 6. Cato, J. User-Centred Web Design, Addison Wesley, 2001.
- van Dongen, B.F., Alves de Medeiros, B.F., Verbeek, B.F., Weijters, A.J.M.M. and van der Aalst, W.M.P. 2005. The ProM framework: A New Era in Process Mining Tool Support. In G. Ciardo and P. Darondeau, editors, Application and Theory of Petri Nets 2005, volume 3536 of Lecture Notes in Computer Science, pages 444–454. Springer-Verlag, Berlin, 2005.
- 8. Etgen, M. and Cantor, J. 1999. What does getting WET (web event-logging tool) mean for web usability. In Proc. of the Fifth Conference on Human Factors & the Web (Gaithersburg, MD, 1999).
- Hong, J, I., Heer, J., Waterson, S., Landay, J,A. 2001. WebQuilt: A proxy-based approach to remote web usability testing. ACM Trans. Inf. Syst. 19, 3 Jul. 2001.
- López J. M., Fajardo I., Abascal J.2007 Towards Remote Empirical Evaluation of Web Pages' Usability. In Jacko J. A.(Ed.): Human-Computer Interaction. Interaction Design and Usability. Part I. LNCS 4550
- 11. Nielsen, J. Usability Engineering, Academic Press, Boston, MA, 1993.
- 12. Obendorf, H., Weinreich, H., and Hass, T. 2004. Automatic support for web user studies with SCONE and TEA. In CHI '04. ACM, New York, NY, 1135-1138.
- Paganelli, L, Paternò, F. 2002. Intelligent analysis of user interactions with web applications. In Proc. of the 7th international conference on Intelligent user interfaces, 2002, San Francisco, California, USA
- Paterson, M., and Dancik, V. 1994. Longest Common Subsequences. In B. Rovan I. Privara and P. Ruzicka, editors, 19th MFCS'94, LNCS 841, pages 127-142, Kosice, Slovakia, August 1994. Springer Verlag.
- Santana, V. F., Baranauskas, M. C. C. 2008. A Prospect of Websites Evaluation Tools Based on Event Logs. In: HCIS 2008 Milan. Springer, 2008. v. 272. p. 99-104.
- Scholtz, J. AND Laskowski, S. 1998. Developing usability tools and techniques for designing and testing web sites. In Proc. of the Fourth Conference on Human Factors & the Web, Basking Ridge, NJ, Jun, 1998.
- Tiedtke, T., Märtin, C., Gerth, N. 2002. AWUSA A Tool for Automated Website Analysis. In: Proc. of the 9th Int. Workshop DSV-IS 2002, Rostock, Germany, pp. 251– 266, 2002.

Beyond Usability: A Methodology to Evaluate the Affective Experience of Interaction with E-Commerce Websites

Troy Abel Virginia Tech, SOVA

203 Draper Road, Blacksburg, VA 24061, USA

tabel@iastate.edu

ABSTRACT

Given the popularity of usability testing, why do people still feel uncomfortable interacting with websites? Could it be because usability testing does not address the user experience but rather tends to deal with efficiency and navigation but seldom with experience? The current implementation of usability research heavily relies on quantitative analysis when the nature of the issue is qualitative. Few studies have adequate scope to include both quantitative and qualitative analysis, while virtually no current Usability Evaluation Method (UEM) incorporates a qualitative component.

Activity theory describes several elements involved in human activity. By incorporating Activity Theory with quantitative and qualitative measures of user experience, the designer will be better able to assess the affective impact of a website design.

The purpose of this paper is to introduce the theory and foundational methodology used to complete the study.

Author Keywords

Activity theory, usability, usability testing methodology, usability evaluation, qualitative data, flow, rubric

INTRODUCTION

Emotions are fundamental in enriching any system interaction [1] and usability testing has become commonplace in the design of new websites to evaluate the success of failure of the system. Psychologists, designers and computer engineers all agree that a system must be usable; however, the evaluation methods currently in place neglect the measurement of the human aspect and behavior during the evaluation.

This research aims at creating a novel methodology to evaluate the human and qualitative aspects of a usability test, and fuse that with the typical quantitative aspects.

Jackob Nielsen's research uncovers several areas of web design that are easily translated into qualitative variables including screen real estate, white space, branding etc. [2] While Csikszentmihalyi's theory of 'Flow,' [3] provides a foundation which to classify each variable of interaction and determine which of three categories it falls: 1) Assisting, 2) Neutral, or 3) Challenge. [2]

Activity theory first introduced into the realm of Human Computer Interaction (HCI) by Kaptelinin & Nardi, [4] provide an interesting conceptual framework with which to work in.

ACTIVITY THEORY

Activity Theory is based in part on the work of Vygotsky, Leont'ev, and has been expanded upon by Yro Engström. In essence, activity theory states that human beings interact with their environment via situations mediated by tools. Through mediation these situations create experiences. "This notion is usually portrayed by what has come to be known as the mediation model of human interactions with the environment." (*Mwanza, 2001, p. 344*).

Leont'ev developed Vygotsky's initial theory further by providing links between social and cultural mediations, resulting in a hierarchical model of human activity.



Figure 1. Activity Triangle Model.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Engström drew upon both Vygotsky and Leont'ev's models of human interaction activity by expanding them to encompass rules, community, and the division of labor. This resulting model has come to be known as the Activity Triangle Model (Figure 1) [5]. Bonnie Nardi and Victor Kaptelinin have applied activity theory to the field of human and computer interaction.

This theory provides an excellent framework through which to explore the interaction between the user, tools, and outcomes with regard to website experience design.

Components of Activity Theory

Activity theory divides an activity of interest into seven components:

- 1. The Activity.
- 2. The Tool.
- 3. The Subject.
- 4. The Object.
- 5. The Rules.
- 6. The Community.
- 7. The Division of labor.

Subject describes the user who is enacting the activity, and the object is the motivation or intended outcome of the enactment. The tool, is situated between the subject (user) and the Object (outcome). It is the device, in this case, a ecommerce website, through which the activity is implemented.

The tool facilitates the efforts of the user to achieve the desired outcome. The rules component mediates the activity; as does any culture, pattern, gender, society, or any other factor that imposes any rules on this interaction.

Community refers to the environment in which the activity is completed, and finally, the rules of labor divide existing roles of in the activity (if necessary). It has the effect of assigning duties to those responsible for completion of each task as well as the division of a larger activity into tasks (if necessary).

THEORY OF FLOW

According to Mihaly Csikszentmihalyi, experience can be described as a journey over the course of a given amount of time [3]. Moreover, "To live means to experience– through doing, feeling, and thinking. Experience takes place in time, so time is the ultimate scarce resource we have" [3]. For this reason, time becomes an essential factor in the interactive system, as well as the concept of optimized emotions, which Csikszentmihalyi calls 'Flow.'

Flow has been described as the point when the optimal levels of challenges (obstacles) and skills (personal) are met. "Flow tends to occur when a person's skills are fully involved in overcoming a challenge that is just about

manageable. Optimal experiences usually involve a fine balance between one's ability to act, and the available opportunities to action" [3]. As skill level rises, challenges must also rise otherwise the user will become bored or apathetic. The optimum level of Flow occurs when the skills are high enough to balance the challenges presented.

When evaluating the usability of a website or system, the relationship between user ability and system challenge must be evaluated to establish the level of Flow for different user groups. One can determine quickly if the design hinders or helps the user during interaction.

METHODOLOGY: MACRO & MICRO ELEMENTS EMOTIONAL QUALITATIVE MEASURES

Activity and human behavior theory provide solid analytical frameworks through which we will explore, identify, develop, and finally deploy an instrument to assess and predict the emotional response and impact of a given webpage. The investigation of 'emotional impact,' as it relates to web page design must begin with an analysis of user experience.

This analysis will define the affective elements of the interactive experience. The investigation, therefore, begins with a content analysis of current websites, as well as the analysis of widely accepted web design guidelines. It continues with the selection of final evaluation components, moves to solidifying the target audience of interest, and finally ends with the creation of an evaluation matrix.

Content analysis allows us to define and identify both successful and unsuccessful patterns and elements in user interaction within web page design. The analysis will focus on the design and implementation of ecommerce homepages. Through this process, it will be discerned which design elements produce desirable interactions.

These design elements will then be reported in the form of an evaluation matrix. The content analysis investigation tool chosen for this purpose is based on research and case studies that set forth precedence in the areas of homepage design and usability guidelines. This precedence has set standard design patterns for various types of site designs, in particularly homepage design.

The homepage components of interest for content analysis were categorized and highlighted according to the following list adapted from Jackob Nielsen's book, *Homepage Usability: 50 Websites Deconstructed*:

- 1. Branding.
- 2. Navigation.
- 3. Advertising.
- 4. Content.
- 5. Other.
- 6. White Space.

After identification of the common content space, a selection of elements was compiled that would later be used to analyze the composition of the homepages. These elements were again chosen from a group defined by Nielson and Schneiderman based on their ability to provide an assessment of the homepage composition and are color, contrast, hierarchy, legibility, scanability, and elemental purpose [6] [7].

The resulting six elements (macro elements) were analyzed over a broad range of e-commerce sites selected at random from the top 100 e-commerce websites published by Google (see Figure 2). Quantitative statistics measuring each macro mean were generated and provided the statistic with which to apply to the case studies.

METHODOLOGY: A.E.E.R. RUBRIC

Activity Theory was used in tandem with content analysis to identify the degree to which Csikszentmihalyi's concept of 'Flow' was present within the composition of each homepage. As described earlier, Flow occurs when the participants' level of obstacles is in correct proportion with their corresponding level of skill for the particular interaction. Obstacles are defined during a website experience as objects or elements that impede the successful interaction to complete a user defined goal. We therefore utilize the theory of 'Flow' to explore the intersection of difficulty and skill level.

In order to achieve a successful rich user experience, we must maximize the motivation level (users desire to interact), while decreasing the amount of obstacles (elements which impede interaction) in hopes of maintaining a positive user experience with any system. Activity Theory provides a framework with which we can base analysis of motivation on and in turn, the necessity of motivation of the user as modulated or created by the tool



Figure 3. AEER Quadrant II for Amazon.com.

(the website), using the skill the user currently possess. The designer must be aware of design precedence whereby not contributing to obstacles during an experience.

By combining these two theories relative user satisfaction can be measured. By quantifying the motivational elements and obstacles present in the homepage a designer can iteratively calculate the relative level of affective impact of the experience. Having captured both quantitative and qualitative data, an instrument was created. This instrument is divided into four quadrants, one measures the quantitative (see Figure 3), one the qualitative data, and the remaining two quadrants contain visual interpretations of these respective values. This resulting instrument is the: Affective Experience Evaluation Rubric (A.E.E.R.) (see Figure 4).



Figure 2. AEER Quadrant II for Amazon.com



Figure 4. AEER Evaluation Matrix (Rotated 90°).

CASE STUDY

An initial case study was conducted in order to test the application of AEER to a live website. Bestbuy.com (United States electronic retailer with brick and mortar, as well as e-commerce stores), and Amazon.com were selected from a random list published by Google of the top ten e-commerce websites.

The resulting completed rubrics revealed each site posed obstacles in relationship to user skill level and required additional design attention to assist and generate a positive experience when interacting with the respective sites.

CONCLUSION

This paper assembles three key concepts and theories and combines them in a unique way to create a rubric which can be used to evaluate the affective impact of an e-commerce website. While one case study has been completed and compiled successfully, a large-scale usability test has commenced to provide internal validity to the resulting evaluation rubric.

Statistical data evaluation methodologies have been intentionally excluded from this extended abstract to allow for expanded discussion between theories and their impact on the resulting methodology.

REFERENCES

1. Brave, S. & Nass, C. (2003). Emotion in Human-Computer Interaction. In J. Julie & A. Sears (Eds.), *The*

Human-Computer Interaction Handbook. (1st. ed., pp. 81-96). Hillsdale: Lawrence Erlbaum Associates

- 2. The criteria were assembled and selected from current guidelines and recommendations provided by Jackob Nielsen in *Designing Web Usability*, Ben Schneiderman in *Research-Based Web Design & Usability Guidelines*, and finally Jakob Nielsen and Marie Tahir in *Homepage Usability*. These sources provided a screen through which criteria were chosen and included as elements in the resulting rubric.
- 3. Csikszentmihalyi, M. (1998). Finding Flow: The Psychology of Engagement with Everyday Life (Masterminds Series). Basic Books.
- 4. Kaptelinin, V. & Nardi, B. A. (2006). Acting with Technology: Activity Theory and Interaction Design (Acting with Technology). The MIT Press.
- Mwanza, D. (2001). Where Theory meets Practice: A Case for an Activity Theory based Methodology to guide Computer System Design. In M. Hirose (Ed.), *Human-Computer Interaction - Interact 2001*. Tokyo, Japan: Ios Pr Inc.
- 6. Nielsen, J. (2000). *Designing Web Usability*. Nielsen, Jakob. Designing Web Usability. Indianapolis, IN: New Riders Publishing, 2000.
- Schniderman, B. (2006). Research-Based Web Design & Usability Guidelines. U.S. Dept. of Health and Human Services. 9-1.

A Framework to Measure User Experience of Interactive Online Products

Katrin Schulze Katrin.Schulze@tu-ilmenau.de Heidi Krömker Heidi.Kroemker@tu-ilmenau.de

Center for Media Technology, TU Ilmenau PF 10 05 65, 98684 Ilmenau, Germany

ABSTRACT

Over the past years, User eXperience (UX) research in the academic community has created various approaches to UX evaluation frameworks. However, industry adopted these approaches rather vaguely into product development. Due to this existing gap between research academics and companies, this paper concerns the question of how UX evaluation can be integrated into the software engineering process of interactive online products. Therefore, influencing factors that need to be measured, including human and system aspects, emotions, a spatiotemporal dimension and motivation, were analyzed. In order to evaluate these factors transparently and to link UX factors to actual product functionalities, existing evaluation methods to collect qualitative and quantitative user data were selected and joint to an evaluation method set. First experiences of using the UX framework and this method set by means of a new web-community concept will be outlined and discussed.

Author Keywords

User experience evaluation, measurement of product qualities, user experience framework, online products.

INTRODUCTION

Nowadays, in scientific and application-oriented research user experience (UX) and the question of its evaluation is shifting to a more essential role in product development. Especially, since the use of interactive online applications has become an integral part of everyday life, users expect simple usable and tangible user interfaces. In order to develop products that meet these requirements and differ from the competitor's solution, an exclusive measurement of useful and usable products is no longer sufficient. Consequently, it becomes more important to evaluate the entire user experience of an interactive product during product development. However, there are still barriers to systematically evaluate and communicate UX that make it difficult to develop and design for a good user experience [1, 8, 16]. Main problems include the "lack of understanding of users, insufficient usage model definitions, too many constraints on the technology, and inconsistency and/or inability to integrate the technology with other parts of the ecosystem" [3].

Hence, the question of how user experience can be defined and what factors influence it needs further standardization for interactive products. Especially for companies that develop interactive products, clearly defined, approved and consistent UX evaluation methods will get more and more indispensable in order to obtain comparable results and to easily include UX evaluation into product development [17]. Though, according to Mattila et al., practical UX evaluation methods require to be valid and reliable, fast, light-weight and cost efficient, applicable for concept ideas, prototypes, and products (more requirements in [17, p.21]).

Therefore, the following paper discusses (1) a framework of UX with its influencing factors and (2) a scalable combined evaluation method set in order to measure these influencing factors by collecting qualitative and quantitative user data and to draw conclusions on approaches to optimize UX. By means of a new webcommunity concept, theoretical elaborations were approved in a laboratory study.

This study attempts to create an UX evaluation framework with a holistic view on UX that can easily be integrated into the software engineering process in order to close a little bit more the gap between research academics and companies.

THE UX FRAMEWORK

The following framework includes a definition of UX and influencing factors that aim at creating a uniform understanding and a basis for UX measurement in a company. Considering that impacts of influencing factors differ from product to product, the framework describes factors that need to be operationalized according to the product. The operationalization based on an example of a web-community will be explained consecutively.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Definition of UX

There are many definitions of UX, but not an agreed one [17]. Existing definitions within HCI name three main components of user experience [1, 8, 16, 23]: (1) *emotion* (consequence of a user's internal state [3, referencing 13, 19 and 20]), (2) *motivation* (causal for activated product experience [5], [21]) and (3) *reflection* (spatiotemporal dimension [1]). By interacting with one another these components make up the user experience as a whole.

Consequently, in this context, user experience is defined as: "the degree of positive or negative emotions that can be experienced by a specific user in a specific context during and after product use and that motivates for further usage".

Influencing Factors of User Experience

To make user experience measurable, its direct and indirect influencing factors need to be known. In Human Computer Interaction human aspects as well as systems aspects seek consideration. For that reason influencing factors are divided into these two main aspects which together influence UX before, during and after product use. UX evaluation of these three moments describes the change of UX from expectations through momentary experiences until the reflective experience. An overview of UX components and including factors is shown in Figure 1. It is assumed that basic human needs are key drivers of product use and quality perception [4, 5, 6]. Fulfillment and frustration of such needs cause user's experience with an interactive product and are mainly influenced by product qualities. According to Mahlke these qualities can be classified in utility, usability, visual attractiveness and hedonic quality [16]. Though, there is no direct correlation of one need to one specific product quality in general which is why direct links between human needs and product qualities need to be identified by analytical and empirical studies. Measurement of UX can be explicitly targeted to evaluate certain aspects of these influencing factors depending on the product goals.





| Product qualities | Operationalization |
|---|---|
| Utility : "is the question of whether the functionality of the system in principle can do what is needed" [18, p.25] | Identity management, relationship manage- ment and information management [24] |
| Usability : describes "the extent to which a product (e.g. tool) can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [9]. | Suitability for the task, self-descriptiveness, conformity with user expectations, suitability for learning, controllability, error tolerance, suitability for individualization [18, 10] |
| Visual Attractiveness: can fulfill human needs and motivate human beings since aesthetic objects contain <i>intrinsic value</i> that is higher than functional value [22]. | Vanguard design, form principles, color principles [14, p.328 cont.] |
| Hedonic Quality : can directly fulfill or awake human needs and is also seen as " <i>be-goals</i> " (e.g. being competent) [6], [7] | Challenge, identity, stimulation, empathy, trust and credibility [2], [6] |

Table 1. Operationalization of product qualities.

OPERATIONALIZATION OF UX FACTORS

In order to have a profound understanding of these influencing factors and to design for user experience during product development, these factors need to be operationalized according to the product type.

For that reason the named product qualities and relevant basic human needs were operationalized according to webcommunities. Operationalizations of product qualities are displayed in table 1 and base on [2], [6], [9], [10], [14] and [24]. Operationalizations of basic human needs were discussed and compiled in an expert group and are shown in table 2.

METHOD

The UX test is subdivided into 14 guideline-supported individual laboratory experiments of 110 minutes duration. The test aimed at collecting **qualitative and quantitative user data**. Test object was a new web-community concept of Deutsche Telekom AG available as an online mock-up.

Participants

Target group of the product were users that use the Internet daily and at least one social network with a personal user profile regularly. All of them know at least one of the portals on which the product aims to be implemented.

| Human needs | Operationalization |
|--------------------------|--|
| Relatedness | communicate fast and easy, find community of interests, integrate myself |
| Influence/ popularity | raise popularity, present myself, adopt different roles, gain influence |
| Stimulation | have fun, be creative, find balance, have a feeling of success |
| Competencies | reduce complexity, avoid stress, gain competencies, optimize time management, reduce work, help myself, orient myself |
| Security | feel secure, feel not being watched, build trust, hand over responsibility, have data control |
| Autonomy | manage information on my own, take over responsibility, help myself, have data control, preserve identity, present myself |

Table 2. Operationalization of basic human needs for a webcommunity.

Fourteen individuals (eight female and six male) from Ilmenau University of Technology participated in the test. All of them matched the target group requirements. They had a mean age of 21 years (Min = 19, Max = 32). Six participants were medium internet users (5-15 hours of internet usage per week) and eight were heavy internet users (more than 15 hours of internet usage per week).

Procedure

Participants were invited via e-mail and each of them got an individual appointment. The test contained (1) a prequestionnaire including demographic data, user behaviour, expectations and the importance of operationalized basic human needs regarding web-communities, (2) typical usability scenarios describing critical incidents and a short free exploration as well as (3) a retrospective questionnaire containing questions regarding need fulfillment. While interacting with the product, participants were asked to "think aloud" in order to gain a better understanding of user's **mental processes** (see [12]) and **emotions**.

Measurement of UX Factors

During the test qualitative data as well as quantitative data was collected. **Qualitative data** regarding need fulfillment, momentary experiences and emotions was collected via observation and open questions in interviews. **Quantitative data** regarding need fulfillment and product qualities were collected via Likert scales and Semantic Differentials. **Product qualities** were measured by seven-stage ranking semantic differentials and observed during task completion. Items to capture product quality were taken from validated item catalogues in [7], [14], [15] and [16].

Retrospective questionnaires aimed at evaluating the **reflective user experience**. Momentary experiences, **emotions** and fulfillment of **basic human needs** and were observed during task completion and during free exploration.

Basic human needs were additionally measured using fivestaged Likert scales with questions resulting from the operationalization. Questions in the pre-questionnaire aimed at evaluating expectations; questions in the follow-up questionnaire aimed at evaluating the actual need fulfillment. This way, a *required profile* and an *actual profile* were determined in order to identify strength and weakness of a product intuitively. Open questions at the end of the test aimed at evaluating the main **emotions** after interacting with the product and **motivation** to use this product in the future. Data from this test can be analysed qualitatively.

CONCLUSION AND FUTURE WORK

Goal of the study was to identify and measure influencing factors of user experience of interactive online products in order to obtain indicators for product optimization and therefore, to include practical UX evaluation methods into the software engineering process.

First experiences have shown that the presented UX framework is applicable during the software engineering processes and that important indications for product optimization can be won. Influencing factors describe a holistic view with human and system aspects on UX in order to obtain important results for product managers, designers and developers. Even though a broad sample size is important for a more profound analysis of quantitative data and only a small sample size (n=14) was used to collect user data in this study, the presented framework was not only more easily practically applicable than with a broader sample size, but **similarity in results** of both qualitative and quantitative data sets became apparent. This finding clarifies, that a small sample size can already be useful to obtain important indicators for product optimization while using the presented UX framework and method set. For instance, the finding that the need for "Stimulation" was only partly fulfilled was discovered during observation, in open questions as well as based on the results of the used Likert Scales. Based on both qualitative and quantitative data clear recommendations to optimize product features and the product on the whole were extracted. An example of a recommendation to fulfill needs for relatedness, stimulation and autonomy is shown in table 3. The example in table 3 describes a weakness in the product quality Hedonic Quality (HQ) and recommends changes regarding the specific product feature "forums". Limitations of the applied method set address some of the requirements for practical UX evaluation methods, which were not completely met yet (see more in [17]).

| Prod. Qual. | Category | Weakness | Recommendation | Priority |
|----------------|----------|--|---|----------|
| HQ | Forums | Not enough topics to allow for extensive communic ation in forums | Offer and maintain much more and various forum topics and add descriptions | !! |

Table 3. Example of a recommendation to improve UX.

The presented UX evaluation method set is **applicable on concept ideas, prototypes and products**. It can therewith be integrated into different stages of product development and only needs to be adapted to the specific stage. Though, requirements regarding validity and reliability need to be proven in further tests. Because of applying the maximum of this scalable combined evaluation method **the test is only partly** *fast, lightweight and cost-efficient and a high level of expertise is required.*

In conclusion, the presented framework is an applicable approach for measuring UX of interactive online products, but further optimization in the scalability of the method set is needed in order to match the requirement of fast, lightweight and cost-efficient practical UX evaluation methods. Matching these requirements makes UX evaluation manageable in a big company. Further work includes the optimization of the presented method set in collaboration with a company and other scientists as well as proving validity of quantitative data with a broader sample size or the exclusive collection of qualitative data. Especially the impact of UX evaluation on the product lifecycle and the customer relationship lifecycle needs to be researched in further studies.

- 1. Battarbee, K. and Koskinen, I. (2005). Co-experience: user experience as interaction. CoDesign. Taylor & Francis, pages 5–18.
- Bauer, H. und Hammerschmidt, M. (2004). Kundenzufriedenheitund Kundenbindung bei Internet-Portalen – Eine kausalanalytische Studie. Konsumentenverhalten im Internet, Seiten 189–214.
- Beauregard, R., Younkin, A., Corriveau, P., Doherty, R., and Salskov, E. (2007). Assessing the Quality of User Experience. Intel Technology Journal, page 78.
- Hassenzahl, M. Eckoldt, K. and Thielsch, M. (2009). User Experience und Experience Design –Konzepte und Herausforderungen. In: Usability Professionals. Fraunhofer Verlag, pages 233-237.
- Hassenzahl, M. (2008): User Experience (UX): Towards an experiential perspective on product quality. In IHM '08: Proceedings of the 20th French-speaking conference on Human-computer interaction (Conférence Francophone sur l'Interaction Homme-Machine), 11-15.
- 6. Hassenzahl, M. (2007). The hedonic/pragmatic model of user experience. Towards a UX Manifesto, pages 10-13.

- Hassenzahl, M., Burmester, M., and Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Mensch & Computer, Pages 187–196.
- 8. Hassenzahl, M. and Tractinsky, N. (2006). User experience – a research agenda. Behaviour & Information Technology.
- ISO 9241-11, E. (1998). Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 11: Anforderungen an die Gebrauchstauglichkeit. Leitsätze. Berlin: Beuth Verlag.
- ISO 9241-110, D. E. (2006). Ergonomie der Mensch-System- Interaktion-Teil 110: Grundsatze der Dialoggestaltung. Berlin: Beuth Verlag.
- ISO/FDIS 9241-210, D. E. (2008). Ergonomics of humansystem interaction–Part 210: Human-centred design process for interactive systems (formerly known as 13407). International Organization for Standardization (ISO). Switzerland.
- 12. Jordan, P. (1998). An introduction to usability. London: Taylor & Francis.
- Khalid, H.M. and Helander, M.G. (2006) Customer Emotional Needs in Product Design. Concurrent Engineering: Research & Applications, pages 197-206.
- Koppelmann, U. (1996). Produktmarketing: Entscheidungsgrundlage für Produktmanager, 5. vollständig überarb. Auflage. Springer.
- 15. Laugwitz, B., Schrepp, M., and Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In Mensch & Computer, 125–134.
- 16. Mahlke, S. (2008). User experience of interaction with technical systems. Diss., Technische Universität Berlin.
- 17. Mattila, V.-V. K., Roto, V., und Hassenzahl, M. (2008). Towards Practical User Experience Evaluation Methods. In International Workshop, page 19.
- 18. Nielsen, J. (1993). Usability engineering. Morgan Kaufmann.
- 19. Norman, D.A. (2004). Emotional Design: Why We Love (or Hate) Everyday Things. New York: Basic Books.
- 20. Ortony, A., Clore, G.L. and Collins, A. (2004). The Cognitive Strucutre of Emotions, Cambridge: Cambridge University Press.
- 21. Rheinberg, F. (2006). Intrinsische Motivation und Flow-Erleben. Motivation und Handeln, 331–354.
- 22. Tractinsky, N. (2005). Does Aesthetics Matter in Human-Computer Interaction? Mensch and Computer 2005: Kunst und Wissenschaft – Grenzüberschreitung der interaktiven Art. München: Oldenbourg Verlag, Seiten 29–42.
- 23. Wright, P. and McCarthy, J. (2004). Technology as experience. The MIT Press.
- 24. Zerfaß, A., Welker, M., und Schmidt, J. (2008). Kommunikation, Partizipation und Wirkungen im Social Web. Halem Verlag.

iHeartrate: A Heart Rate Controlled In-Flight Music Recommendation System

Hao Liu

hao.liu@tue.nl

Jun Hu j.hu@tue.nl Matthias Rauterberg g.w.m.rauterberg@tue.nl

Technical University of Eindhoven, HG 2.44, Den Dolech 2, 5612AZ, Eindhoven, The Netherlands

ABSTRACT

Travel by air, especially long distance, the enclosed environment of the aircraft cabin causes discomfort and even stress to flight passengers. In this paper, we present a new heart rate controlled music recommendation system. Heart rate is used as a stress indicator. If the user is stressed and his/her heart rate is higher/lower than normal, the system recommends a user preferred calm/uplift music playlist to transfer the heart rate back to normal; An user experiment is done to validate our system. The experiment results not only show that our system keeps the user's heart rate at normal but also show that the passenger's stress level can be reduced by our system.

Author Keywords

Recommendation system, heart rate control, music preference, music recommendation.

ACM Classification Keywords

H.5.5 Sound and music computing

INTRODUCTION

Travel by air, especially long distance, is not a natural activity for humans. Many people experience some degree of discomfort and even stress when flying. Excessive stresses may cause the passenger to become aggressive, over-reactive and even endanger the passenger's health [1].

In this paper, we present a heart rate controlled in-flight music recommendation system. Except based on the user's music preference, music recommendation is also controlled by the user's heart rate to keep his/her heart rate within the normal range. If the user does not like the recommendation, he/she can decline the recommendation and re select the music himself/herself, based on interactions between the user and the system, the system can learn and adapt to the user's latest music preference. The rest of this paper is organized as follows. Firstly, the heart rate controlled music recommendation system is introduced. Then, the user experiment to validate the system design is given. Finally, Conclusions are drawn.

SYSTEM DESIGN

A. Heart Rate

Heart rate is a measure of the number of heart beats per minute. For an adult (age 18 and over), his/her normal heart rate at rest is 60-100 beats per minute. If the heartbeat rhythm is disrupted and is beating lower than 60, it is called a bradycardia. A slow heart beat may cause dizzy spells, blackouts, breathlessness or tiredness. If the heart beats higher than 100 beats per minutes, it is called a tachycardia. When the heart beats rapidly, the heart pumps less efficiently and provides less blood flow to the rest of the body, including the heart itself.

B. Music and Heart Rate

Peter Sleight found that listening to music with a slow or meditative tempo has a relaxing effect on people, slowing breathing and heart rate. Listening to faster music with a more upbeat tempo has the opposite effect - speeding up respiration and heart rate [2]. Steelman looked at a number of studies of music's effect on relaxation where tempo was varied and concluded that tempos of 60 to 80 beats per minute reduce the stress response and induce relaxation, while tempos between 100 and 120 beats per minute stimulate the sympathetic nervous system [3].

C. System Framework

Figure 1 presents system framework design. The system starts by observing the passenger's current heart rate that it wishes to control. This step of perception creates an internal representation of the passenger's heart rate situation. The information in this representation then must be processed in order to determine whether the passenger's heart rate is normal or not; based on this interpretation, refer to the user profile, the system then recommends on a personalized music playlist. The passenger himself/herself is an adaptive system. His perception creates an internal representation of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

the music service. This perception affects the passenger's heart rate. During this process, the passenger's heart rate may also influenced by a set of variables which in the control system called disturbances. The change in the passenger's heart rate is again perceived by the system, and this again triggers the adaptation process we have described, thus closing the control loop. In figure 1, if the system recommends music that the passenger does not like, he can reject the recommendation and re select the music. During this process, the system logs the interactions between the passenger and the system. By mining on the log information, the system can learn and adapt to the passenger's latest music preferences.



Figure 1. System framework.

D. System Architecture

Figure 2 shows the main components that make up the new heart rate controlled music recommendation system. The whole architecture is divided into five abstraction levels from functionality point of view. The lowest level is the resource level. The second layer is the resource manager layer which includes music service manager, heart rate manager and user profile manager. The third layer is the database layer which constitutes by a database. It acts not only as a data repository, but also enables the layers and the components in the layers loosely coupled. The fourth layer is the adaptive control unit layer which includes user feedback log, adaptive inference and user preference learning components. It is used to mediate between the user



Figure 2. System architecture.

profile, heart rate and available music to provide personalized music playlist keep the user's heart rate within normal range. The fifth layer is the interface layer. The passenger interacts with the system interface to get personalized music services.

USER EXPERIMENT

A. Setup

The user experiment was conducted in the SEAT flight simulator. The SEAT simulator consists of a small scale aircraft cabin residing on a moving platform, a project section to simulate the aircraft cabin outside sky, and a control room section. The interior of the aircraft cabin is divided into an economy class section, a business class section, a galley and a lavatory. The projection section include a beamer hangs above the aircraft cabin and a projection wall next to the aircraft cabin. The control room is equipped with the state of art of computers to support inseat computers in the aircraft cabin and long haul flight simulation.

B. Test Subjects

Twelve subjects were invited to participate in user experiments. Six were allocated to the controlled group and others were allocated to the treatment group. The age of the control group ranges from 21 to 33. The age of the treatment group ranges from 23 to 32. Each group is composed by three females and three males.

C. Procedure

A real world KLM KL0895 flight from Amsterdam Schipol international airport (6:20PM) to Shanghai Pudong international airport (10:45 AM <Shanghai time>, 4:55AM <Amsterdam time>) is simulated in our flight simulator on 31st (Friday), July 2009 (the controlled group user experiment) and 7th (Friday), August 2009 (the treatment group user experiment).

A professional flight attendant is invited to serve the flight cabin. Figure 3 is the procedure of the flight simulation.

Figure 4 is a snapshot of our user experiment. Figure 5 shows the Emfit heart rate sensor we have used [4]. It can be embedded in the seat to measure the user's heart rate non intrusively.



Figure 3. Procedure.



Figure 4. User experiment.



Figure 5. Heart rate sensor.

D. Results

Five patterns have been summarized. These include two uplift patterns, two calm patterns and one keep pattern. Figure 6 is one of the uplift patterns. In Figure 6, after the user listens to the recommended music playlist for four minutes, his/her heart rate is uplifted back to the normal range. Figure 7 is a calm pattern. For the details, see [5].



Figure 6. Uplift pattern.



Figure 7. Calm pattern.

Figure 8 is the stress level represented by heart rate variability. The computing is based on five minutes of heart rate. The method used is welch's periodogram where the window is 62s with 50% overlap. The difference between the control group and treatment group is significant (ANOVA, P = 0.031).

Figure 9 is the self report stress scales [6] by the passenger each hour. The difference between the control group and treatment group is significant (ANOVA, Sig. =0.001).

Giving the test subjects a high level of long haul flight presence is the precondition of the trustable of the user experiment data. Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another [7]. In this



Figure 8. Heart rate variability: LF/HF.



Figure 9. Stress scale.

paper's context, presence means the "passenger's" subjective experience of being in the long haul flight; even when the "passenger's" is physically sitting in set up. The presence questionnaire in [7] is customized to measure the user presence. It is filled in by the test subjects right after the user experiment.

Table 1 presents the result based on the twelve test subjects' answers. All the five questions scored above 3.5.

| | N | Mean | Std. deviation |
|-----------------------|----|--------|----------------|
| BeingThere | 12 | 4.0000 | 0.73855 |
| RealFlight | 12 | 3.7500 | 1.21543 |
| LaboratoryOrSomewhere | 12 | 3.5833 | 0.79296 |
| LaboratoryFlight | 12 | 3.9167 | 0.79296 |
| SitLabOrFlight | 12 | 3.5000 | 1.08711 |
| Valid N (listwise) | 12 | | |

| | Table | 1. | Subjects' | presence | results. |
|--|-------|----|-----------|----------|----------|
|--|-------|----|-----------|----------|----------|

CONCLUSION

In this paper, we present a new heart rate controlled music recommendation system. It can regulate the user's heart rate within the normal range with user preferred music playlist recommendation. If the user's heart rate is higher than the normal heart rate, the system recommends his/her preferred calm music to the user to transfer his/her heart rate back to normal; If the user's heart rate is lower than the normal heart rate, the system recommends his/her preferred upbeat music to uplift his/her heart rate back to normal. We implemented the system. The user experiment has validated our design concept.

ACKNOWLEDGMENTS

This project is sponsored by the European Commission DG H.3 Research, Aeronautics Unit under the 6th Framework Programme, under contract Number: AST5-CT-2006-030958.

- 1. World Health Organization (2005). Travel by air: health considerations. Retrieved March 1, 2008 from World Health Organization's Web site: http://whqlibdoc.who.int/publications/2005/9241580364 _chap2.pdf
- 2. Bernardi, L., Porta, C., Sleight, P. Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart*, 92, 4 (2006), 445-452.
- 3. Steelman, M. Relaxing to the beat: music therapy in perioperative nursing. *Today's OR Nurse*, 13 (1991), 18-22.
- 4. EMFIT company. www.emfit.com.
- 5. SEAT project. <u>www.seat.id.tue.nl</u>.
- 6. Bartenwerfer, H. Einige praktische Konsequenzen der Aktivierungstheorie. Zeitschrift für experimentelle und angewandte Psychologie, **16**, (1969), 195-222.
- 7. Usoh, M., Catena, E., Arman, S., Slater, M. Using Presence Questionnaires in Reality, *Presence: Teleoperators and Virtual Environments*, 9, 5 (2000), 497-503.

Measuring Motor Actions and Psychophysiology for Task Difficulty Estimation in Human-Robot Interaction

Domen Novak

domen.novak@robo.fe.uni-lj.si

Matjaz Mihelj matjaz.mihelj@robo.fe.uni-lj.si Jaka Ziherl jaka.ziherl@robo.fe.uni-lj.si

Andrej Olensek andrej.olensek@robo.fe.uni-lj.si Marko Munih marko@robo.fe.uni-lj.si

University of Ljubljana, Laboratory of Robotics and Biomedical Engineering Faculty of Electrical Engineering, Trzaska cesta 25, SI-1000 Ljubljana, Slovenia

ABSTRACT

In this paper, we describe a method for estimating task difficulty in human-robot interaction using a combination of motor actions and psychophysiology. A number of variables are calculated from kinematics, dynamics, heart rate, skin conductance, respiration and skin temperature. Discriminant analysis of the variables is used to determine whether the user finds the task too easy or too hard. The discriminant function is recursively updated with Kalman filtering in order to better adapt to the current user. The method was tested offline in a task with 20 subjects. In cross-validation, nonadaptive discriminant analysis yielded a classification accuracy of 80.2 % while adaptive discriminant analysis yielded a classification accuracy of 84.3 %.

Author Keywords

Human-robot interaction, virtual reality, sensor fusion, psychophysiology, biocooperative robotics.

ACM Classification Keywords

H5.2. User interfaces, I2.9. Robotics

INTRODUCTION

In the last few decades, robots have been developed that interact with humans in many different environments. For instance, humanoid robots provide entertainment and companionship while haptic robots assist patients in motor rehabilitation. However, studies have shown that it is difficult to fully evaluate and interpret the interaction between a robot and a human [1]. While most robots are equipped with sensitive force and position sensors that can measure the user's motor actions, such sensors cannot reveal information about the user's subjective feelings: stress, engagement etc. Such information could be estimated through the use of so-called psychophysiological measurements. Defined as measurements of physiological responses to psychological stimuli, these have been extensively used for user state estimation in various situations. For instance, users' emotional responses to computer games are reflected in their heart rate, skin conductance and skin temperature [5].

Once motor actions and psychophysiology are measured, they can be used in a biocooperative feedback loop: while the user interacts with the robot in order to perform a task, the task difficulty is adjusted in order to avoid frustration (task is too hard) or boredom (task is too easy). The principle of such a feedback loop is illustrated in Figure 1.



Figure 1. Basic principle of task difficulty adaptation.

Such a feedback loop was successfully created using only psychophysiological measurements in a flight simulator [2]. However, in that system, task difficulty was estimated somewhat arbitrarily: if physiological signals exceeded a manually set threshold, the task was considered too hard. For practical use, a more advanced method of task difficulty estimation is required. One possible method is discriminant analysis, a statistical method for classification of multidimensional data into two or more classes. Two classes ('too easy' and 'too hard') should be sufficient for basic task difficulty estimation.

Since psychophysiological responses exhibit large interindividual differences, an optimal task difficulty estimation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

method would be able to adapt to a user as the task progresses and the system obtains data about that particular user. A variant of discriminant analysis, Kalman adaptive discriminant analysis [6], can be used here. With this method, the classifier is created offline from training data and recursively updated online after every new data point. The system can thus adapt to a particular user. The principle of such an approach is shown in Figure 2.



Figure 2. Task difficulty adaptation with task difficulty estimator updating.

We examined the suitability of using adaptive discriminant analysis of motor actions and psychophysiology for task difficulty estimation. Our first hypothesis was that adaptive methods would be more accurate than nonadaptive ones. Our second hypothesis was that classification based on both motor actions and psychophysiology would be more accurate than classification based on a single type of data.

MATERIALS AND METHODS

Task

For our task, we used a scenario previously used for reaching and grasping exercise in rehabilitation robotics [3]. A photo of a subject performing the task is shown in Figure 3. In the centre of the screen, there is a table sloped toward the subject. At the beginning of the task, a ball appears at the top of the slope and starts rolling downward. The subject's goal is to catch the ball using a haptic robot (the HapticMaster, manufactured by Moog FCS) before it reaches the lower end of the table. Once the ball is grasped, the subject must then hold the ball and place it in a basket above the table. Once the ball is dropped into the basket or falls off the table, another ball appears at the top of the table and the task continues. The haptic robot's grasping device allows the subject to feel each virtual item. Seven different difficulty levels were implemented, with each higher difficulty level featuring smaller and faster balls so that they were harder to catch.

Experiment Procedure

Twenty students and staff members (16 males, 4 females) from the [removed for review] participated in the study. Mean age was 27.3 years, standard deviation 4.6 years.

Upon arrival, the task was explained and demonstrated to the subject. The physiological measurement equipment was attached. The subject rested for two minutes, then performed the task for six two-minute periods (12 minutes total). Within each period, the task difficulty was constant.



Figure 3. A subject performing the task using a haptic robot (1) and grasping device (2) while his/her arm is supported by cuffs (3). The screen (4) shows a sloped table, a ball (5) and a basket (6).

At the end of a period, the subject was asked whether he or she would prefer the difficulty of the task to increase or decrease. The difficulty of the task then changed randomly by one or two levels in the selected direction. This randomness was introduced in order to expose subjects to a wider range of difficulty levels.

Input Variables for Classification

The variables used for classification were divided into two groups: motor variables and psychophysiological variables. A feature vector was defined as the vector of all variables from a single time period from a single subject.

Motor variables describe how well subjects did and how they moved during a particular time period. Measured using the robot's force and position sensors, they include variables such as the percentage of balls caught by the subject, the subject's mean velocity in different directions, and the mean forces exerted by the subject.

Physiological signals were sampled at 600 Hz using a g.USBamp amplifier (g.tec Medical Engineering GmbH). The electrocardiogram was recorded using disposable surface electrodes placed on the torso. Skin conductance was measured using a g.GSR sensor (g.tec). The electrodes were placed on the second and third fingers of the nondominant hand. Respiratory rate was obtained using a thermistor-based SleepSense Flow sensor placed beneath the nose. Peripheral skin temperature was measured using a g.TEMP sensor (g.tec) attached to the fifth finger of the nondominant hand.

Thirteen variables were extracted from the physiological signals. From the electrocardiogram, mean heart rate as well as seven time- and frequency-domain measures of heart rate variability were extracted. Detailed information about these variables is available in an extensive paper [4]. From the skin conductance signal, three variables were extracted: mean skin conductance, skin conductance response frequency and mean skin conductance response amplitude. Skin conductance responses are defined as transient increases in skin conductance whose amplitude exceeds 0.05 μ S. From the respiration signal, two variables were extracted: mean respiratory rate and standard deviation of respiratory rate. From the skin temperature signal, mean temperature was extracted.

Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a well-known method for feature extraction and classification, used to find a linear discriminant function that best separates data points into two or more classes. The discriminant function is built using a set of feature vectors (training data), each of which has a known class label assigned to it. This discriminant function is then used to determine the best class label for new feature vectors.

As previously noted, the goal of the discriminant analysis was to classify a measurement as either too easy or too hard. Thus, there were two possible classes. Two methods of discriminant analysis were used. The first was the classic, nonadaptive LDA, which is well-described in statistical literature. The second method was Kalman adaptive linear discriminant analysis (KALDA), an adaptive version of the LDA in which the weights of the discriminant function are recursively estimated online using a Kalman filter as new data becomes available. The Kalman gain varies the update coefficient and changes the adaptation speed depending on the properties of the data. Detailed equations can be found in [6].

Classifier Fusion

While it is possible to use discriminant analysis to build a single, multivariate classifier using all the input variables at once, another option is to build a separate univariate classifier for each input variable. While the accuracy of any individual univariate classifier would be low, fusing the large number of classifiers may result in high accuracy. For classifier fusion, the result of each classifier was weighed according to its estimated accuracy, which was estimated from previously obtained training data. For instance, a classifier that was able to correctly classify 100% of the training data would be weighted with a factor of 1, a classifier that was able to correctly classify 75% of the training data would be weighted with a factor of 0.5, and a classifier that was able to correctly classify 50% or less of the training data would be weighted with a factor of 0. Classifier fusion is then done using the following formula:

$$C = \sum k_i \cdot c_i$$

where $\mathbf{c}_{\mathbb{I}}$ is the class assigned to the feature vector by classifier *i*, $\mathbf{k}_{\mathbb{I}}$ is the weighting factor of classifier *i*, and \mathbf{C} is the final assigned class. If $\mathbf{c}_{\mathbb{I}}$ can either be -1 (class 1) or +1 (class 2), the feature vector is assigned to class 1 if \mathbf{C} is equal to or less than zero and assigned to class 2 if \mathbf{C} is greater than zero.

Both multivariate classifiers with no classifier fusion and weighted vote fusion of univariate classifiers were tested.

Cross-Validation

To test the accuracy of our classifiers, we used leave-oneout cross-validation. For a classifier, the entire data set was split into the test data (all data from one subject) and the training data (all data from all other subjects). The classifier was built using the training data, then validated using the test data. This procedure was repeated as many times as there were subjects, with each subject's data used as the test data exactly once. The accuracy rate of a classifier was calculated as the number of correctly classified feature vectors divided by the number of all feature vectors. A feature vector was considered to be correctly classified if the classifier's estimate (too easy or too hard) was the same as the choice that the subject had made.

Since the data was available offline, adaptive classifiers were tested as follows. The first feature vector from each subject (i.e. from the first time period of a session) was classified using the initial classifier obtained from the training data. Then, the classifier was recursively updated using this feature vector and the choice that the subject had actually made. The updated classifier was tested on the second feature vector from each subject, once again updated and so on. The weakness of such an approach is that, in online task difficulty estimation, the choice that the subject had made would not be available. This problem is further explored in the Discussion section.

| | motor. | psychophys. | both |
|------------------------|--------|-------------|--------|
| multivar. LDA | 79.3 % | 60.3 % | 71.1 % |
| weighted univar. LDA | 79.3 % | 61.6 % | 80.2 % |
| multivar. KALDA | 79.3 % | 66.9 % | 72.7 % |
| weighted univar. KALDA | 81.0 % | 76.0 % | 84.3 % |

 Table 1. Classification accuracy rates for different classification methods.

RESULTS

Table 1 shows the accuracy rates for different classifier types and different input data (motor actions, psychophysiology or both). In the multivariate classifiers, a single classifier is made using all available inputs. In the weighted univariate classifiers, one classifier is made for each input variable and the classifiers are then fused as described in the 'Classifier fusion' section.

DISCUSSION

Comparison of Different Classifiers

In all cases, KALDA yields higher classification accuracy than LDA. This confirms the hypothesis that adaptive methods improve classification accuracy. In weighted fusion of univariate classifiers, using both motor actions and psychophysiology yields higher classification accuracy than using only one data source. This is in agreement with our second hypothesis that multiple data sources improve classification accuracy. Interestingly, in the case of multivariate classifiers, nonadaptive classification using both data sources produces a worse result than using only motor actions. This may be because the data dimensionality is large, so it is difficult to find a single robust discriminant function. Weighted vote fusion of univariate classifiers should be more robust since each individual discriminant function covers only one dimension.

While nonadaptive task difficulty estimation using motor actions already gives a classification accuracy of almost 80 %, nonadaptive classification using psychophysiology yields an accuracy of less than 62 %. However, adaptive methods increased the accuracy of psychophysiology-based classification to 76 %. In our case, the task was only performed for a total of 12 minutes, and increasing the length of the task may allow even greater improvement in classification accuracy when using psychophysiological measurements. Nonetheless, results suggest that, at least in haptic interaction, motor actions should be used as a primary data source with psychophysiology providing supplementary information.

Naturally, perfect classification accuracy should not be expected. Several subjects occasionally expressed the desire to stay at the same difficulty level, but this option was not available. Additionally, it is not certain whether subjective choices are always perfectly reflected in measurable responses and thus whether completely accurate classification is even theoretically possible.

Use in Online Task Difficulty Adaptation

Since our classifiers can be used to determine whether a task is too easy or too hard, they can be directly used for online task difficulty adaptation. If the task is too hard, the system should decrease the difficulty. If the task is too easy, the system should increase the difficulty. Methods of increasing or decreasing difficulty must be defined in advance and can range from simple to very complex.

In our implementation, the user's direct input (task is too easy / too hard) was used to update the KALDA classifiers. However, in a real-world application, this information would not be available and the update process would need to use its own estimate of the current class rather than the actual class. This would need to be done carefully since such an approach can also amplify classification errors. If an incorrect class estimate is used to update the classifier, the classifier will become worse. One way to address this would be to generate a measure of how 'reliable' the estimate is. The system would then only update the classifier if the estimate was sufficiently reliable. A simple variant of this has already been tested and resulted in a classification accuracy of 72.1 % with only psychophysiological inputs (compared to 76.0 % when the user's actual input is available). Another possibility would be for the system to explicitly ask the user for input if certain potentially erroneous trends are detected (e.g. if the classifier repeatedly estimates that the task is too easy even though the user has reached a very high difficulty level).

CONCLUSION

We have demonstrated a classification method that can be used to estimate task difficulty in human-robot interaction based on motor actions and psychophysiology. The classifier can be recursively updated as new data becomes available, allowing it to gradually adapt to a particular user.

ACKNOWLEDGEMENTS

This work was funded by the [will be entered later if paper is accepted] and additionally supported by the [will be entered later if paper is accepted].

- 1. Bethel, C. L., Salomon, K., Murphy, R. R. and Burke, J. L. Survey of psychophysiology measurements applied to human-robot interaction. *16th IEEE International Conference on Robot and Human Interactive Communication* (2007), 732-737.
- Haarmann, A., Boucsein, W. and Schaefer, F. (2009). Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight. *Applied Ergonomics*, 40 (2009), 1026-1040.
- 3. Novak, D., Ziherl, J., Olenšek, A., Milavec, M., Podobnik, J., Mihelj, M. and Munih, M. (2010). Psychophysiological responses to robotic rehabilitation tasks in stroke. *IEEE Transactions on Neural Systems* and Rehabilitation Engineering, in press.
- 4. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17* (1996), 354-381.
- 5. van Reekum, C. M. and Johnstone, T. Psychophysiological responses to appraisal dimensions in a computer game. *Cognition and Emotion*, *18* (2004), 663-688.
- 6. Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R. and Pfurtscheller, G. Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. *IEEE Transactions on Biomedical Engineering*, 54 (2007), 550-556.

Evaluating Multimodal Human-Robot Interaction: A Case Study of an Early Humanoid Prototype

Gudberg K. Jonsson Human Behavior Laboratory University of Iceland Skipholt 50, 105 Reykjavik, Iceland gjonsson@hi.is

ABSTRACT

Multimodal natural behavior of humans presents a complex yet highly coordinated set of interacting processes. Providing robots with such interactive skills is a challenging and worthy goal and numerous such efforts are currently underway; evaluating the progress in this direction, however, continues to be a challenge. General methods for measuring the performance of artificially intelligent systems would be of great benefit to the research community. In this paper¹ we describe an approach to evaluating human-robot multimodal natural behavior. The approach is based on a detailed scoring and spatio-temporal analysis of the structure and patterning of live behavior, at multiple temporal scales, down to the decisecond level. The approach is tested in a case study involving an early virtual robot prototype, Gandalf, which is capable of real-time verbal and non-verbal interaction with people. Our analysis includes a comparison to a comparable human-human dyadic interaction scenario. Our main objective is to develop a methodology for comparing the quality and effectiveness of human-robot interaction between a wide variety of such systems. Early results indicate that our approach holds significant promise as a future methodology for evaluating complex systems that have a natural counterpart.

Author Keywords

T-patterns, human-robot interaction, multimodal dialogue, real-time, turn taking.

INTRODUCTION

In the fields of robotics and artificial intelligence work dealing with human-agent interaction, such as pet robot

Kristinn R. Thorisson

School of Computer Science University of Reykjavik Menntavegi 1, 101 Reykjavík, Iceland thorisson@ru.is

development for entertainment, robot humanoid applications [11], interactive teaching systems [4], and so on, has been gradually increasing. Domains where human behavior understanding is crucial (e.g., human-computer interaction, affective computing and social signal processing) require advanced pattern recognition techniques to automatically interpret the complex behavioral patterns of human-machine interaction. This is a challenging problem where many questions are still open, including the joint modeling of behavioral cues taking place at different time scales, the inherent uncertainty of machine-detectable evidence for human behavior, the mutual influence of people involved in interactions on each other, the presence of long-term dependencies and the important role of dynamics in human behavior understanding.

Earlier studies using the pattern detection approached described below have investigated the spontaneous play between the human and the AIBO robot and compared the temporal structure of the interaction with dog and AIBO in both children and adults [6]. The results indicated that both children and adults terminated T-patterns more frequently when playing with AIBO than when playing with the dog puppy, which suggest that the robot has a limited ability to engage in temporally structured behavioral interactions with humans. The authors argue that, as with other human studies, the results indicate that the temporal complexity of the interaction is good measure of the partner's attitude and conclude that more attention should be given to the robots' ability to engage in cooperative interaction with humans.

The RoboCup research community aims to developing artificial agents that will be able to mimic human behavioral patterns during soccer games and to meet this they have argued that there is a need for measuring and comparing the emerging behavior across populations (human vs. artificial) and to develop and standardize a particular pattern detection system that could be used by all research groups and that would further serve as a measure of efficiency of research improvements. For this purpose the t-pattern detection approach, described below, has been suggested with the argument that is has already been successfully applied to the analyses of human-animal and human-robot interactions, and real-life human soccer matches [2]. The

¹ This work was funded in part by the European Union FP7 program through the project HUMANOBS – Humanoids that Learn Socio-Communicative Skills Through Observation, grant number 231453.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. User interacting with Gandalf. Gandalf's behavior proceeds at natural human-human real-time dialogue speeds (notice glance in response to deictic gesture).

objective of the present work is to develop a general methodology for evaluating multimodal human-robot interaction. The approach is based on the analysis of the complex structure/patterns of verbal and non-verbal behavior in multimodal human-humanoid interaction. The interaction is compared to comparable human-human interaction. The study is a part of an ongoing and broader research concerning the development of a prototype for evaluating multimodal human-robot interaction where the goal is to develop new cognitive architectural principles to allow intelligent agents to learn socio-communicative skills by observing and imitating people.

MULTIMODAL HUMANOID SYSTEM

As an initial system to study we have chosen the Gandalf system - an early prototype of an artificial agent capable of broad multimodal real-time interaction with people [12,13]. This system was chosen both because of its breadth of behaviors and number of modes presented both in the input and the output, as well as the number and quality of video recordings available.

Gandalf is a communicative humanoid built in the Ymir framework at M.I.T. between 1992 and 1996 [12,13]. It represents a distributed, modular approach that can be used to create autonomous characters capable of full-duplex multimodal perception and action generation. Gandalf is capable of fluid turn-taking and unscripted, task-oriented dialogue; he perceives natural language, natural prosody and free-form gesture and body language, and conducts natural, multimodal dialogue with a human. Gandalf, an expert in the solar system, can perceive and interpret several thousand utterances related to the topic; two kinds of general manual gesture (deictic and iconic); contextual body, head and gaze direction; and speech prosody. He contextually generates several thousand utterances related the topic, many types of facial expressions, two types of manual gesture (beat and deictic), as well as head and gaze direction. Computer-naïve users' rated Gandalf highly on believability, language ability and interaction smoothness [12]: In less than a minute people communicate naturally

and take turns efficiently. In evaluation questionnaires collected from users Gandalf's "interactivity" is rated somewhere between that of a dog and a real human. The original Ymir/Gandalf system ran on 8 networked workstations; more recent incarnations of Ymir have been used in increasingly complex dialogue systems [4] and run on humanoid robots, including Honda ASIMO [11].

CODING AND DATA ANALYSIS

Ten dyadic interactions between human user and Gandalf were transcribed and analyzed using the Theme software [7,8,9]. The results were compared to results obtained from several studies on human-human dyadic interactions involving interaction between doctor-patient; friends, strangers and students [5].

The interactions were transcribed using ThemeCoder. A category system for non-verbal behavior was adopted from McGrew [10], and verbal categories from Bromberg & Landré [1]. The transcribed records were then analyzed using Theme 5.0 [7,8,9]. The basic assumption of this methodological approach is that the temporal structure of a complex behavioral system is largely unknown, but may involve a set of particular type of repeated temporal patterns (T-patterns; [7,8,9]) composed of simpler directly distinguishable event-types, which are coded in terms of their beginning and end points (such as "boy begins deictic gesture" or "girl ends speaking"). The kind of behavior record (as set of time point series or occurrence times series) that results from such coding of behavior within a particular observation period (here called T-data) constitutes the input to the T-pattern definition and detection algorithms.

Within a given observation period, if two actions, A and B, occur repeatedly in that order or concurrently, and found more often than expected by chance, they are said to form a minimal T-pattern (AB), (assuming as h0 independent distributions for A and B, there is approximately the same time distance (called critical interval, CI) between them). Instances of A and B related by that approximate distance then constitute occurrence of the (AB) T-pattern and its occurrence times are added to the original data. More complex T-patterns are consequently gradually detected as patterns of simpler already detected patterns through a hierarchical bottom-up detection procedure (see a simple example in Fig. 2). Pairs (patterns) of pairs may thus be detected, for example, ((AB)(CD)), (A(KN))(RP)), etc. Special algorithms deal with potential combinatorial explosions due to redundant and partial detection of the same patterns using an evolution algorithm (completeness competition), which compares all detected patterns and lets only the most complete patterns survive. As any basic time unit may be used, T-patterns are in principle scaleindependent, while only a limited range of basic unit size is relevant in each concrete study. This methodology is explained in detail in Magnusson 1996 paper [7].

The results for the analysis depend on two factors, the source material itself, and how well it is coded, and the



Figure 2. The lower part of this figure shows a simple realtime behavior record containing a few occurrences of several event types, a, b, c, d, & e, indicating their respective instances within the observation period. The upper line is identical to the lower one, except that occurrences of k and w have been removed. A simple t-pattern (abcd) then appears, that was difficult to see when the other events were present.

settings of particular analysis parameters. The coding was done by experienced coders with years of experience in using the Theme software¹. We have chosen to code parameters that were a special target of the system design. As multimodal coordination is a key feature of the Gandalf system, we have included features from gaze, manual gesture, head movement, and speech. An analysis of the timing of events is of course part of the detailed comparison, as the T-patterns method analyzes spatiotemporal patterns. The following parameter settings were used for all analysis: Minimum number of occurrences set at 3 and significance level set at .0005 (other search values set at default: [7,8,9]).

RESULTS OF ANALYSIS

A high number of temporal patterns were detected in the Gandalf data set involving turn-taking. The number, frequency and complexity of detected patterns indicate that behavior was highly synchronized in all situations. This synchrony was found to exist on different levels, with highly complex time structures that extended over considerable time spans where some of the patterns occurred in a cyclical fashion. The Gandalf data is most similar to human-human scenarios that involve interviews and interrogations. The reason is that Gandalf is designed to be a passive "guide to the solar system" - the dialogue exclusively via human-initiated proceeds dyadic interactions. Many of the patterns seen are the same as those found in the human-human data from interviews/interrogations. Patterns involving coordinated gaze and gesture were found to be similar to human-human dialogue but some were found exclusively in the Gandalf scenarios and others in both human-human and Gandalfhuman dyads. For those patterns the content and order of the turn-taking events was the same but differences were



Figure 3. Example of a complex pattern found in the humanagent dyads with 10 events occurring in the same order with significantly similar time interval between each event type occurrence: 1. Human ends looking at big screen. 2. Human begins to look at agent. 3. Human begins to ask agent for information. 4. Agent begins to look at user. 5. Agent begins to reply. 6. Human ends looking at agent. 7. Human begins to look at big screen. 8. Agent begins finger tapping. 9. Agent ends looking at user. 10. Agent begins to look at big screen.

found in the interval between even types. Preliminary results indicate that the interval between question and answer was at least 15% less in friendship dyads than in the Gandalf scenarios. The interval between the turn-taking events was found to be more similar between the Gandalf scenarios and interview and stranger dyads.

On average the Gandalf scenario patterns seem to be somewhat slower than human-human, which can likely be explained by a longer average duration between critical elements of the turn taking system, especially the speech recognition². A closer comparison to human-human data suggests that the Gandalf data has the highest similarity to dyads involving interaction between (human) friends, higher than doctor-patient, dialogue between strangers. Turn-taking in the Gandalf data reaches a "mean level" of patterning quickly, as is the case in dyadic interaction between friends [5]. In all Gandalf dyads we find examples of a "patterning-growth" period, were the patterns in many cases developed into highly complex structures, comparable to human-human in complexity.In the Gandalf-human dyads the human initiates over 99% of patterns detected and seems to controls the growth and synchronization or "beat" of the interaction. In that sense the human-agent interaction has some similarities with doctor-patient interaction. The duration and patterning of manual gesture in relation to speech and turn-taking, seems to be highly involved in the patterning of the turn-taking system in all situations analyzed. In human-human dyads we find a richer repertoire of gestures - higher number of different gestures used, even though the functionality seems to be the same.

¹ The inter-observer reliability of coding has been assessed in prior work and was not measured here; our prior results indicate that scores was 0.74 for all classes of behavior, but over .85 for "looking behavior" and "verbal behavior".

² Gandalf's speech recognition is the largest bottleneck, taking on average around 1.5 to 2 seconds to process the utterances. Other processes typically take only a fraction of that to produce output, such as e.g. Gandalf's ability to gaze in the direction of pointing, but sometimes a processes producing visible behaviors are serially dependent.

We analyzed the coordination of head, hand and gaze in relation to speech and turn-taking. When looking at simple and short turn-taking patterns results indicate that similar structures are detected across all dyads analyzed. More complex patterns are though detected in the human-human dyads than in the Gandalf scenario, partly explained by the limited repertoire of behavior displayed by Galdalf. When looking at head-turn and gaze, apart from different interval detected between event in the Gandalf scenarios and human-human, we also find that the structure of Gandalf's looking behavior is more similar to humans with moderate and high self-esteem and extraverts [5] than those with low self-esteem, even though the frequency of "looking at partner" behavior is less. Human gaze patterns of people with low self-esteem and/or introverts are of a lower frequency and duration than that of those who have high self-esteem and/or are classified as extroverts [5].

METHODOLOGY: RESULTS

The preliminary investigations described have reinforced the authors' belief that T-pattern identification has good potential as an effective research tool in AI/Robotics. One reason is the fact that the T-pattern detection correctly identified the turn-taking patterns by the system, and allowed us to compare them to comparable naturally occurring data at a fine level of detail. In addition, the method detects more complex structures than has been possible before and, as in the present study, it detected patterns indicating when the human-agent turn-taking pattern were about to fail. One potential concern is that the results are strongly affected by the choice of parameter settings for the T-pattern analysis, which could make comparisons between researchers more difficult. In the present study we were careful to set these parameters at the same settings for all analysis. In the future it should be investigated to link them to the actual data, to make crossevaluations also possible.

CONCLUSION

Our approach for studying human-robot interaction shows great promise. It is especially relevant when - as we do here - it is possible to compare the results to real human-human data of comparable circumstances. The results are in line with prior work using T-patterns and the Theme software for analyzing spatio-temporal patterns. More experience must of course be collected on the use of these tools, and they must be applied to a broader range of systems. We believe that the approach presents the way towards the development of a set of methods and indexes that will improve quantitative and qualitative comparison between large dataset of different artificial intelligence systems. The identification of complex and repeated patterns, which are not identifiable through simple observation, has great benefits for the development of an index aimed to advance the continued development of artificial agents and robots.

- 1. Bromberg, M. & Landré, A. Analyse de la Structure Interactionnelle et des Stratégies discursive dans un talkshow. Psychologie Francaise, *38*, 2 (1993), 99-109.
- Hadzibeganovic, T., Van den Noort, M.W.M.L., Cannas, S.A., Bosch, M.P.C., Jonsson, G.K., & Magnusson, M.S. The missing neurocognitive and artificial general intelligence bases of RoboCup research: What still needs to be done before 2050? Proceedings of the 3rd Austrian RoboCup Workshop 2008. Villach, Austria.
- 3. Johnson, W. L., Vilhjalmsson, H. and Marsella, M. Serious Games for Language Learning: How Much Game, How Much AI? 12th International Conference on Artificial Intelligence in Education, 2005).
- Jonsdottir, G. R. and Thórisson, K. R. Teaching Computers to Conduct Spoken Interviews: Breaking the Realtime Barrier With Learning. IVA '09, 446–459; LNAI 5773.
- Jonsson, G.K. Personality and Self-Esteem in Social Interaction. In From Communication to Presence: Cognition, Emotions and Culture towards the Ultimate Communicative Experience. Edited by Giuseppe Riva et al. IOS Press, 2006. ISBN 1-58603-662-9.
- Kerepesi, A., Kubinyi, E., Jonsson, G.K., Magnusson, M.S., Miklósi, Á. Behavioural comparison of humananimal (dog) and human-robot (AIBO) interactions. *Behavioral Processes*, 73 (2006), 92-99.
- Magnusson, M.S. "Hidden Real-Time Patterns in Intra- and Inter-Individual Behavior: Description and Detection." *European Journal of Psychological Assessment*, 12, 2 (1996). 112-123.
- 8. Magnusson, M.S. Discovering hidden time Patterns in Behavior: T-patterns and their detection. *Behavior Research Methods, Instruments & Computers, 32* (2000), 93-110.
- Magnusson, M.S. Structure and Communication in Interaction. In G. Riva, M.T. Anguera, B.K. Wiederhold, F. Mantovani (eds.) 2006. From Communication to Presence: Cognition, Emotions and Culture Towards the Ultimate Communicative Experience. Útgefandi: IOS Press October 1, 2006.
- McGrew, W.C. An Ethological Study of Children's Behaviour. New York: Lawrence Erlbaum Associates, 1972.
- Ng-Thow-Hing, V., K. R. Thórisson, R. K. Sarvadevabhatla, J. Wormer, T. List. The Cognitive Map architecture for facilitating human-robot interaction in humanoid robots. *IEEE Robotics & Automation*, 16, 1 (2009), 55-66.
- Thorisson, K. R. Communicative Humanoids: A Computational Model of Psycho-Social Dialogue Skills. Unpublished Ph.D. Thesis, Media Laboratory, Massachusetts Institute of Technology, 1996.
- Thorisson, K. R. A Mind Model for Multimodal Communicative Creatures and Humanoids. International *Journal of Applied Artificial Intelligence*, 13, 4-5 (1999), 449-486.
Measurement of Feet-Ground Interaction for Real-Time Person Localization, Step Characterization, and Gaming

Chris Varekamp chris.varekamp@philips.com Patrick Vandewalle patrick.vandewalle@philips.com

Philips Research, High Tech Campus 36 5656AE Eindhoven, The Netherlands

ABSTRACT

We introduce a setup for detailed and unobtrusive feet mapping in a room. Two cameras arranged in a stereo pair detect when an object touches ground by analyzing the occlusions on a fluorescent tape that is attached to the baseboard of a room. The disparity between both cameras allows localization of a person's feet and the calculation of step size and walking speed. People are separated from furniture such as chairs and tables by studying the occlusion duration. We present and discuss data-association and filtering algorithms and the algorithms that are needed for presence detection, step characterization and gaming.

ACM Classification Keywords

H5.1.b Artificial, augmented, and virtual realities; H.5.m. Information interfaces and representation: Miscellaneous, I.2.10.a 3D/stereo scene analysis

INTRODUCTION

Our aim is to create a low-cost measurement setup for longterm installation in homes such that walking behavior can be studied on natural home surfaces or gaming can be done in-home. We introduce a measurement setup that allows unobtrusive measurement for this purpose. The setup consists of two cameras (stereo pair) on one side and an elongated fluorescent tape on the other side of a room. Both the cameras and the tape are mounted near the floor. The tape height is only 2 cm to be minimally disturbing someone that is present in the room. The measurement geometry is shown in Figure 1. The setup does not require on-body sensors or markers. We calculate and visualize in real-time the 2D position of feet and other objects that touch the floor. When observed over time, this measurement provides information on the walking behavior of people. Our measurement setup has potential applications ranging from presence detection to in-home gaming. Most existing systems that measure feet-ground

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

interaction are intended for professional use, such as systems that measure gait. We are interested in in-home activity monitoring using for instance step size as an indicator and not for clinical research. The gaming industry is working on unobtrusive solutions for gaming. Notably, Microsoft plans to use a depth sensor such that a game controller is no longer required [1].

This allows for the use of gestures, and also for kicking a ball as we will also demonstrate in this paper. Currently we cannot compare the accuracy of calculated feet positions with those calculated by the Xbox sensor since that sensor is not on the market at the time of writing. Video based human gait extraction typically exploits strong prior shape information. The authors in [2] train a 12 parameter model and fit that to extracted silhouettes. They use strong prior information in an articulated 2D model to compensate for noisy silhouettes (e.g. from background subtraction outdoor). However, they consider only walkers moving perpendicular to the camera. For our soccer game (see further) the player can also face the camera. The use of cameras in combination with a background that has



Figure 1. Top view of a room. Two cameras are mounted a distance *B* apart with the origin and camera 1. Image coordinate *v* is assumed to lie parallel to world coordinate *y*. World coordinate *z* points towards the tape. An object that touches ground occludes the reflective tape. The object position $(y^{(1)}, z^{(1)})$ on the ground is calculated from the shift $(y^{(1)}, y^{(2)})$ in the image *v*-coordinate of the occluded area

between the two camera images.

specified optical properties is known from the movie industry (blue screen or green screen) and from the monitoring of inventory [3]. We propose a new background based technique for a room that is minimally intrusive and can be used to calculate the 2D position of people and characteristics such as step size and walking speed. In the remainder of this paper we discuss the measurement setup, treat the required signal processing and develop specific application algorithms.

MATERIALS AND METHODS

Measurement Setup

In our experiments, we use two QuickCam Fusion webcams (Logitech). The cameras capture images at a resolution of 640×480 pixels at 30 frames per second. The cameras are placed 15 cm apart and mounted 1.5 cm above the ground. At the opposite side of the room we attach a red fluorescent tape (HPX). The tape is over 3 m long and has a height of only 2 cm. We implement algorithms in C building on data structures, video input/output and plotting functions from the open source computer vision library *OpenCV* originally developed at Intel and now hosted by Willow Garage [4].

Tape Detection

The fluorescent tape serves as a known background against which objects such as human feet are observed. We construct a vector of tape pixels for each camera using a fixed threshold that differs for each color channel (rgb). We use $r = \{119...255\}$ and $g,b = \{0...99\}$ as selection criterion for 'tape pixels'. Since the tape is red, the red channel is required to take large values whereas the green and blue channels are required to take small values. In general, the selection of these values depends both on the camera properties, illumination, and on the specific tape that is used. Since the tape touches the ground surface, and the ground surface is flat, the tape pixels lay on a straight line in each camera image. We fit this line to the tape pixels in each image. During this fitting step we clear the carpet of objects. Currently, carpet roughness, which could influence our measurement, is not accounted for.

Object Segmentation

For each tape pixel we detect occluded line segments using again knowledge of the tape color. The start and stop coordinates of each line segment are stored. Figure 2 shows two camera images with the detected line segments that in this case correspond to a foot that occludes the fluorescent tape.

Disparity Estimation

In order to produce a map of a person's feet on the ground we need the disparity of each object between the two cameras. Here we account for the possibility of a false detection due to image noise. If each image contains the same number of objects we associate objects in the two cameras using their order of occurrence along the tape. For each object we calculate the average pixel *v*-coordinate in both cameras (see the geometry in Figure 1). We then



Figure 2. Left and right stereo images of a a walking person with one foot touching ground. The corresponding occlusion segments of the red fluorescent tape are shown in green below each images of the stereo pair.

calculate disparity $v^{(2)} - v^{(1)}$ of the corresponding points between the two cameras of the stereo pair.

Calculation of Object Position and Size

The observed disparity allows the calculation of depth z. The camera model geometry in Figure 1 is a simplified version of the model given by Forsyth and Ponce [5, p.29-31]. Using this simplified model, the *z*-coordinate follows as:

$$z^{(1)} = \frac{Bk}{v^{(1)} - v^{(2)}},\tag{1}$$

where *B* is the distance between the cameras [m] and *k* is a constant that depends on the camera focal length and pixel size. The superscript ⁽¹⁾ in the above notation stresses that the origin of our coordinate system lies at the lens centre point of the first camera. When *z* is known, the *y*-coordinate follows as:

$$y^{(1)} = \frac{z^{(1)}v^{(1)}}{k}.$$
 (2)

We can now create a 2D map showing a top view of the room with all objects that touch the ground. Such a map is shown in Figure 3. The sizes of the circles can differ since one foot can be fully on the ground, while the other foot is in the process of being lifted. We estimate object size from the end points v_{start} and v_{stop} of a line segment:

$$\Delta^{(1)} = \frac{z^{(1)}}{k} \left| v_{\text{start}}^{(1)} - v_{\text{stop}}^{(1)} \right|.$$
(3)

Calibration

Our aim is to be able to map feet positions in a room's coordinate system. We need to determine the baseline B and the camera parameter k. Although B can be known when the cameras are mounted in fixed positions, we consider it unknown since the two cameras may need to be installed



Figure 3. Top view of the room with the position of two feet on the ground. During step analysis, the two feet are recognized as one single step and connected by a line.

separately in the baseboard of a room, in which case calibration follows later.

To estimate *B* and *k*, black calibration cylinders are placed at known coordinates (y,z). Given these known coordinates, we first solve for parameter *k* by minimizing the sum of squared errors in the *y*-coordinate. Using equation (2) the least-squares estimate for *k* is:

$$\hat{k} = \frac{\sum_{i=1}^{N} z_i^{(1)} v_i^{(1)}}{\sum_{i=1}^{N} y_i^{(1)}}.$$
(4)

Given that k is now known, the baseline B can be calculated after inserting the estimated value for k in equation (1). The least squares estimate for B then follows as:

$$\hat{B} = \frac{\sum_{i=1}^{N} z_i^{(1)}}{\hat{k} \sum_{i=1}^{N} \left(\frac{1}{v_i^{(1)} - v_i^{(2)}} \right)}.$$
(5)

Note that the above least-squares method is sub-optimal since it does not minimize the Euclidian distance between the true and estimated (y,z)-coordinates. However, the above approach that treats each coordinate separately has the advantage of providing a closed-form solution.

Over-Time Association and Filtering

We link feet measurements over time to remove noise. This temporal linking is also necessary to determine when a foot is placed on the ground and when it is lifted. Data association is explained in Figure 4. Let **p** denote the 2D location of an object. Now assume that two objects are present at time t and that one new measurement **q** is detected at time $t+\Delta t$. The new measurement **q** lies close to **p**₂ and far from **p**₁. We use the new measurement to update and object's location if it lies less than 10 cm from that object. Thus **q**₃ is used to update **p**₂ but not to update **p**₁ (see Figure 4). After this data-association step, **p**₂ is updated using a low-pass filter. In this case the estimate for **p**₂ is updated according to:

$$\mathbf{p}_{2}^{(t+\Delta t)} = \mathbf{p}_{2}^{(t)} \boldsymbol{\alpha} + \mathbf{q}^{t+\Delta t} (1-\boldsymbol{\alpha}), \qquad (6)$$

where α =0.1 controls the amount of filtering. Note that position **p**₁ stops to exist since no new measurement has been found in its vicinity. The same filtering operation is



Figure 4. Illustration of the over-time data association of detected positions for the purpose of temporal analysis and filtering.

also applied to update the size estimate. Currently we do not explicitly treat the feet of multiple people with the consequence that errors occur when multiple persons come close.

Step Detection

During a typical step, first the heel touches ground, then the full foot, and finally when the foot is lifted, the foot leaves the ground via the toes. These events correspond to different spatial positions on the ground. We use a lead time of 0.2 s (6 frames on average) to allow the position filter to do its work before we plot the foot position. We have found that this lead time is necessary to avoid introducing multiple positions during the step that do not correspond to a foot touching ground. The occurrence of these non-touching positions could be due to a small clearance distance [6]. The straight line that we fit to the 2 cm high tape will probably lie approximately 1 cm above the ground surface. A smaller clearance is common as shown in [6, p.193, fig.3]. To recognize a step we introduce a minimum step size of 0.3 m. If a detected location has a duration that exceeds the lead time and lies further than 0.3 m from the previous foot location then we recognize it as the next foot of a step.

APPLICATION ALGORITHMS

In this section we develop various application algorithms to show how the measurements must be interpreted for practical use. All application algorithms run in real-time on a conventional laptop.

Person Localization

Presence detection is relevant for home and building



Figure 5. Temporal occlusion analysis as a basis for the separation of dynamic and static objects. One dynamic object is ignored because it has one end-point for which the position cannot be determined.



Figure 6. Positioning a person's feet in a room that is cluttered with furniture. Due to occlusion a foot position may be missed. However, if detected, the positions are not influenced by the presence of static objects.

automation such as lighting control. We wish to position people in a room that is possibly cluttered by tables and/or chairs. With the processing discussed so far, this can be problematic. A person's foot placed behind a chair can result in a single silhouette instead of two separate ones and the estimated position will lie somewhere between the foot and the chair. To solve this problem we store the duration in seconds that a 'tape pixel' has already been occluded. We then produce two object segmentations: one containing occlusions with duration longer than 5 seconds and another that consists of occlusions with duration of less than 5 seconds (Figure 5). Each segmentation is input to position/size estimation and data-association/filtering. Figure 6 shows an example classification where a person passes behind a chair.

Soccer

We implemented a real-time soccer game. To realize a virtual ball-kick, the velocity vector of each foot is calculated. To increase accuracy we exploit the typical smooth leg motion and filter the velocity vector temporally:

$$\mathbf{v}^{(t+\Delta t)} = \mathbf{v}^{(t)}\boldsymbol{\beta} + (\mathbf{p}^{(t+\Delta t)} - \mathbf{p}^{(t)})(1-\boldsymbol{\beta}).$$
(7)

where β =0.5 weighs the contribution from the previous foot motion vector v^(t+\Delta t). With this filtering operation we achieve a realistic kick. If one of the two feet hits the virtual ball, we transfer the foot-velocity vector to the ball and let it move according to the motion vector at the time of contact (Figure 7).

Tap Dance

We pre-record sounds that resemble the effect of a *heel* hitting a wooden floor and a different sound for a *toe* hitting a wooden floor. The sounds play in real-time in response to the feet actions. To determine whether heal or toe hit the



Figure 7. Kicking a virtual soccer ball (green circle) with one of two feet (blue circle) during a simple soccer play.

ground, we compare the *z*-coordinate of each foot when it hits the ground with the mean *z*-coordinate of both feet that we track over time. If the tap dancer faces roughly in the negative *z*-direction, the classification works satisfactorily.

CONCLUSIONS AND FUTURE WORK

We have introduced a low-cost setup for measuring groundfeet dynamics. The technique can determine presence, foot location, step size, walk direction and speed and can be meaningful for in-home gaming.

The setup was tested using an image resolution of 640×480 . The image resolution could be further increased to achieve higher accuracy of feet maps. The proposed measurement system has a number of advantages over the use of traditional camera setups. First, the use of reflective tape as background makes the measurement robust. Second, placement of both camera and tape near the floor allows straightforward selection of the events of interest, i.e. walking patterns. Third, the required processing is rather limited, which could allow for energy efficient measurement. There are also some drawbacks of the current setup. For instance, tape color, camera sensor, and algorithm thresholds need to match. For flexible use, we need a higher robustness to variations in lighting. Switching the room lighting on or off can currently stop the application (e.g. soccer) from functioning properly. A remaining issue is camera calibration since the measurement accuracy must be known for applications that measure step size or walking speed. Our simple camera model ignores small rotations. In future work, we wish to investigate the use of a (removable) calibration pattern on the tape and estimate camera rotations in addition to baseline B and scale parameter k. Another remaining issue is the timing of captured image frames. The cameras in our current setup cannot be triggered for accurate timing of frames. Such cameras exist and should be used for applications that require temporal accuracy.

REFERENCES

- 1. Project Natal, <u>www.xbox.com/en-US/live/projectnatal/</u>.
- Zhou, Z., Prugel-Bennett, A., and Damper, R.I. A Bayesian Framework for Extracting Human Gait Using Strong Prior Knowledge. *IEEE. Trans. on Pattern Anal. and Machine Intelligence*, vol.28, no.11, pp. 1738-1752.
- Cato, T.R. and Zimmerman, T.G. Using Cameras to Monitor Actual Inventory. US Patent Application 2009/0121017.
- 4. OpenCV at Willow Garage: http://www.willowgarage.com/pages/software/opencv
- 5. Forsyth, D.A. and Ponce, J. *Computer Vision: A Modern Approach.* Prentice Hall, 2003.
- Begg, R., Best, R., Dell'Oro, L., Taylor, S. Minimum foot clearance during walking: Strategies for the minimization of trip-related falls. *Gait & Posture*, 25 (2007), 191-198.

Web Based Measuring System for Health Monitoring at Home

Luca Rizzi luca.rizzi@cetma.it Cinzia Dinardo

Nicola Savino

Gianluca Ciullo

CETMA Consortium s.s.7 Appia Km 706+030 - c/o Cittadella della Ricerca 72100 Brindisi, Italy

ABSTRACT

The aim of this study is to improve the design of a monitoring system for early dehospitalization of patients suffering from chronic diseases, excluding advanced stage conditions and those in death risk, and also ensuring continuous assistance at home with monitoring, interaction and intervention. We actually reached the goal of completely reproducing hospital visits at the household, moreover the system could be used at a hospital in the same way of standard "vital signs monitor" networks. The developed system integrates wired vital signs sensors, adhoc networking, video communication system and web portal technology to allow remote monitoring of patient's vital-sign status and intervention with therapies by activating medical equipment onboard. An important goal of this study is to monitor the effectiveness of the therapy applied on the patient. Furthermore the video - interaction channel will be an important instrument for early diagnosis of cognitive diseases in the elderly.

Author Keywords

Telemedicine, e-Health, vital signs monitor, health network, ageing, user interfaces, touch monitor.

ACM Classification Keywords

H.1.2 User/Machine Systems: Human factors, H.3.5 Online Information Services: Web-based services, H.5.2 User Interfaces: Prototyping

INTRODUCTION

The H@H (Hospital at Home) Patient Monitoring System combines the proved power of video health assistance with remote patient monitoring. Some systems are already available on the market and have been described in many research projects [1], in this article we will discuss benefits that could be achieved thanks to innovative design solution.

The patient monitoring, coupled with interactive video, enables clinicians to monitor patient's health status, identify instability, provide real time remote clinical intervention in order to stabilize patients by activating the medical system available onboard, applying oxygen-therapy and medicalair and vacuum. The system offers all the advanced monitoring features of "patient vital signs monitor" in realtime with the additional ability to turn on an interactive video session when patient's condition demands it. When the patient's diagnosis, acuity level and condition changes it can create situations where a real-time interactive video visit is needed. The H@H patient monitoring system offers to the healthcare organizations the ability to implement new strategic capabilities. The application connected to a server station in dedicated hospitals or a clinician's office on one side and with a large number of H@H systems, in patient's homes on the other, using ordinary telephone lines or broadband connection.

The following will describe some of the most important aspects that were developed to improve usability of the system, reproduce visits and create *at-home measuring session*, like at a hospital.

SECTIONS

The device is really innovative compared to others already available on the market because it allows a continuous monitoring (24 hours on 24) as well as remote intervention. Doctors and health operators can administer therapies to the patient by activating medical equipments onboard. Before activating therapies it will be very important to ensure a correct control of the status of the patient.

Vital Sings Monitoring

The remote monitoring of a patient based at home involves the vital parameters control which represents the state of health and could be done in the same way in which it takes place in protected Hospitals. To fully implement the objective it is necessary to bring to the patient's home all

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. WEB based networks for early dehospitalization and patients monitoring.

the instruments necessary to measure his physical state, and all the technologies necessary to verify the outcome of the applied therapy, to command medical gases valves or to apply vacuum. All the system should be easy to use, effective in the monitoring of therapy and comfortable even for household installation [2]. For example, following some patients and healthcare suggestions, we decided to implement the vital signal status with wired sensors like at a hospital. This solution avoids recharging problems and would be more useful for in-hospital application besides allowing better sensors diagnosis in case of system's errors.

The entire project aims to examine, in terms of performance, how we can achieve a system for distance monitoring of patients with chronic disease but not in risk of death. With the designed system we would be able to define and create a new protocol of interaction between central hospitals and local hospitals and implement a solution that integrates hardware and software to respond to those needs. Doctors, Surgeons and Nurses in general can perform daily visits with remote access. Few health operators could follow several number of patients distributed on the territory.

Video-Communication for Patient Control

Thanks to the dedicated video-call system the H@H allows monitoring of the effectiveness of the therapy. The idea is to improve the quality of video-communication using High Quality cams, thus permitting a real "remote visit" into patient's homes. A good quality video-signal allows the doctor to see the details, to evidence problems and to better understand the emotional state of the patient. The patient is also advantaged by the better video-signal together with the use of a large monitor (15") making it possible for one to feel more involved in their care protocol as well as to interact with their doctor. Monitoring the patient during therapy is an application which can evidence problems and identify the ability and the independence of the patient in their self-care. The monitor will be connected to the system



Figure 2. Arm positions for patient and caregivers.

with articulated arm that allows the patient to feel like in real *on-bed visit*, the same monitor could also be used by caregivers.

In addition to cognitive decline, the video interaction channel can identify, by continuous monitoring and recording the evaluation of patient's status, at an early stage the signals for many elderly cognitive disease like: Dementia, Mild Cognitive Impairment, Alzheimer's Disease which are usually expressed by significant changes in personality and behavior [3] (e.g., apathy, disinhibition/social misconduct, mood changes, lack of empathy, poor judgment).

A future development of the system could include a test to monitor the cognitive disease also an exercise, some entertainment and games to improve memory and to slow the progress of the disease.

Easy-to-Use Graphic Interface

Interactive product design should be based not only on the interface graphic appearance but mostly on its usability. An excellent graphic interface should be essential in functionality as well as properly organized in each graphical element. All functionality should be placed as a result of defined visual perceptual hierarchy [4,5].



Figure 3. Touch screen interface.

Furthermore the touch screen technology involves all the problems related while the visual perception is not supported by tactile perception, simply reminding that "touch screen buttons" have no tactile feed-back. In order to obtain a high level of usability the project of the system is developed following the "Perceptibility Principle" that suggest highlighting, and making easy to perceive, all information that is necessary for product use [6].

To meet this principle we define commands and information's hierarchy through:

- Requirements analysis;
- Task analysis
- Scenarios identification
- Use of Conventions.

We finally designed the graphic interface thanks to a clear commands hierarchy definition, in-depth knowledge of different typology of users who will interact with the system and a clear definition of tasks and scenarios. Particularly the analysis of tasks has determined an important reduction of actions that users will have to play during the interaction, allowing an overall simplification of the structure. We actually believe that using constant users feed-back we can have an important improvement of the interface. The most interesting thing is that the feed-back is not only active (driven from the user), but can be monitored also thanks to health operators' collaboration. In that case the role of the video-signal will be fundamental for any future analysis of user actions.

CONCLUSION

With this new system it will be possible to reduce healthcare expenses and to improve patient quality of life, ensuring continuous assistance at home thanks to the active monitoring, interaction and intervention. The continuous measuring of vital signs correlated with video supervision of the activities and response of patients to the applied therapy can provide all the information that hospital personnel needs to achieve early diagnosis. Following a shared protocol, like the HL7 standard [7], the system provides continuous monitoring, executing the video-talks, by "administering therapies", requiring diagnostic tests and if necessary requiring intervention or hospitalization. All the data will be stored and can be easily exchanged with the Hospital Data System [8]. Furthermore the household workstation will be easy to install and to configure. The remote services offered will be differentiated according to the telematics infrastructure available guaranteeing the minimum system requirements in every situation.

The H@H bedside System is a product development project of Industrial Design and Information Technology departments of CETMA consortium [9]. The analysis of the feasibility and the implementation of the first prototype have been carried out by Item Oxygen S.r.l. [10].

REFERENCES

- H. Alemdar, C. Ersoy, Wireless sensor networks for healthcare: A survey, *Comput. Netw.* (2010), doi: 10.1016/j.comnet.2010.05.003
- 2. Jef Raskin "Interfacce a misura d'uomo", Apogeo 2003
- 3. Tecnologie informatiche e utenza debole. La progettazione ergonomica dei siti web e delle postazioni di lavoro per i disabili. *Il Sole 24 Ore Pirola* (2002) Codice EAN: 9788832447552
- 4. Progetto Lettura Agevolata www2.comune.venezia.it/letturagevolata
- 5. Hyperlabs empowering users www.hyperlabs.net
- BAECKER, R., & Buxton, W., Readings in Human-Computer Interaction: A Multidisciplinary Approach, Morgan-Kaufmann Publishers, Los Altos, CA, 1987.
- Teresa Andreoli, Psy.D. Psychology Assistant Cognitive Disorders Among the Elderly, Lic #PSB 31633, BCIA-C #3949 Fellow *Brain Therapy Center* www.brain-injury-therapy.com/articles/dementia.htm
- 8. Health Level Seven International global authority on standards for interoperability of health information technology. www.hl7.org
- 9. Industrial Design and Information Technology departments of CETMA consortium Brindisi-Italy. www.cetma.it
- 10. Item Oxygen S.r.l., Altamura (Bari)-Italy. www.item-medical.it

All figures are designed by CETMA consortium using images from Office Online ClipArt (http://office.microsoft.com/clipart).

Towards the Use of Psychological Variables in User Profiling

Ricardo Luiz

ricardo.luiz@ydreams.com

Marilia Felismino marilia.felismino@ydreams.com

Edifício YDreams,

Madan Parque - Sul, P-2825-149 Caparica, Portugal

Marta Castela

martavcastela@gmail.com

Pedro Marques

pedro.a.r.marques@gmail.com

Faculdade de Psicologia e de Ciências da Educação, Alameda da Universidade, P-1649-013 Lisboa, Portugal

ABSTRACT

The scope of this study is the development of a questionnaire that can assess psychological variables relevant in the context of human-technology interaction. The final goal would be to use these variables as a basis for user-profiling tasks in the specific case of YDreams' Showroom. In this sense, we propose three variables computer self-efficacy, computer anxiety and interaction with avatars - that can be relevant in the prediction of the interaction between humans and novel technologies, such as YDreams' interactive installations based on Augmented Reality. The possibility of extracting these variables indirectly (that is, without the questionnaire) through different tasks and/or time latencies is explored, and some examples are proposed. The questionnaire was developed for the specific case of YDreams' Showroom, but this method could be applied more widely to the field of humancomputer interaction (HCI), and therefore used in different contexts.

INTRODUCTION

This project was developed by YDreams, SA in collaboration with the Faculty of Psychology of the University of Lisbon. YDreams is a global company that is redefining the concept of interactivity, with a focus on the exciting field of Augmented Reality technologies. Over the last few years, YDreams has been developing full-scale interactive environments (from stores to museum

exhibitions), products and intellectual property, combining technology and design. This company has a showroom that is frequently visited by customers, schools and media, among others. These visits are guided by salespeople that present the same interactive installations presented there but in a random order for all visits. As the users are naturally different from each other, the aim of this project is to study the best way to create an immersive environment in the showroom that adapts the visit according to the different user types, in order to maximize their satisfaction. To achieve that, there are several areas of research that are being studied and applied, such as storytelling, user modeling, artificial intelligence, augmented reality, emotive agents, among other. This paper presents a component of the user modeling research.

The scope of this study is the development of a questionnaire used to assess psychological variables relevant in the context of human-technology interaction, for further use in user-profiling tasks. The questionnaire was developed for the specific case of YDreams Showroom.

The role of the questionnaire would be to extract psychological variables to be used as a tool in user profiling. As such, we started by looking for relevant variables inspired by the theory of reasoned action [1], namely, variables that could predict behavioural intention, and therefore could be assessed (since we can't assess behavioural intention itself). The chosen theoretical guideline was Venkatesh's [7] model of perceived ease of use. This model was adapted from TAM - Technology Acceptance Model - and considers a perspective of anchoring and adjustment in the conception and change of perceptions of ease of use through time. The model of perceived ease of use argues that people will create their perceptions of the ease of use of a system based on different anchors that relate to their general beliefs about computers and their use. For that purpose, they propose the existence

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

of the following anchors: computer self-efficacy (CSE), computer anxiety (CA), computer playfulness (CP) and perception of external control [5]. Venkatesh suggests that anchors guide the initial judgment but experience with the system will lead to an adjustment, on the basis of the usability of the system, and on perceived enjoyment [5]. Regardless of this, the role of computer self-efficacy and perceived external control remains pretty strong, even after experience with the system [5].

Trying to adapt the Technology Acceptance Model (TAM) [6,7] from the technology implementation in organizational settings framework to a tour in a showroom of novel technological applications (including Augmented Reality), we selected the variables that would fit the best into this new context and in future user-profiling applications: computer self-efficacy (CSE), computer anxiety (CA) and computer playfulness (CP). Due to our intention of integrating an AR virtual agent/avatar as a tour guide in the Showroom visits, we also added an "Interaction with an Avatar" (IA) scale to the questionnaire. To do so, we adapted two scales of the Nomura, Kanda, Suzuki & Kato's Human-Robot Interaction questionnaire [3], due to the similarities between robots and AR-based avatars: there is a general reluctance in people's willingness to interact with AR avatars and robots alike, as both are perceived as novel communicational entities [3]. From Nomura and colleagues' scales (NARS - Negative Attitudes toward Robots Scale and RAS - Robot anxiety Scale) we selected, adapted and translated the items that seemed to fit better in our approach, ending up with our own "interaction with avatars" (IA) scale.

METHOD

The questionnaire was built using a 6 degree Likert scale to assess the subjects rating of all items of the CA, CSE and IA variables. The CA scale had 5 items, the CSE scale consisted on 10 items and the IA had 12 items. The CSE and the IA scales had substantially more items due to the novelty of their use and adaptation to the language and the context of the study. This section was available online at "http://spreadsheets.google.com/viewform?formkey=dGt6a E5ma0c3YVc0ZTVaQ1NSR242Tnc6MA". The CP scale consisted on a 7 items approach and was addressed online as well but as you argue ahead this as deleted from the original questionnaire so is no longer available. The scale was simpler (yes/no) and context driven so subjects didn't have to change from one assess criteria to another thus creating some unnecessary noise in the data. This CP scale was translated and adapted from the original from Venkatesh [5] "The following questions ask you how you would characterize yourself when you use computers: spontaneous, unimaginative, flexible, creative, playful, unoriginal, uninventive...".

We addressed the online questionnaire via email to 62 persons, 17 of them working in YDreams (which we considered to be "high technological proficiency" users, as

opposed to the other 45 subjects, which we considered to be "normal technological proficiency").

With these responses, we analysed the questionnaire to figure if all the scales were working as separate and consistent components and therefore were a good measure of what they proposed to measure.

RESULTS

We started by performing a factor analysis in order to ascertain if the questionnaire items saturated in the corresponding dimension/variable and to what degree they did so. The factor analysis showed that items from computer self-efficacy, computer anxiety and interaction of avatars loaded in separate components. In contrast, computer playfulness items' didn't show such a consistency in its saturation within the factors. In addition, we calculated the Cronbach alphas of the original scales and found that all shown high values, except for the Cronbach alpha for computer playfulness, that was significantly lower than that of the other scales (see Table 1). Because of this, this scale was deleted from the questionnaire. This variable seemed very promising, but really didn't work in our study. In future studies it would be interesting to try to tap this variable in a different way.

| | CSE | CA | СР | IA |
|-------------------|-----|-----|-----|-----|
| Cronbach's alphas | .89 | .85 | .72 | .91 |

Table 1. Cronbach's alphas results of all scales.

We also performed a 4 factor analysis (CA, CP, CSE and **IA**) that allowed us to collect the most relevant items for each factor and afterwards we correlated the items of the check them scales to cross between groups (normal/Ydreamers). The results allowed us in general to choose 3 or 4 items per scale (CA, CSE and IA) after the deleting of the CP from the analysis. On a trial 5 factor analysis we were presented with the possibility of a fraction in the IA scale that suggested evidence that confronting and having to interact "physically" with an avatar and having to communicate with it are 2 different approaches with somehow separated psychological representations. Due to lack of time we weren't able to research on this finding any further.

In summary, computer self-efficacy, computer anxiety and interaction of avatars became clear different components in the factor analysis and showed high Cronbach's alphas, proving to be good measures of what they propose to measure.

CONCLUSIONS AND FUTURE WORK

In this study we managed to find robust measures of psychological variables that are valuable in the broad context of the human interaction with novel interactive installations mediated by avatar-like virtual guides. These measures will help predicting ease of use and satisfaction with technological applications and are prone to be transferred to more indirect (and also simple and elegant) ways of data gathering.

Although it is well-known that psychological variables have been extensively used in affective computing [2] or even in user profiling itself [for an example with personality variables see 4], the use of specific cognitive and attitudinal variables directly involved in user's satisfaction and perceived ease of use of specific applications (in our case, YDreams' showroom interactive installations) is an important innovation as a tool to enhance user experience.

With a validated version of the questionnaire, our goal would be to find and validate tasks that could assess these variables through an indirect manner. In that sense, it would be required to think of applications or aspects of them that could somehow tap these psychological variables. Regarding computer self-efficacy, we thought that giving subjects the possibility of watching a demo before a task could somehow relate to this variable. In this sense, a person with high self-efficacy would have a lower probability of choosing to watch the demo, while a person with low self-efficacy would have a higher probability of choosing to watch the demo.

Concerning the interaction with avatars (IA) variable, there is a possibility that it could be tapped by trying to get the user to respond to an instruction of the avatar. By measuring the time latency between the instruction and the user's action it might be possible to get information about the user's ability to understand the avatar as a member of an interaction, which is capable of communication and understanding it's context.

Regarding computer anxiety, it might be hard to tap this variable with some sort of physiological measure without being too intrusive. Nevertheless, hesitation could be an indicator of it, in the sense that some specific time latencies might be a good clue to it. In addition, overall performance could also work as a clue to this variable, in the sense that if a subject feels more anxious that might compromise he's performance. In fact, overall performance could be thought of as a combination of the influence of all variables (and others not considered, of course). In this sense, a high overall performance could be an indicator of: high selfefficacy, low anxiety, and good interaction with avatars. With this next step, it would be possible to minimize user's physical and conscious input of information. To do this, the theoretical rationale of the variables should always be followed and the technological applications should be scrutinized so that characteristics that allow the tapping of these psychological constructs (opportunities to measure reaction times, difficulty level selection, need for demonstration/help videos, etc.) can be exploited.

As future work, this kind of user information could be used to feed algorithms for adaptive learning environments, providing the user with a personalized response from the environment and virtual agents without being intrusive and without the need for physiological measures.

REFERENCES

- Ajzen, I. From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), *Actioncontrol: From cognition to behavior*, (1985), Heidelberg, Germany, 11-39.
- Hudlicka, E. Affective Computing for Game Design. In Proc. GAMEON-NA, 2008, McGill University, (2008), 5-12.
- 3. Nomura, T., Kanda, T., Suzuki, T., & Kato, K. Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes toward robots. *Transactions on Robotics*, 24 (2008), 442-451.
- Nunes, M., Cerri, S. & Blanc, N. Towards User Psychological Profile. In Proc. VIII Simpósio sobre Fatores Humanos em Sistemas Computacionais IHC (2008), 196-203.
- 5. Venkatesh, V. Determinants of Perceived Ease of Use: Integrating Perceived Behavioral Control, Computer Anxiety and Enjoyment into the Technology Acceptance Model. *Information Systems Research*, 11 (2000), 342-365.
- Venkatesh, V., & Bala, H. Technology Acceptance Model 3 and a Research Agenda on Interventions., *Decision Sciences*, 39 (2008), 273-315.
- 7. Venkatesh, V. & Davis, F.D. "A model of the antecedents of perceived ease of use: Development and test". *Decision Sciences*, 27 (1996), 451-481.

MAESTRO: Orchestrating User Behavior Driven and Context Triggered Experience Sampling

Alexander Meschtscherjakov alexander.meschtscherjakov@sbg.ac.at Wolfgang Reitberger wolfgang.reitberger@sbg.ac.at Manfred Tscheligi manfred.tscheligi@sbg.ac.at

Christian Doppler Laboratory for "Contextual Interfaces", HCI & Usability Unit, ICT&S Center, University of Salzburg, Salzburg, Austria

ABSTRACT

In this paper we present "MAESTRO" a novel concept to study behavior in realistic environments based on the experience sampling method (ESM). It constitutes a variation of ESM by giving researchers the possibility to dynamically and remotely "orchestrate" experiencesampling studies for evaluating usage behavior in various different contexts. MAESTRO allows user behavior driven and context triggered experience sampling by using a client-server architecture. So far a variation of different computerized ESM tools have been introduced capable of delivering questionnaires triggered by contextual information on mobile devices. We propose to enhance ESM by not only using context and user actions as a trigger for ESM questions, but also by exploiting long term user behavior and usage patterns for shaping personalized ESM questions to different types of users. Our concept lays the foundation of measuring user experience not only in a mobile but in a magnitude of different contexts. As a proof of concept we present an implementation of our approach on BlackBerry devices along with initial results of a case study validating this approach.

Author Keywords

Experience sampling method, in-situ evaluation, user experience measurement, ESM tool, methodology.

ACM Classification Keywords

H.5.2 User Interfaces: Evaluation/methodology

INTRODUCTION

The evolution of computing from single user desktop computer located in offices to mobile and ubiquitous devices that permeate the entire environment has sparked the need for in-situ evaluation methods. Theoretical works like Hutchins' "Cognition In the wild" [3] have encouraged researchers to leave the lab and conduct their studies in the field. Therefore psychological self-reporting methods like the Experience Sampling Method (ESM) [5] have been adopted to a wide range of research fields. ESM is a psychological method to capture participants' experiences in everyday life. This is achieved by triggering self-reports (e.g. diary entries, predefined questions on paper notes) from the participants.

Recently there has been a shift from paper-and-pencil ESM to more sophisticated computerized tools. Le et al. [6] have explored and compared various early ESM tools (ESP, iESP and PMAT) for the behavior research method community. Since ESM is a method for collecting information not only about the content of the daily life of individuals, but also tries to capture different context factors, the requirements for these tools are twofold: on the one hand researchers have to capture input from the users' subjective experience and on the other hand they need to record objective information about the context. Furthermore context information can also be used to trigger dynamic samples. This possibility was first identified by Intille [4] and demonstrated in the Context-Aware Experience Sampling toolkit (CAES). This approach can be referred to as context-contingent ESM. Our approach is to enhance ESM by not only using contextual information or user actions as a trigger for ESM questions, but also by exploiting users' usage behavior over a period of time for shaping personalized ESM questions to different types of users. This means that user behavior driven and context triggered ESM enables the researcher:

- to trigger ESM questions based on the recorded user behavior in the past and dynamically adapt these ESM questions whenever the user changes his / her behavior.
- to log and monitor in real-time comprehensively and flexible user behavior together with meaningful context information.

RELATED WORK

Experience sampling has been applied to various studies of mobile user behavior. Mehl et al. [7] invented the Electronically Activated Recorder (EAR), which automatically recorded audio samples thorough the day in order to reflect people's daily lives. Another recent project

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

is TUMCAT [11] a test bed for user experience for mobile context-aware applications. Beside the previously mentioned CAES [4] different experience sampling tools have been presented. SocioXensor [10] as an example is a toolkit, which exploits hardware sensors and software capabilities located on the mobile device to enable contextcontingent ESM. One of the most prominent ESM tools so far is MyExperience [2], which is developed for Windows Mobile devices and requires installation of a Microsoft SQL database. On both tools the logging information and the experience sampling data are deployed locally on the mobile device itself. Most of existing tools have been implemented as standalone tools on Windows CE or Palm OS platforms.

Our approach follows a different paradigm. We propose a shift from the client (the mobile device) to the server. We therefore only need a lightweight client software and no database installed on the mobile devices. This allows developing relatively easily lightweight clients on a multitude of mobile platforms in various contexts. A similar solution is Momento proposed by Carter et al [1]. Contrary to our approach Momento does not support ESM questions triggered by an analysis of previous user behavior. Reimers and Steward [9] used SMS text messaging as a simple means for responding in ESM studies. This approach has the advantage that no client at all is needed on the other hand neither user behavior or context triggered ESM is possible.

USER BEHAVIOR DRIVEN AND CONTEXT TRIGGERED EXPERIENCE SAMPLING

The MAESTRO Concept

The above-mentioned approaches have successfully shown that it is possible to trigger ESM questions on certain contextual parameters. The contextual trigger has two purposes: (a) it determines the point in time, when an ESM question is asked and (b) it chooses an adequate context dependant ESM question. (a) gives researchers the possibility to react to a user interaction immediately and therefore minimizes the recall bias. (b) provides researchers with the possibility to ask context and usage behavior specific questions.

The MAESTRO approach includes both and adds another dimension. We want to be able not only to react to singular specific context information or an actual usage behavior but to respond dynamically to different user behavior patterns. To learn these patterns it is necessary to log the usage behavior over a period of time, and then dynamically assign appropriate ESM questions contingent on a long-term user behavior. It is therefore possible to provide different questions to different user groups, without knowing - a priori - which participant belongs to which group (e.g. frequent SMS users vs. infrequent SMS users). It is possible to dynamically adjust to a change in usage behavior (e.g. the participant used two write on average one Email per day and now he writes on average 25 Emails per day). It is also possible to predict the probability of a certain user behavior and then react with corresponding ESM questions. Another advantage of this approach is the possibility to conduct a multi-phase study, where the results of the first part of the study can impact the setup of the subsequent parts of the study (even dynamically). Using traditional tools would require collecting all mobile devices used for the ESM study in order to update the study setup.

The MAESTRO Architecture

The MAESTRO concept follows a client-server architecture to master user behavior driven and context triggered experience sampling. Logged events are not stored in a local database but are immediately sent to a web server, which is one of the key differences to other ESM tools (e.g. MyExperience [2]). Each of these events can serve as a trigger for the user behavior driven and context triggered ESM. Additionally, the researcher can define rules whether the specific event triggers an ESM question and - based on the user history - determine which questions are asked.

Generally the MAESTRO concept can be applied to various contexts. In the following we describe an example from the mobile context: MAESTRO has logged a certain user behavior over a period of time. E.g. the user has written on average three short messages (SMS) per week and is therefore categorized as an infrequent SMS user. One day the user behaves different than before and sends 35 short messages on one day. Whenever the user sends a short message a "SMS sent"-event is sent to the server. Based on predefined rules (e.g. change of SMS usage behavior) it is decided whether an ESM question is sent to the user. The appropriate question is then displayed on the client accompanied with an acoustic and tactile signal. The user is prompted to answer the question immediately. This has the advantage of minimizing recall bias. Since most events on the mobile device are user triggered, questions are likely to be asked in moments when the user is interacting with the device. This minimizes the burden for the user and can increase the answer rate. If the user answers the question, this answer is immediately sent to the server. Depending on the answer an additional question may be asked.

BLACKBERRY IMPLEMENTATION OF MAESTRO

For the purpose of validating the feasibility of our approach we implemented MAESTRO on BlackBerry mobile devices. The implementation consists of a light-weight client software and a web application including a database on a web server (see Figure 1). The client was built in Java 2 Micro Edition especially adapted for BlackBerry devices. It is designed as a system module, which starts automatically with the BlackBerry OS. The client's task is twofold: firstly it sends predefined events via GPRS / EDGE to the web server. This is done in the background and happens without explicit user interaction. If at the time of the event no connection is available (e.g. too weak signal strength), the data is stored internally and will be sent as soon as possible together with the next event. The second



Figure 1. The MAESTRO Architecture.

task is to direct the BlackBerry's internal web browser to a web site to display the questions defined by the researcher, which allows the user to answer them.

The MAESTRO web application is implemented with PHP and MySQL. To setup a study, the researcher has to define questions, answers and structure of the questionnaire, and a configuration file, which holds the triggers for questions. When the server receives an event from the mobile client it stores it automatically in the database (1). The server determines whether the event triggers a question and which question. The particular question ID is sent back to the client (2), where the BlackBerry internal browser is opened and directed to request a URL containing the question ID from the web application (3) and then displays the question to the user (4). The user's answer is send back to the server and stored in the database (5). Depending on the answer the next question is sent to the user.

CASE STUDY

We now present a case study to illustrate how MAESTRO can be used to measure the usage behavior and the corresponding experience of users, while interacting with their BlackBerry devices. The study included 20 participants (10 m, 10 f, age between 22 and 56 years). The study lasted for eight weeks. During this time period each participant was provided with a BlackBerry device. On each device the client software was installed. Using a BlackBerry Enterprise Server (BES) it is possible to install the software without having the participant physically come to the lab. This reduces the effort for the researcher and the participant alike. Setting up the client involves selecting events, which the researchers want to be logged during the study. This is achieved by editing a configuration file on the client, which can also be remotely updated during the study over the air. We balanced the number of ESM questions to avoid to be too burdensome for the participants. We decided to react to events three times a day - in the morning, in the afternoon, and in the evening. Once the user answered the first question, he was presented with five to



Figure 2. ESM questions as displayed on the BlackBerry.

six follow-up questions. Due to the fact that we only asked questions as a reaction to an event on the mobile device, we were able to present mobile device usage context specific questions. During the study we analyzed the logged data and the collected answers. This gave us the possibility to react to problems and adapt questions depending on the usage behavior.

From a user's perspective a typical experience sampling, where the context trigger was activated by sending a text message looked like the following: After the user sent the text message at 10:30 a.m. the MAESTRO client logged contextual information (e.g. time, user id) and sent it to the server. Based on different rules (e.g. user history) it was decided whether a sample was triggered or not. Initially all users had the same set of rules. The MAESTRO toolkit gave us the possibility to react to certain user behavior patterns in the past and ask different questions based on this behavior. This possibility clearly sets MAESTRO apart from any other existing tool. One rule was for example that only a single set of questions were asked between 9 and 12 a.m. Another rule concerned the last time a text message related question was already asked. Another rule was whether the user behavior has changed significantly or not. If all the rules applied the set of questions were displayed on the BlackBerry device. Sample questions are illustrated in Figure 2.

The results of this study are out of the scope of this paper and can be found in [8]. To give you an impression about the power of MAESTRO we report that we logged all in all 127.255 events (e.g. email sent, contact updated, keyboard unlocked etc.). The number of actual fully answered set of questions was 2.041. Altogether we had answers for 10.216 single questions, which means that we got in average nine answers per day per participant. This huge amount of data was analyzed by using prepared SQL-scripts.

DISCUSSION

Our results show that MAESTRO enables user behavior driven and context triggered ESM studies. We showed that it might not only log usage data but also measure subjective user experience during actual everyday mobile device usage. Both the possibility to react to a certain user behavior over time and the context triggering proved to be a very fruitful approach. Our user tailored ESM approach also reduced the burden on the user to answer too many questions since user behavior driven and context triggered ESM minimized the number of questions being asked. Individual interviews with the participants conducted after the study showed that the study with this setup as no burden. Some participants even said, that it would not have bothered them, if the study had continued. Only one participant stated the he was happy that the study ended. Based on our experiences in this case study, we now discuss advantages and limitations of MAESTRO in relation to other ESM approaches:

User behavior driven ESM: MAESTRO gave us the possibility to exploit long term usage behavior for shaping personalized ESM questions to different types of users.

Remote Installation: With the use of BES it is possible to install the MAESTRO client remotely onto BlackBerry mobile devices over the air.

Remote Setup Configuration: Once the client has been installed the experience sampler setup and the context logger setup can be configured via the server. This makes it easy to include a large number of participants and to change the study setup for follow-up studies.

Real Time Data Management: Since MAESTRO stores study data into a database on the server as it is generated the researcher has immediate access to the data. Thus the data can be backed up and analyzed in real time. This enables the researcher to assess whether the mobile device is still in use and if the MAESTRO software is running. Additionally the users' compliance with the study can be verified.

Limitations: The most severe limitation of MAESTRO is the need for a constant wireless network connection to be able to send logged data to the server and trigger questions. If no connection is available, the events are logged locally, but no questions can be triggered. The big amount of data traffic potentially leads to high connection fee, and may be particularly problematic when the user travels abroad. The constant network connection also reduces the battery life time cycle.

CONCLUSION AND FUTURE WORK

We have presented MAESTRO a user behavior driven and context triggered experience-sampling approach along with an implementation on BlackBerry devices. Overall our approach gives the researchers the possibility to dynamically "orchestrate" experience-sampling studies for evaluating usage behavior. At the moment MAESTRO focuses on the specific mobile device usage aspect of the entire context. MAESTRO lays the foundation of logging contextual information and measuring user experience in various different contexts. For the future we plan to get a holistic view of the situation and therefore be able to trigger questions even more precisely. For example within the car context it will be possible that the user is asked an ESM question dependant on different car relevant data (e.g. while waiting in front of a red traffic light). Another context in which we will deploy our concept is the factory. Dependent on operator behavior we aim at understanding reasons for certain operator decisions. Furthermore we will port MAESTRO to different platforms and release an open source version.

ACKNOWLEDGEMENTS

The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged (Christian Doppler Laboratory for "Contextual Interfaces").

REFERENCES

- 1. Carter, S., Mankoff, J., and Heer, J. Momento: support for situated ubicomp experimentation. In *Proc. CHI* 2007, ACM Press (2007), 125-134.
- Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., and Landay, J. A. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proc. MobiSys 07*, ACM Press (2007), 57-70.
- 3. Hutchins, E. *Cognition in the wild*. Cambridge, MA: MIT Press (1995).
- Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., and Bao, L. A context-aware experience sampling tool. In *CHI '03 Ext. Abstracts*, ACM Press (2003), 972-973.
- 5. Larson, R., Csikszentmihalyi, M. The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, *15* (1983), 41-56.
- 6. Le, B., Choi, H. N., Beal, D. J. Pocket-sized psychology studies: Exploring daily diary software for Palm Pilots. *Behavior Research Methods*, *38* (2006), 325-332.
- Mehl, M. R., Pennebaker, J. W., Crow, M. D., Dabbs, J., Price, J. H. The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods*, 33 (2001), 517-523.
- Meschtscherjakov, A., Obrist, M., Mirlacher, T., Lankes, M., and Tscheligi, M. Evaluating the emotional attachment to mobile devices & services. Tech. rep., HCI & Usability Unit, ICT&S Center, University of Salzburg, 2008.
- Reimers, S., Stewart, N. Using SMS text messaging for teaching and data collection in the behavioral sciences, vol.41 (2009), 675-681.
- ter Hofte, G. H. Xensible interruptions from your mobile phone. In *Proc. MobileHCI* '07, vol. 309. ACM Press (2007), 178-181.
- 11. Vermeeren, A., and Kort, J. Developing a testbed for automated user experience measurement of context aware mobile applications. In: *Proceedings of the COST294-MAUSE Workshop NordiCHI '06*, (2006).

Multi-Method Analysis of the Relationship Between Individual's Space-Time Behavior, Built Environment and the Evolution of Cognitive Representations: Application of Tracking Technologies and Internet Surveys

Anastasia Moiseeva

Faculty of Architecture, Building and Planning Eindhoven University of Technology PO Box 513, Vertigo 8.16 5600 MB Eindhoven, the Netherlands A.Moiseeva@tue.nl Tel: +31 40 247 3044

Harry Timmermans

Faculty of Architecture, Building and Planning Eindhoven University of Technology PO Box 513, Vertigo 8.16 5600 MB Eindhoven, the Netherlands H.J.P.Timmermans@tue.nl Tel: +31 40 247 5096

ABSTRACT

This paper summarizes some key concepts and data collection methods used in a study which aims at better understanding how cognitions of urban environments evolve over time as individual learn about the environment by executing activity patterns. A GPS device is used to trace respondents over a longer period of time. These traces are semi-automatically interpreted to derive data on activity-travel patterns. A Web-based survey serves to complement these behavioral data with data on cognitions.

Author Keywords

Cognitive process, learning component, GPS tracking, space-time behavior.

INTRODUCTION

Research in urban planning and design, environmental psychology and related disciplines suggests that often the relationship between characteristics of the built environment and comfort, satisfaction, behaviour etc are less strong than one would like to believe [6]. It can be argued that the built environment is the spatial manifestation of human decision making. Many of these decisions are related to the way in which we perceive space, evaluate the elements of space, and imagine its potential use [1, 4]. It is important to realize that all experiences with elements of environments external to the individual take place within a framework of space and time. Although there is a wealth of studies on travel and spatial behaviour in

general, much less studies have focused on space-time behaviour in public and semi-public spaces. There is a literature in environmental psychology, but this tends to focus on psychological concepts; hence the specific timespace relationship is typically not the core of analysis.

A fundamental assumption in this study is that the cognitive processes (such as perceiving, learning, formation of cognition, evaluation and cognitive maps) operate to produce in people an individual spatiotemporal awareness about environments, and mediate the relation between urban environment and behaviour. The approach taken in this study is to combine qualitative and quantitative analyses for systematically analyzing space-time behaviour and underlying psychological factors such as development of personal cognitions and value judgments (overall evaluation) toward a number of public and semi-public spaces.

We are trying to capture the dynamic evolution of the cognitive process (learning component) because learning has a major influence on the type, complexity and intensity of cognitive representations.

METHOD

GPS Tracking and Automatic Imputation of the Behavioral Patterns

In the context of our study it is crucial to understand the level of interaction (behavioral component or experience) one has had with all of the choice locations in order to determine associated cognitions and value judgments, their intensity and dynamic development. Tracking technology such as GPS logger Bluetooth A+ is utilized to collect data during 3 months on choice locations and activity-travel patterns.

The authors have developed a system, called TraceAnnotator and a prompted recall instrument, which

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

was designed for semi-automatic imputation of activitytravel patters [5]. The system has been developed to process automatically multi-week GPS traces. For the purpose of this project, the configuration of TraceAnnotator has been designed in a way that two main processes are handled: (*i*) imputation of transportation modes and activity episodes, whereas a Bayesian belief network is a core of the imputation process; (*ii*) imputation of an activity type where GPS data are fused with personalised and general land use data. The core of the TraceAnnotator is a Bayesian belief network that uses a network of input variables, such as speed, acceleration, possession of a transportation mode and etc, to classify the outcome variables of interest (activity location and trips with associated transportation modes).

A total of 7 different types of transportation modes are considered: walking, running, bike, motorbike, car, bus and train. For public transport, we consider only transportation modes, which are available in the study region (bus and train). Two types of the activity stops are defined: when a person conducts an activity at a certain location (activity episode) and when a person conducts an activity inside a building (activity inside building).

To increase imputation accuracy of an activity location and type of behaviors at the location, additional variables such as time of the day, and day of the week, are added as input variables into the Bayesian belief network. The results of a pilot study have indicated that activity-travel patterns can be derived from GPS traces with a high degree of certainty.

For the validation process of modes and activities episodes generated by TraceAnnotator the Internet-based prompted recall instrument was designed. A specially designed web application allows survey participants to upload multi-days GPS traces (<u>http://bw-dssv02.bwk.tue.nl/</u>). After uploading multi-day GPS traces, TraceAnnotator processes the data. As a result, Activity Agendas, arranged by the date (yymm-dd), are generated for every day. Respondents are requested to check the information for every Agenda, make necessary changes, save these changes and confirm the Agenda.

It is essential for behavioral research that we are able to impute from the GPS tracers the locations of origins and destinations, trip length, trip duration, departure and arrival times and modes for all individual trips and their corresponding trip purposes. As a result space-time behavioral patterns are described in terms of duration, frequency, type of the behavior, sequence and temporal dynamics in the use of place.

Internet Survey

The physical attributes of the built environment as well as the cultural and social components (constructs) and aesthetic quality are fundamental in forming individual's cognitive representations and value judgments toward public locations. An internet survey was designed to measure personal awareness, perceptions, cognitions and overall evaluation of certain public locations.

To allow analyses how cognitions change as individuals learn the city, respondents were asked to complete the survey in the beginning, in the middle and in the end of the tracking study.

The first group of questions involves questions concerning cognitive processes. Several questions are related to individual awareness about location attributes, perception of the location attributes and symbolic meaning of the location. Other questions are intended to understand personal considerations (reasoning) why the location was chosen.

The second group of questions is targeting toward the evaluative component: personal satisfaction with the location in general and evaluation of certain behavior in general and at particular locations.

The third group of questions consists of questions related to the learning component such as frequency of visiting and degree of interaction.

The web survey gives an understanding how personal evaluations and preferences toward location differ among people and between locations. Because the measurements are repeated, changes over time as people learn about their environment can be noted.

METHODS COMBINATIONS

The results of web survey combined with space-time behavioral patterns of individuals give an understanding how the level of experience and familiarity influences the formation of personal cognitions during the time hence allow to better understand and predict behavioral movement and space-time behavior.

The internet survey will provide information concerning the accuracy and range of recalling attributes (characteristic) of the location and evaluation of these attributes by participants (importance of these attributes for participants). For instance, on the question – could you identify social, physical and aesthetic attributes that strongly correspond to the image of this location, evoke the memory about this location – participants have to mark how these attributes characteristic for a specific location. The format for a five-point Likert scale is used:

- 1. very characteristic for this location
- 2. somewhat characteristic for this location
- 3. not characteristic for this location
- 4. I do not know I do not notice
- 5. no opinion

The physical attributes relate to the built environment, the built environment includes land-use patterns, large- and small-scale built and natural features (e.g. architectural details, quality of landscaping) and the transportation system (the facilities and services that link one location to another). Cultural and social characteristics relate to the social and cultural atmosphere of the location, diversity of activities, risk to meat unpleasant encounters, safety of walking around and etc. Aesthetic quality ascribes artistic value of the location, beauty of place and views on architecture.

At the same time the GPS traces will provide very detailed information how much time has been spend at these destinations (location), the transport mode used and the route followed.

Because individuals are traced for a longer period of time, these data also allow detecting temporal variation in activity travel patterns. Which destinations are visited repeatedly, which ones have been visited for the very first time. Assuming that cognitive representation evolves as a function of the execution of activity-travel patterns, the awareness of places, the accuracy of recalling attributes values and the degree of detail or lack therefore will likely change over time and these changes if any will be reflected in the web-survey data.

In addition to input to statistical analysis, a model of how cognitive representations are formed over time can be developed. Cognitive representations reflect the awareness of people of choice alternatives in their environments. This model can therefore be linked to a model of activity-travel behavior. Assuming that such behavior is contextdependent, choice heuristics can be derived from the data on activity-travel behavior, using data mining algorithms. An example is Albatross, developed for the Ministry of Transport [2]. This model is however static. The rules embedded in this system may however be made dynamic by allowing for specifying and simplification of rules.

In principle, alternative tree induction algorithms may be used for this purpose. However, we have found the best results for Chaid-based methods. This tree induction method allows one to identify the rules that describe which choices are made under which conditions. It relies on the *Chi-square* test to determine the best next split at each step, and generates a decision tree by splitting subsets of the space into two or more nodes repeatedly, beginning with the entire data set. To find the best split at any node, it evaluates each predictor variable and merges any allowable pair of categories of that predictor variable if there is no statistically significant difference within the pair with respect to the target variable. The process is repeated for each newly created group until no more significant splits are found. This process of extracting the rules is the same as the one used in the original ALBATROSS model.

Arentze and Timmermans [3] have suggested a method to derive elasticity information from the rule-based models to facilitate interpretation.

The principle is to use the model to predict for each condition variable a frequency crosstable of the levels of the condition variables versus the levels of the target variable in columns. The frequency table for a given condition variable is generated by applying the model as many times as there are levels of the condition variable. The frequency distribution across actions of the action variable predicted under that setting is recorded. Repeating this process for each level of the condition variable yields a frequency cross table of the condition variable against the action variable. The impact of the condition variable is then measured as the Chi-square for this frequency table. Apart from impact size, they also suggested a measure of the direction of impact.

Creating dynamic rules, which would represent a shift in the scripts people use to use the city, can be done by consistently checking whether a specification or detailing of the condition states would improve the descriptive accuracy of the choice rules. Similarly, simplification can be handling by testing whether a merge of conditions states would not lead to any statistically significant reduction accuracy.

REFERENCES

- Amadeo, D., Golledge, R. An Introduction to Scientific Reasoning in Geography. Melbourne, FL: Krieger Publishing, 1975/1986.
- 2. Arentze, T., Timmermans, H. A Learning-Based Transportation Oriented Simulation System. In. *Transportation Research Part B*, 38 (2004), 613-633.
- Arentze, T., Timmermans, H. Measuring Impacts of Condition Variables in Rule-Based Models of Space-Time Choice Behavior: Method and Empirical Illustration, In *Geographical Analysis*, 35 (2003), 24-45.
- 4. Golledge, R. *Spatial Behavior: A Geographic Perspective*. New York, The Guilford Press, 1997.
- 5. Moiseeva, A., Jessuran, J and Timmermans, H.J.P. Semi-Automatic Imputation of Activity-Travel Diaries Using GPS Traces, Prompted Recall and Context-Sensitive Learning Algorithms. *In Transportation Research Record: Journal of the Transportation Research Board*, 2010 (to be appeared).
- 6. Stead, D., Marshall, S. The relationship between Urban Form and Travel Patterns. An international Review and Evaluation. *European Journal of Transport and Infrastructure Research 1*, 2 (2001), 113-141.

Appraising the Aesthetics of Human Movement: An Application to Contemporary Dance Using a Motion Capture System

Marta Castañer

INEFC-Lleida, University of Lleida, Partida Caparrella s/n. 25192, Lleida, Spain mcastaner@inefc.es **Carlota Torrents** INEFC-Lleida, University of Lleida, Partida Caparrella s/n. 25192, Lleida, Spain ctorrents@inefc.es Gaspar Morey

INESCOP, C/ Selleters, 25 – Polígono C'an Matzarí, 07300 Inca, Palma de Mallorca, Spain gmorey@inescop.es

Toni Jofre

INESCOP, C/ Selleters, 25 – Polígono C'an Matzarí, 07300 Inca, Palma de Mallorca, Spain tjofre@inescop.es

ABSTRACT

The present study describes a way of appraising the aesthetic meaning of dance motor skills generated by professional modern dancers. Using a motion capture system (Vicon MX), both real images and 3D stick figures were simultaneously obtained from 96 trials of dance motor skills performed by four experienced contemporary dancers. Subsequently, 101 students of Physical Activity and Sport Sciences used semantic differentials to appraise the aesthetic value of each type of motor dance skill, comparing the virtual and real images of dancers.

M. Teresa Anguera

Faculty of Psychology, University of Barcelona,

Campus Mundet, Vall d'Hebrón. Barcelona, Spain

tanguera@ub.edu

Author Keywords

Aesthetic movement, aesthetic perception, motion capture, contemporary dance.

INTRODUCTION

Every choreographic production is susceptible not only to being observed from the theatre stalls as something ephemeral, in the successive appearance and disappearance of bodies, but can also be observed in a scientific and objective way as a genuine block of text that offers an exhaustive and lasting description of the text written by bodies in movement.

Davies [3] draws upon Laban's legacy of movement analysis or Labanotation, stating that space is related to the extent of the 'bubble' in which we move, referred to by Laban as the kinesphere. Dance is dynamic in both space and time, and is always subject to aesthetic appraisals based on stimuli, as well as to subjective perception (see Depatri, 2009; Figure 1). The present study focuses on motion capture so as to compare it with real images, the aim being to determine whether the dancer's real body and image influences the aesthetic appraisal of dance.

The system (Vicon MX) we used makes use of retroreflective markers placed on defined points of the dancer's anatomy (PlugInGait marker set), thus enabling his/her dance motor skills to be reconstructed in three dimensions and to be observed from any perspective (Figures 1 and 2).

By generating a faithful virtual reproduction of motion, this approach enables us to identify which kinesic dance styles are most highly appraised. Specifically, we compared stick and real images from a total of 96 trials related to the essential dance motor skills of displacements, turns, jumps



Figure 1. Lines of dance movements (Depatri et al., 2009).



Figure 2. Retro-reflective markers on a dancer's anatomy, and real and 3D stick images from motion capture.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 3. The eight essential dance motor skills analysed.

and balance (Figure 3), which were chosen according to the criteria of Gallahue & Cleland-Donnelly [8] and Castañer et al. [1].

Aims

In order to focus on the aesthetic appreciation of human dance we observed and judged eight essential motor dance skills involving displacements, turns, jumps and balance. Although a dancer's image, as well as his/her expression, can modify the perception of motion, the influence of these features is unknown. However, these features disappear when observing only the animation of a stick figure obtained via 3D motion capture.

The aim of the study was therefore to establish the influence of non-kinematic parameters on aesthetic judgement in dance, this being achieved by comparing virtual and real images of dancing bodies in order to determine whether the dancer's real body and image influences the aesthetic appraisal of dance.

METHODS

Four experienced contemporary dancers performed three repetitions of eight dance motor skills in a space measuring $(3 \times 4 \times 2.5)$ m. Retro-reflective markers were attached to defined locations of each dancer's anatomy, as established by the PlugInGait marker set (VICON). This enables us to obtain stick-figure animations of each danced skill.

The dancers performed eight motor skills associated with contemporary dance. Each skill was repeated in three different kinds of movement, according to Laban's criterion of muscular tension (strong, mild and moderate). A total of 96 trials were simultaneously filmed with a traditional video camera and ten 3D cameras.

A total of 101 students of Physical Activity and Sport Sciences observed all 96 animations using the Polygon Viewer software (Figure 4), this being done over two sessions spaced one week apart. The students used semantic differentials [7] (Figure 5) to appraise levels of aesthetic beauty in the observed movement. One semantic



Figure 4. Screen capture of Polygon Viewer software.

differential was completed by observing the virtual images of dancers obtained via the motion capture system. A week later a second semantic differential was completed by visualising the real filmed images of each dancer.

| Beauty | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Ugliness |
|--------|---|---|---|---|---|---|---|----------|
|--------|---|---|---|---|---|---|---|----------|

Figure 5. Polarity of aesthetic appraisal used as a semantic differential.

Instrument

The following instruments were used:

- Vicon Mx motion capture system with ten cameras operating at 125 Hz.
- Vicon Nexus 1.4 (VICON) capture software. Analysis software: Ad hoc routines developed within MATLAB (The Mathworks Inc).
- Traditional filming using a video camera.
- Semantic differentials [7] applied *ad hoc* to the aesthetics of human movement [2], in which opposing aspects related to the beauty of aesthetic appreciation were appraised.

DATA ANALYSIS

Correlations between the ratings of stick figure and video representations showed an r of about 0.5 (Table 1), the exceptions being the two turns and jump 2 (a jump performed while turning). By way of an example, Figure 6 shows the histogram of ratings corresponding to stick figures and videos, considering all skills together.

| | R | p | Mean score 3D | Mean score real |
|----------------|--------|--------|------------------|--------------------|
| All skills | 0,4783 | 0,0001 | 3,7 | 3,6 |
| Displacement 1 | 0,4855 | 0,0001 | 2,8 | 2.9 |
| Displacement 2 | 0,4588 | 0,0001 | 4,0 | 3,7 |
| Balance 1 | 0,5362 | 0,0001 | 3,8 | 4,1 |
| Balance 2 | 0,5135 | 0,3690 | 3,2 | 3,3 |
| Turn 1 | 0,3321 | 0,0001 | 4,2 | 3,8 |
| Turn 2 | 0,2399 | 0,0001 | 3,9 | 3,4 |
| Jump 1 | 0,5309 | 0,0001 | 4,5 | 4,0 |
| Jump 2 | 0,3597 | 0,0070 | 3,6 | 3,5 |

 Table 1. Correlations between the ratings for stick figure and video representations.



Figure 6. Ratings corresponding to the stick figures and for the videos.

CONCLUSION

Analysing the two semantic differentials (one of the elided images and another of the real filmed images) allows us to determine the prior experience of observers when it comes to appraising the aesthetic value of the danced motor skill series.

The data analyses show that although ratings are correlated due to kinematic parameters, the influence of non-kinematic parameters is also present. The larger differences corresponding to turns may be due to the dynamics and expressivity of these skills. It is likely that in this case, expressivity has had a greater influence on subjective judgment.

Stick figures are awarded higher scores, probably because some subtle movements (position of the hands, tension in the arms, etc.) are hidden when using 3D images. Indeed, the fact that these subtle movements are not reflected in the stick images, as opposed to what occurs when observing the movement of real bodies, illustrates that although aesthetic experience probably depends on the subjective perception of the observer, it is also induced by particular stimuli or their properties.

REFERENCES

- 1. Castañer M., Torrents C., Anguera M.T., Dinušová M. and Jonsson G.M. Identifying and analyzing motor skill responses in body movement and dance. *Behavior Research Methods*. 41, 3 (2009), 857-867.
- 2. Castañer, M., Torrents, C. and Dinušová, M. Specific Motor Components of Contemporary Dance Productions. A Semantic Differential Tool. 21st *World Congress on Dance Research*, Athens, 2007.
- 3. Davis, E.: Beyond dance. Laban's Legacy of Movement Analysis. New York: Routledge, 2001.
- Daprati, E., Iosa, M., and Haggard, P. A. Dance to the Music of Time: Aesthetically-Relevant Changes in Body Posture in Performing Art. *PLoS ONE*, *4*, 3 (2009), e5023. doi:10.1371/journal.pone.
- Jacobsen, T., Schubotz, R.I., Höfel, L. and Cramon, D.Y. Brain correlates with aesthetic judgement of beauty. *Neuroimage*, 32, 1 (2006), 486-487.
- 6. Leder, H., Belke, B., Oeberst, A. and Augustin, D. A model of aesthetic appreciation and aesthetic judgements. British Journal of Psychology, *95* (2004), 489-508.
- Osgood, Ch. E., Suci, J. And Tannembaum, P.H. *The* measurement of meaning. *Illinois*. University of Illinois Press, 1957.
- 8. Gallahue, D., Cleland-Donnelly, F. *Development of physical education for all children*. Champaign, IL: Human Kinetics, 2003.

Using Motion Sensing to Study Human Computer Interaction in Hospital Settings

Katherine M. Sellen

kate.sellen@utoronto.ca

Mark Chignell

mark.chignell@utoronto.ca

University of Toronto 10 King's College Road, Toronto, Ontario, Canada, M5S 3G4

Jacob Pendergrast

University Health Network 700 University Avenue, 8th Floor South Room 8-18, Toronto, ON M5G 1Z5 jacob.pendergrast@uhn.on.ca Sharon E. Straus University of Toronto 10 King's College Road, Toronto, Ontario, Canada, M5S 3G4 sharon.straus@utoronto.ca

Jeannie L. Callum

Sunnybrook Health Sciences Centre, B211, 2075 Bayview Avenue, Toronto, ON, M4N 3M5 jeannie.callum@sunnybrook.ca

Allison Halliday

Sunnybrook Health Sciences Centre, B211, 2075 Bayview Avenue, Toronto, ON, M4N 3M5 allison.halliday@sunnybrook.ca

ABSTRACT

Tracking someone's behavior patterns, interactions with technology, social interactions, or a combination of these in a medical setting can pose significant challenges. This paper provides an example, from an evaluation of new technology in the hospital setting, of how using motion sensor technology and video capture to provide context, combined with software logs, can meet ethical and practical challenges, and improve the depth and breadth of health informatics evaluation techniques.

Author Keywords

Motion sensing, ethnography, medical informatics, user studies.

ACM Classification Keywords

H5.m. Miscellaneous, J. Computer Applications, J.3 Life and medical sciences: Medical information systems

INTRODUCTION

Many technology interventions to improve healthcare safety and efficiency are the subject of research, formal evaluations studies, and impact assessments. This is especially needed for new technologies that can be disruptive and have unintended impacts on workflow and communication [4]. For researchers studying the process of uptake and use of a new technology or researching a new protocol for technology use, observing the workplace and actions of healthcare professionals, and the actions of patients and their caregivers, is the norm. Often this involves ethnographic methods, observation, and shadowing or following participants. Researchers might be aiming to record someone's behavior patterns, interactions with technology, social interactions, or a combination of these.

The nature and structure of medical work, particularly in the hospital setting, can pose a significant barrier to carrying out such research. Often research requires:

- Tracking individuals and/or objects over a wide range of time and space such as across an entire hospital
- Recording events that are infrequent and happen at unpredictable times of the day or night
- Recording frequent discreet events such as using a particular drop down menu.

Relying on human observers to be present and record such events can be impractical, unrealistic, subjective, and may result in very little data that is quantifiable. Using video can provide a persistent record from which to sample, measure, quantify, and code, and can be accessed by multiple members of the research team. It can also be used by participants as a training aid or for process improvement as a product of the research process [7]. Video can provide a record from which measurements can be made, behaviors can be coded, or insights shared. It is a commonly used tool in other settings but presents significant problems in the healthcare environment. Some of the challenges are:

- Medico-legal environment
- Staff attitudes towards capture
- Managing consent

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

• Managing video data

Researchers who report on their research experiences often document challenges posed by the medico-legal environment of hospitals [3]. Staff may be resistant to having events captured in detail in case a medical error occurs and the captured event is used as evidence in court or for internal investigations.[3] The legal department may be unwilling for errors and near misses to be recorded in case, at a later date, an error is shown to have occurred more than once [3]. There are also the physical challenges of avoiding capturing data on patients, visitors and others in the hospital environment. This is especially important when using video or audio capture that is general and unspecific (field of view is not restricted) in semi-public spaces such as emergency rooms or recovery wards. The issues of how to inform participants, obtain consent, and manage communication with participants can become a barrier to research [8]. We aim to provide an example of how adding motion sensing capabilities to video capture in the hospital setting can overcome some medico-legal challenges and enhance possibilities for more focused, in-depth research.

BACKGROUND

There are an increasing number of examples of successful uses of video analyses in medical settings to analyze behavior, for education, training, quality management, and human factors and ergonomic analyses [3,7]. In this paper we describe the use of motion sensing to support video capture and analysis in an evaluation study of a new technology for blood delivery to operating rooms. The use of motion sensing together with video builds on work by other researchers to develop protocols for the successful use of video in sensitive hospital environments such as operating rooms, intensive care units, and emergency rooms.

STUDY SETTING AND DESCRIPTION

The use of "just-in-time" cross matching of blood for use in operating room settings is advocated as a patient safety practice and should provide greater process efficiency, decreasing the number of both unnecessary cross matches and transfusions [9]. The Neoteric BloodTrack® Self Serve System for remote allocation of blood is an example of a "just-in-time" system and has been installed in the OR suite of three Toronto hospitals. This device is networked to the blood bank laboratory information system and allows for point-of-care cross matching. The purpose of the research for which motion sensing was used is to evaluate the impact of the Neoteric Blood®Track Self Serve System on blood utilization, blood delivery, work practices and behavior of staff during blood issuing and delivery.

The Neoteric BloodTrack® Self Serve System consists of a database and a PC with a monitor at the blood bank/transfusion lab that is used to identify and track blood units and a remote (at the operating suite) fridge and kiosk. Once blood units have been indentified and tracked in the

blood bank/transfusion lab they are loaded into a 9 drawer locked fridge (called a HemoNine) that is situated in the operating suite close to the operating rooms. Alongside the HemoNine is a kiosk with a touch screen and label printer. When blood is needed in an OR an employee approaches the fridge and identifies herself by scanning an ID badge. The software then prompts the employee for information about the patient, looks up the blood type for that patient and unlocks the fridge. The employee can open only 1 drawer and pick out 1 unit of blood (each drawer is for a separate blood type). The employee scans the unit and the software checks the unit type with the patient type. If this information matches correctly, a patient label is printed for the blood unit. The employee then performs a second safety check scan. This process is repeated if another unit of blood is needed.

Research Challenge

The study required analysis of staff using the new system both in the blood bank and in the operating suite. Loading of the remote fridge in the operating area occurs anytime between 5:00 am and 8:00 am after blood bank staff have identified blood requirements. This task can take anywhere from 20 minutes to an hour and occurs approximately once every 5 days. Staff in the operating suite use the remote fridge to obtain blood for patients between 1-20 times per week depending upon the hospital. This task takes an average of 60 seconds, between 9:00 am and 4:00 pm Monday to Friday. Using traditional observational techniques this might require a researcher to be present in the operating suite area for several weeks in order to capture a 1 second interaction with the new system. Considering that the research study required comparison across sites it was calculated that a minimum of 120 days of observations would be required to capture a comparable number of errors between sites. An observational approach without capture of errors would not be useful.

One challenge of providing effective technologies for



Figure 1. Example of video camera position in the operating suite.

medical settings is whether the new technology can meet the requirements of critical events or events in crisis situations that may not occur very frequently and are unpredictable [2]. An example in the context of this study would be an unpredicted big bleed of a patient undergoing emergency surgery at 2 am. Therefore the researchers needed to be able to capture such critical events that might occur outside of the systems typical usage pattern.

Although hospital management was supportive of a study that captures and analyzes errors in using the new system, staff were resistant to video capture particularly of faces. It was also important to avoid capturing patient identifying data such as names and medical numbers as obtaining patient consent for use of this information would be impractical in an operating room setting. In order to overcome these challenges we decided upon using motion sensing to control video capture which also had some advantages for meeting ethical requirements and meeting staff concerns.

VIDEO AND MOTION SENSING SOLUTION

Small fixed cameras (4) were installed to capture interactions continuously over the capture period of 120 days. These cameras were placed to capture paper tools, blood unit scanning, and computer screen interactions on the desk that is used most frequently for issuing to the new fridge. Cameras (4) were also installed in the operating suite area where the HemoNine fridge is in use. These cameras were Lorex CCD color cameras with 3.6mm wide angle lenses. One dome camera was also used with built in audio capture. Cameras were intended to capture meta interaction data such as time taken to issue a unit of blood, placement of blood units on desks, interruptions to the retrieval process, user errors, and number and type of tools required for the issuing process. Care was taken in the installation of the cameras to make sure that faces were not in full view i.e. camera field of view included the new fridge, kiosk, desk, computer and desk. Figure 1 shows a typical setup for the cameras in the operating suite. Cameras were connected to a Lorex security digital recorder device (DVR) with 500 GB hard drives. Lorex DVR included software for motion sensing with a motion detection area of 16 x 12 grids per camera and 4 adjustable variables for motion detection sensitivity.

RESULTS AND DISCUSSION

In the setting where we implemented the motion sensor supported video capture we faced some specific challenges related to the dynamics of the tasks usually encountered in blood bank procedure and operating suite task flows.

Working with Dynamic Tasks and Timeframes

We identified 6 different types of task that would require study. Across these tasks, task frequency ranged from once every 45 minutes to once every 7 days, task durations ranged from 55 seconds to 45 minutes, and tasks did not follow an easy pattern of working hours but could occur at any time of day or night. This is not dissimilar to task dynamics encountered in other hospital settings such as intensive care units or emergency rooms. What these hospital settings have in common is that their tasks share attributes that make traditional task observation methodologies, either using human observation or human triggered video, impractical. Some of these attributes are:

- Unpredictable start times
- Round the clock activity
- Wide ranging task durations and frequency
- Active or moving tasks that occur across work spaces

Motion sensors allowed us to trigger video capture for events whenever they occurred. This allowed us to study all task types. It also allowed us to increase the timeframe for the video capture from 5-7 days of continuous observation to 120 days of discreet task observations. In this way we have been able to capture tasks that are representative of user behavior in the early days after implementation of the new technology, as well as tasks representative of user behavior after the technology has been in place for some time and use of the system has become routine. We were also able to capture critical incidents whenever they occurred e.g. use of the remote fridge for trauma patients who were brought in to surgery during the night, or, for transfusion requests during routine day surgeries when unexpected bleeding occurs.

Ethical and Legal Challenges

Motion sensing is particularly suited to achieve research protocols for video capture. Protocols can be developed to include practices that minimize medico-legal concerns and increase staff comfort with video recording. Some of these include; limiting view angles, restricting audio recording, facilitating optional deletion of video footage, restricting duration of videoing, and targeting specific tasks or interactions [5, 7]. In our study we used motion sensors targeted to very specific segments of each of the cameras' field of view. For example, we set up small areas of motion sensitivity for each camera field of view where staff would be interacting with the technology intervention under study such a scanning area and the door handle of the blood fridge. Combined with our multiple camera coverage we were able to trigger video capture for both very short durations and for only relevant interactions. The multiple camera approach enabled us to avoid, as much as possible, capturing staff head shots (a point of sensitivity with staff members).

Managing Consent

The motion sensing technology that we used also records time and date of the video trigger. With this information an efficient optional delete/non-consent procedure can be implemented. For this to work we asked non-consenting participants to anonymously fill a request for certain date and time segments to be deleted. This procedure satisfied both ethical review board concerns, increased staff comfort with the video, and was easily implemented.

Motion Sensing and Video Analysis

Motion sensors can trigger video capture over the full length of a pilot or research study. This allows more specific events to be collected for thorough analysis and interpretation including infrequent and critical incidents. Since a large number of specific events are captured it is then possible to use a sampling technique to select video clips for analysis. We have combined video clips with software log data that records user errors to identify clips for analysis.

LIMITATIONS AND FUTURE SOLUTIONS

The solution we have described was implemented using readily available technology most often used for security surveillance of commercial property such as stores and restaurants. There are other solutions that might provide more control to both the researcher and the participant. RFID tags could be used by consenting participants to trigger video capture in addition to motion sensor triggers [10]. Motion sensed event data (such a timestamp data) could be linked to software logs to provide automatic event/clip information and identification by task type, duration and participant characteristics.

CONCLUSION

The use of motion sensing to trigger video capture in hospital based studies has many advantages for the research and the participants. As well as the significant opportunities it provides for increasing the efficiency and focus of video capture (i.e. targeted to specific events/tasks), we have described the advantages to the researcher of using event trigger data to enhance analyses by providing larger more specific video clip libraries from which to sample for analyses. Whilst motion sensing has been used in other settings for many years, for health informatics research this study implementation represents a significant step towards addressing the challenges of behavioral research in the medical setting.

ACKNOWLEDGEMENTS

We acknowledge the help of the nursing staff and blood bank staff at each of our study sites.

REFERENCES

- 1. Ammenwerth, E., and de Keizer, N. An inventory of evaluation studies of information technology in health care trends in evaluation research 1982-2002. *Methods of Information in Medicine*, 44, 1 (2005), 44-56.
- 2. Bardram J.E. Designing for the dynamics of cooperative work activities. Proc. CSCW 1998, ACM Press (1998), 89-98.
- 3. Gelbart, B., Barfield, C., and Watkins, A. Ethical and legal considerations in video recording neonatal resuscitations. *Journal of Medical Ethics*, *35*, 2 (2009). 120-124.
- 4. Koppel, R., Metlay, J.P., Cohen, A.B. et al, Role of computerized physician order entry systems in facilitating medication error. *Journal of the American Medical Association*, 293, 10 (2005), 1197-1203.
- 5. Mackay, W.E. Ethics, lies and videotape.... *In Proc. CHI 1995*, ACM Press (1995), 138-145.
- 6. Mackenzie C.F. and Xiao Y. Video Analysis in Health Care. *Handbook of Human Factors and Ergonomics in Health Care and Patient Safety*. Lawrence Erlbaum Associates, Publishers, Mahwah, NJ. (2007), 663-678.
- 7. Mackenzie, C.F. and Xiao, Y. Video techniques and data compared with observation in emergency trauma care. *Quality & Safety In Health Care*, *12*, 2 (2003), II51-II57.
- 8. Mackenzie, C.F. and Xiao, Y. Video analysis for performance modeling in real environments: Methods and lessons learnt. *Proc. HFES 43RD Annual Meeting*, Human Factors and Ergonomics Society (1999), 237-241.
- 9. Staves, J. Electronic remote blood issue: a combination of remote blood issue with a system for end-to-end electronic control of transfusion to provide a "total solution" for a safe and timely hospital blood transfusion service. *Transfusion*, 48, 3 (2008), 415-424.
- 10. Taneva, S., and Law, E. In and out of the hospital: The hidden interface of high fidelity research via RFID. *Proc. INTERACT 2007.* ACM Press (2007), 624-627.

A High-Resolution System for Recording the Daily and Lifetime Behavioral and Movement Patterns of Individual Tephritid Fruit Flies

James R. Carey¹*

¹Department of Entomology One Shields Ave., University of California Davis, California 95616 USA jrcarey@ucdavis.edu

Leopoldo Robles

National Institute for Astrophysics, Optics and Electronics - INAOE Mexico

Jane-Ling Wang¹

janelwang @ucdavis.edu

Sige Zou

Laboratory of Experimental Gerontology, National Institute on Aging, Baltimore, Maryland 21224, USA zous@grc.nia.nih.gov

> Amy Morice¹ aemorice@ucdavis.edu

Pablo Liedo

El Colegio de al Frontera Sur Tapachula, Chiapas, Mexico pliedo@ecosur.mx

Hans G. Müller¹

hgmueller@ucdavis.edu

Donald Ingram Pennignton Biomedical Research Center, Louisiana State University Baton Rouge, Louisiana 70808 USA Donald.ingram@pbrc.edu Nikos Papadopoulos University of Thessaly, Greece nikopap@uth.gr

ABSTRACT

We developed a system in which the lifetime behavior and movement patterns of *Anastrepha ludens*—a Tephritid fruit fly slightly larger than a housefly and commonly known as the Mexican fruit fly—maintained in 27 cubic cm cages is recorded every fifth of a second for one minute three times per hour for 24 hours each day throughout their lives. The data generated from the behavioral monitoring system (BMS) include date, time, 1-of-6 behaviors, and locations in 3-dimensional space (XYZ location). This yields over 12 million data-points for a single fly that lives 100 days. The raw data enable researchers to analyze: (1) behavior including frequency, sequence, age and diel patterns, and circadian rhythm; (2) movement including age and diel patterns, rates, and pathways; and (3) location within the cage by age and time-of-day.

BEHAVIORAL MONITORING SYSTEM

The hardware of the BMS for monitoring behaviors of mexfly consists of a pair of cameras to acquire left and right video images in real time, a nine-cage tray to house individual flies in each cage, visible and infrared lights. The cameras are connected to an image processor for short term storage of video images, which are fed to a computer for behavior detection and long-term data storage. The system is configured as follows at the beginning of the experiments and then each time after fresh water and food are provided to flies. The cameras and lighting are adjusted to capture clear left and right images of each cage in the center of video. The fixed positions of water and food in the images are captured respectively and recorded into the system. The fly is recognized as a white image in the dark background of the cage. The rate of sampling by cameras is 5 frames or images per second. The sampling time per recording can be adjusted between the ranges of 10 to 60 seconds. The quantity of images or frames in the video will be a function of the sampling time per recording. Once the positions of water and food, sex of the flies, date of birth, comments, and sampling time are loaded, the recording can be initiated. The BMS records 3-D coordinates of individual flies from each video frame and automates classification of six behaviors, termed as resting, moving, walking, flying, feeding and drinking, for each bout or image frame, based on changes of 3-D coordinates between two constitutive frames and relative positions from the food and water sources. The BMS generates two ethograms in the text format every day, a behavior file containing date, time, XYZ coordinates, speed and behavior classification at each bout of imaging, and a statistics file containing date, time, total activity level, frequency and activity level of each behavior per sampling period.

^{*} Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

PROOF OF CONCEPT

Behavioral Experiments

Two BMS were used to monitor lifetime behaviors of 16 mexfly females with each BMS holding four females on the full diet and four females on sugar only diet. The flies were randomly assigned to and individually housed in a cage of the 9-cage tray in each BMS. Fresh food and water were provided to flies from the bottom of each cage once a week. The recording was initiated at the eclosion of the flies and stopped at day 134 when approximately half of the flies were dead. The video sampling time for each cage was 60-second long.

Lifetime Behavior Patterns and Dietary Effects

We used two BMS systems to continuously track eight mexfly females on the full diet and eight females on sugar only diet for 134 days when more than half of the mexflies were dead. The lifetime recording was disrupted for short periods of time when the system needed re-adjustment due to introduction of fresh food and water once a week or occasionally power outage. For each fly, we used the behavior ethograms from the BMS to compute the frequency and activity level of each of the six behaviors in every hour of the day in function of the age in days.

Using these hourly behavior patterns, we first computed frequency of each of the six behaviors over the age or over the time of the day to assess age-related or circadian rhythm patterns respectively. Several age-related patterns could be identified from the data depicted in the figures based on the outcome of recordings for 16 females on two different diets. Flies on both diets spend more than 60% of their time on resting every day over their lifetime. The next most frequent daily behaviors are walking and moving, which, combined, count approximately 25% of a fly's daily time. Flies on sugar diet have significantly higher walking frequency than those on the full diet. Flying, feeding and drinking occur at very low frequencies, which counts less than 5% of all the behaviors in most of the days. The frequency of daily resting and moving remain relatively constant with little fluctuation over the lifetime of flies, suggesting that there are no age-associated changes for these two behaviors. The frequency of daily walking reaches the peak around the age of 20 days and then gradually decreases with age for flies on both diets, suggesting that walking frequency is an age-related parameter. Overall, the frequency of daily flying declines with age after reaching the peak around the age of 20 days, while the frequency of daily feeding and drinking appears to increase with age.

We also analyzed the circadian rhythm patterns for the hourly frequency of each behavior. During night time between 7 pm when the light is off and 7 am when the light is on, flies spend almost all of their time (>95%) on resting, presumably in sleep, while other behaviors are rare. During daytime, resting is still the most frequent behavior, which occupies more than 55% of the flies' time; walking is the next most frequent behavior (up to 40%), followed by moving (up to 20%); feeding, Drinking and flying are rare (less than 5% combined). Walking and flying show hourdependent changes and reach the peak period in late afternoon from 3-7 pm, while the other behaviors display relatively constant rates during daytime. During daytime, flies on sugar diet walk more frequent that those on the full diet with the peak level differing by approximately 50%. However, the highest walking frequency occurs at 6 pm for both mexflies on sugar and the full diet.

Lifetime Activity Patterns

It is reasonable to assume that the daily total distance a fly travels reflects how active this fly is every day. To further assess the lifetime activity level of flies, we calculated the distance a fly travels in each hour every day. By averaging the hourly total distance values from individual flies, we generated age-related distance patterns and circadian distance patterns for flies on the full or sugar only diet. The daily total distance gradually increases in the first 15 days, reaches the peak at the age of approximately 20 days and then gradually declines with age for flies on both diets. This suggests that daily total distance is a good aging marker. The daily distance is generally higher for flies on sugar diet than those on the full diet with the peak distance levels differing by ~60%. The circadian distance patterns reveal that overall patterns are similar between flies on both diets with ~90% of the daily distance traveled is during daytime. Similar to the circadian walking frequency patterns, ~50% of the distance traveled is in late afternoon from 3-7 pm and the distance peaks at 6 pm for flies on both diets. In addition, the peak distance level for flies on sugar is ~40% higher than those on the full diet.

DISCUSSION

We described a behavior monitor system capable of automating lifetime behaviors. Each BMS tracks the 3-D positions of nine individually housed mexflies once every few minutes over their lifetime and generates ethograms with automated classification of six behaviors. This BMS can be easily adopted for a regular research lab as it is made with video cameras and a computer commonly found in a biology lab. The software converts video images of large file size into text files, which allows long-term continuous recording and storage by a standard lab computer. In addition, the software automates classification of six basic behaviors of mexflies, which makes it user-friendly for biologists to analyze data without much additional programming. Although many of the existing video systems have potentials to conduct lifetime long experiments, our BMS is the first high resolution machine vision system demonstrated to have the capacity to record lifetime 3-D positions of an animal and automatically classify lifelong behaviors for months.

Identification of Vertical and Horizontal Movement Patterns in Cod Behavior

Gudberg K. Jonsson

Human Behavior Laboratory, University of Iceland Skipholt 50, 105 Reykjavik, Iceland gjonsson@hi.is Vilhjalmur Thorsteinsson The Marine Research Institute Skulagotu 4, 121 Reykjavik, Iceland villi@hafro.is

Gunnar Gudni Tomasson

School of Science and Engineering, Reykjavik University Nautholsvik, Menntavegi 1, 101 Reykjavík, Iceland gunnargt@ru.is

ABSTRACT

The Atlantic cod is one of the most important commercial species known. 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. 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 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. A number of specific temporal patterns were also identified within and across vertical movements of individual cod. Future objective is to explore these patterns in relation to environmental factors and horizontal location.

Author Keywords

Cod, Gadus morhua, behaviour, DST tags, tidal wave model, T-patterns.

INTRODUCTION

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 current study discusses a new approach to analyze behavior such as horizontal and 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 (see www.patternvision.com) 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

The data was collected using the Data Storage Tags (DST centi series) developed by Star-Oddi (see www.star-oddi.com). The DST centi is a small underwater data logger, available with sensors for underwater temperature and depth logging. 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. All measurements are time related, utilizing a real time clock inside the DST.

Each event is defined by start time and end time, and also a label indicating the meaning of the event. For example, an event may indicate that the tag was located at a certain depth level, or that temperature increased over a certain time interval. The 3 types of events used in the current study are:

- Level-based events that are based on the range of the measurement data.
- Speed-based events, which indicate changes in the data.
- Acceleration-based events, which indicate changes in speed over time.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Event basing depth data involved partitioning timeseries into defined periods/events.

Tidal Wave Model

A numerical model has been developed and set up to predict sea level variations and tidal currents along the Icelandic coastline, taking into account both the astronomical and meteorological forcing. The model is run on an operational basis at the Icelandic Maritime Administration to predict sea level and tidal currents in Icelandic coastal waters using a weather forecast from the European Center for Medium Range Weather Forecasts. The model is based on the two-dimensional part of the Princeton Ocean Model (POM) [4], solving the nonlinear shallow water equations numerically using a staggered finite difference scheme.

RESULTS

Tidal Location of Fish in the Ocean Around Iceland

Knowledge of the whereabouts of tagged fish between tagging and recapture is important in research on commercial fish stocks. Indirect location of fish with DST



Figure 2. Example from an experimental application of this method tracing a cod tagged in Icelandic coastal waters using the database of tidal components generated by the tidal model.



Figure 3. An example of one pattern of vertical movements found in the time series. The pattern occurs 12 time during the observation period with same order and approximately the same time interval between event types.

tags measuring water pressure is possible by isolating the tidal variation of sea level (pressure) from other variations of pressure, such as those due to change of depth by the fish. These, together with the tidal model, may then be used to find the location within the ocean with matching tides to those measured by the fish. Tidal location of fish has been attempted in the North Sea with good results [1].

The existence of a tidal model and in particular the existence of a database for amplitude and phase of the most important tidal components for all the ocean around Iceland make it possible to attempt tidal location of tagged fish around Iceland. Figure 2 shows results from an experimental application of this method to fish tagged in Icelandic coastal waters using the database of tidal components generated by the tidal model described here. Preliminary results show that this method can indeed be used with the current database and that the accuracy obtained in location of the fish will be acceptable. However, improvements of the database and the tidal location method being applied are planned in the near future to improve the accuracy of the predicted locations.

Patterns of Vertical Movements

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 3. A number of specific temporal patterns were also identified within and across individual cod vertical movements

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. A 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, b) behavioral patterns related to temperature and depth data, c) behavioral patterns related to observed patterns of wind speed and direction, tidal stage, lunar stage, day length, and other local oceanographic features, and d) behavioral patterns related to location of the cod in the Icelandic waters.

REFERENCES

1. Hunter, E., Aldridge, J.N., Metcalfe, J.D. and Arnold, G.P. Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables. I. Tidal location method. *Marine Biology*, 142 (2003), 601-609.

- 2. Kurlansky, M. Cod: a biography of the fish that changed the world. Walker and Company, New York, 1997.
- Pálsson, Ó.K. and Thorsteinsson, V. 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 (2003), 1409-1423.
- Blumberg, A.F. and Mellor, G.L. A description of a three-dimensional coastal ocean circulation model. In *Three-Dimensional Coastal Ocean Models*, N. S. Heaps (Ed.), American Geophysical Union, Washington, DC, (1987), 1-16.

Measuring 3D Arm Movements for Activities of Daily Living

Thomas Haslwanter

Upper Austria University of Applied Sciences Dept. of Medical Technology, Garnisonstr. 21 4020 Linz, Austria thomas.haslwanter@fh-linz.at

ABSTRACT

Typical movements of the upper extremity in everyday life show a large variability between different, healthy subjects. In order to find reference arm movements for monitoring the progress of rehabilitation after stroke, we investigated the similarity of arm movement patterns on both sides of the body in healthy subjects for typical activities of daily living.

The movements of the upper extremity and the trunk, respectively, were recorded with a video-based 3D motion analysis system. Different statistical analyses of these data showed that the variability between the left and right body side of a subject is of the same size as the variability between different subjects. In other words, taking the movement trajectory of the healthy arm as a reference for the affected arm is not better than using the mean trajectory of a population as a reference, and simpler assessment criteria should be used.

Author Keywords

Arm movements, limb movement, activities of daily living, 3-dimensional, video, rehabilitation, quantitative assessment.

ACM Classification Keywords

J. Computer Applications, J.3 Life and Medical Sciences: Health

INTRODUCTION

While quantitative evaluations of leg movements are well established, for example in the analysis of gait [1], no such standards are available for arm movements. This lack of standards is caused mainly by two reasons: first, while the lower extremities are used predominantly for locomotion, hands and arms have to fulfill much more variable tasks. And second, the 3-dimensional analysis of arm movements Jürgen Waldhör

Upper Austria University of Applied Sciences Dept. of Medical Technology, Garnisonstr. 21 4020 Linz, Austria j.waldhoer@gmail.com

is quite complex, and has only recently started to be addressed [2]{Rau, 2000 6396 /id}. This makes the rehabilitation of arm movements, for example after stroke, more difficult: What should we aim for when we want to achieve a "normal" arm movement? We hypothesized that while the variability of arm movements is large between different subjects, it might be significantly smaller within one subject, when the movement kinematics of the right and left arm are compared for the same, well defined task.

METHODS

Subjects

In total twenty students voluntarily participated in our investigations. A pilot study, which was used to optimize the arm movement tasks suitable for our investigation, comprised eight male and two female subjects (mean age 27.6 +/- 6.9 yr). Based on the subjects' own estimation, three of them were classified as "dominantly left handed", and seven as "dominantly right handed". In the main study seven male and three female subjects participated (mean age 26.0 +/- 3.8 yr). All subjects except one were dominantly right handed. All of them were healthy, with no upper extremity complaints.

Paradigms

For the upper extremity there is no single most relevant functional task determinable. Instead, researchers often specify a set of "activities of daily living" (ADLs) for their investigations. For our goal – finding repeatable reference arm movements – we chose arm movements similar to those during typical ADLs. To limit the degrees of freedom (DOFs), we chose to measure only the orientation of the upper arm relative to the thorax. Inclusion of lower arm and hand movements would add additional DOFs, further increasing the variability and complicating the analysis.

For each subject both upper extremities were measured for all three tasks specified below. These tasks were based on earlier studies [2] but were slightly modified to be more representative for ADLs. For all tasks, the subject was sitting in a chair without armrests. The subjects were instructed to keep their torso upright with no contact to the back of the chair. The thighs should be parallel and the knees roughly at an angle of 90 deg. To ensure repeatable starting conditions, each hand started palm down, on the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

ipsilateral knee. Each task was repeated five times. The subjects were instructed to execute the repetitions at a moderate speed that should be similar for the individual repetitions and body sides, respectively. The subjects also had to remain in the start and end position for approximately one second. The following three tasks were measured:

- *Reaching (Task 1)* Subjects started as described above. The end point was reached when the index finger touched a point marked on a metal bar in front of the subject. The point was located 10 cm above the top of the subject's head, in the plane that goes through the shoulder joint and parallel to the mid-sagittal plane, and at a distance where the subject could reach it by almost fully extending the elbow (Figure 1, left). This task represents activities such as taking something from a shelf.
- *Hand to contralateral shoulder (Task 2)* Again, subjects started as described above. The subjects were instructed to touch the contralateral shoulder with the tip of the index finger (Figure 1, center). This task represents all activities near the contralateral shoulder, e.g. zipping up a jacket or washing the arm-pit.
- *Hand to hip pocket (Task 3)* Again subjects started in the starting position as described above. The end position was reached when the hand was placed on the ipsilateral hip pocket (Figure 1, right). This task represents reaching the back and accomplishment of perineal care.

Recording System

Movements of the upper extremity and torso were recorded using the LUKOtronic AS202 system (Lukotronic, Innsbruck, Austria), a video-based motion analysis system, for tracking the positions of active LED-markers attached to the body segments. This system allows tracking of actively controlled infrared markers with sample rates up to 1200 Hz, within a measurement range of up to 5-7 m. The system has the advantage that it is easily portable. The motion capture unit is connected to a laptop via USB and requires no additional power supply.

For our study, all data were sampled at a frequency of 100 Hz. Since our study focused on the relative movement of the upper limb with respect to the trunk, we decided not



Figure 1. Arm movement tasks (left to right): Pointing to a target 10 cm above the head; pointing to the contralateral shoulder; moving the hand to the hip pocket.



Figure 2. Markers attached to upper arm with a modified blood pressure cuff (left) and to the torso (right).

to monitor the complete anatomical chain (thorax - upper arm - lower arm - hand), but only thorax and upper arm. A total of six markers were attached to the subject - three on the upper arm und three on the trunk (Figure 2).

ANALYSIS

Description of Limb Movement

The data analysis was performed in Matlab (The Mathworks, Natick, Mass, USA), with programs developed in our group. Our investigation focused on the orientation of the arm relative to the torso. For each task, a representative trajectory was calculated by averaging the repetitions for each subject, body side and task.

We described the orientation of the upper arm relative to the trunk by using quaternions [3]. Compared to the more commonly used Euler angles, this method has two big advantages: it does not require a decomposition of the current orientation into a sequence of three consecutive rotations about arbitrarily chosen axes; and it avoids the problem of "gimbal lock" [4], which is unavoidable when using Euler angles.

To describe the orientation of the upper arm with respect to the torso, we define a torso-fixed coordinate system, with the x-axis pointing forward, the y-axis to the left, and the zaxis upward.

Evaluation of Limb Movement

Since the subjects performed the movement repetitions at different speeds, a normalization with respect to time is necessary before averaging. This is also essential for the subsequent comparison of different trajectories. An unavoidable consequence of time normalization is the loss of velocity information.

We perform the normalization with respect to time in two steps. First, the trajectory of each repetition is normalized to a length of 2000 data points. Depending on the duration of the repetition this can mean a shortening or a stretching of the data set. This normalization is accomplished by interpolation. Secondly, the duration of the phases in the trajectory is normalized. Therefore, each trajectory is divided into five phases that are defined by four characteristic points. These points have to be marked manually by the operator (for an example see Figure 3). Since we have five repetitions for each movement, this



Figure 3. Single trajectory of the dominant arm, with the four characteristic points indicated by black circles.

marking results in five time coordinates for each point. Taking the median of these coordinates enables us to calculate the average duration for each phase. With these average durations the phases of all repetitions are normalized by interpolation. For a quantitative comparison of two trajectories a parameter has to be defined that describes the deviation between them. To obtain one single parameter, we decided to take the area between the two trajectories, as shown in Figure 4. This can be done either for each component separately; or it can be calculated by determining the difference between trajectories as a curving ribbon in 3-dimensional space.

Statistical Analysis

To minimize the variability in each subgroup, we distinguished between "dominant" and "non-dominant" arm. We analyzed three different pairings of data:

- Trajectory of subject's dominant arm compared to the trajectory of subject's non-dominant arm.
- Trajectory of subject's dominant arm compared to the mean trajectory of all dominant arms in the group.

Trajectory of subject's non-dominant arm compared to the



Figure 4. We characterized the deviation between two trajectories by the area between them (here marked in gray).

mean trajectory of all non-dominant arms in the group.

First we checked if our data come from normal distributions. This was done with the Lilliefors test, which is an adaption of the Kolmogorow-Smirnow test. The Matlab-command *lillietest* tests the null hypothesis that the data set comes from a normal distribution, against the alternative that it comes from a different distribution.

We tested the following two hypotheses:

- *Hypothesis 1:* The deviation between the trajectory of a subject's non-dominant arm and the average trajectory of all measured non-dominant arms is not greater than the deviation between the trajectory of a subject's dominant arm and the average trajectory of all measured dominant arms.
- *Hypothesis 2:* The deviation between the trajectory of a subject's dominant arm and the average trajectory of all measured dominant arms is not greater than the deviation between the trajectories of a subject's left and right arm.

In all our tests the significance level was set to p < 0.05.

RESULTS

With one exception, all the pairings for all the tasks were normally distributed, which justifies our subsequent use of a t-test for the statistical data analysis.

To our surprise, none of the null hypotheses stated above could be rejected at a 5% significance level. Although the p-values for Task 3 are smaller than for the Tasks 1 and 2, the results do not give strong evidence that Task 3 is more appropriate for the evaluation of arm movements. Furthermore, there is no evidence that the separate analysis of the x, y and z component has any advantage over the three dimensional analysis.

DISCUSSION

We know that upper limb movements of the right and left side are to some extent coupled [5]. We also know that bilateral movement training can have significant beneficial effects in stroke rehabilitation [6;7]. What we don't know is how best to quantify arm movements, especially arm movements that have been recorded in three dimensions

| Task | Нуро | dev 3d | dev x | dev y | dev z |
|------|------|--------|-------|-------|-------|
| 1 | 1 | 87.1 | 67.4 | 92.0 | 73.2 |
| 1 | 2 | 66.2 | 87.9 | 23.6 | 56.6 |
| 2 | 1 | 70.1 | 83.6 | 72.0 | 31.4 |
| 2 | 2 | 79.0 | 35.9 | 87.5 | 67.7 |
| 3 | 1 | 51.2 | 52.9 | 77.5 | 25.0 |
| 3 | 2 | 52.0 | 31.9 | 52.3 | 54.6 |

Table 1. Results of the t-test for the main study, given in %. If the Hypothesis were wrong for a given Task, the corresponding value in the table would have to be less than 5%. (3D). We also don't know how to optimize the training of unilateral arm movements during rehabilitation. The reason for these difficulties lies in the variability and complexity of arm movements, and in the lack of standards for recording and analyzing arm movements in 3D space [8].

We speculated that the variability of arm movements might be reduced by a) restricting the recording and analysis to relative movements between upper arm and thorax, during selected typical activities of daily living; b) eliminating temporal variability by normalizing the executed movements in time; and c) by referencing the arm movements to the movement of the contra-lateral side in healthy subjects, not to average movements of a normal population. This expectation was bolstered by the finding of Mackey et al, that in children with hemiplegic cerebral palsy, the range of motion (ROM), timing, and peak angular velocity of the unaffected arm were similar to the corresponding parameters of the dominant arm of healthy control children [9]. Similarly, Macedo and Magee found little difference in the ROM between the dominant and nondominant arm in healthy subjects [10].

Somewhat to our surprise, the results of our investigations could not verify the hypotheses that we tested: it could not be shown that taking the movement pattern of the healthy arm as a reference for the affected arm is any better than using the mean trajectory of a population as a reference. Our study looked only for clear, reproducible correlations in the movement kinematics of right and left upper arm, which could provide a firm basis for new rehabilitation programs for arm movements. The detection of subtle correlations between right and left arm, or of gender related differences of movement kinematics, would require the inclusion of a larger number of subjects.

As handedness does not seem to be affected much by age [11;12], we speculate that the differences between our results and those from investigations looking only for difference in the ROM [9;10] are caused by the difference in the analysis: our analysis included the whole 3D trajectory of the movement, thereby highlighting any differences in any trajectory feature or component, while previous analyses focused on more global parameters such as ROM.

In summary, it can be stated that even for a well-defined set of movements that resemble activities of daily living, there is a large variability in the movement trajectories. Since both the individual variability between a person's left and right arm and the variability within the group are that large, meaningful reference parameters for a quantitative evaluation of arm movements have to be chosen carefully. Simple parameters like ROM are more likely to be successful than more detailed trajectory components.

REFERENCES

- 1. Perry, J. Gait Analysis: Normal and Pathological Function Slack Incorporated, 1992.
- 2. van Andel, C. J., Wolterbeek, N., Doorenbosch, C. A., Veeger, D. H., and Harlaar, J. Complete 3D kinematics of upper extremity functional tasks. *Gait Posture*, 27, 1 (2008), 120-127.
- 3. Haslwanter, T. Mathematics of three-dimensional eye rotations. *Vision Res.*, *35*, 12 (1995), 1727-1739.
- 4. Hanson, A. J. Visualizing Quaternions, Elsevier, 2006.
- 5. Swinnen, S. P. Intermanual coordination: from behavioural principles to neural-network interactions. *Nat. Rev. Neurosci.*, *3*, 5, (2002), 348-359.
- Stewart, K. C., Cauraugh, J. H., and Summers, J. J. Bilateral movement training and stroke rehabilitation: a systematic review and meta-analysis. *J. Neurol. Sci.*, 244, 1-2, (2006), 89-95.
- 7. Goble, D. J. The potential for utilizing inter-limb coupling in the rehabilitation of upper limb motor disability due to unilateral brain injury. *Disabil. Rehabil.*, 28, 18 (2006), 1103-1108.
- 8. Rau, G., sselhorst-Klug, C., and Schmidt, R. Movement biomechanics goes upwards: from the leg to the arm. *J. Biomech.*, *33*, 10 (2000), 1207-1216.
- 9. Mackey, A. H., Walt, S. E., and Stott, N. S. Deficits in upper-limb task performance in children with hemiplegic cerebral palsy as defined by 3-dimensional kinematics. *Arch. Phys. Med. Rehabil.*, 87, 2 (2006), 207-215.
- Macedo L. G., and Magee, D. J. Differences in range of motion between dominant and nondominant sides of upper and lower extremities. J. Manipulative Physiol. Ther., 31, 8 (2008), 577-582.
- 11. Rueckriegel, S. M., Blankenburg, F., Burghardt, R., Ehrlich, S., Henze, G., Mergl, R., and Hernaiz, D. P. Influence of age and movement complexity on kinematic hand movement parameters in childhood and adolescence. *Int. J. Dev. Neurosci.*, *26*, 7 (2008), 655-663.
- 12. Carlier, M., Dumont, A. M., Beau, J., and Michel, F. Hand performance of French children on a fingertapping test in relation to handedness, sex, and age. *Percept. Mot. Skills*, 76, 3, 1 (1993), 931-940.

ACKNOWLEDGEMENTS

This study was supported by the grant L425-N15 from the Fonds zur Förderung der wissenschaftlichen Forschung (FWF), and by the INTERREG IVC program of the European Union.

Review of Real-time Visual Driver Distraction Detection Algorithms

Christer Ahlström christer.ahlstrom@vti.se Katja Kircher katja.kircher@vti.se

Swedish National Road and Transport Research Institute (VTI), 581 95 Linköping, Sweden

ABSTRACT

Many incidents and crashes can be attributed to driver distraction, and it is essential to learn how to detect distraction in order to develop efficient countermeasures. A number of distraction detection algorithms have been developed over the years, and the objective of this paper is to summarize available approaches and to describe these algorithms in a unified framework. The review is limited to real-time algorithms that are intended to detect visual distraction.

Author Keywords

Driver distraction, eye tracking, inattention, detection algorithm.

INTRODUCTION

Driver inattention and distraction are major contributors to road incidents and crashes [1, 2]. Even so, drivers continue to engage themselves in distracting tasks such as using their mobile phones or navigation systems, eating, grooming and tending to their children. Advanced driver assistance systems may change this pattern by alerting the driver when his or her attention diverts from the road.

Driver distraction can be defined as "the diversion of attention away from activities critical for safe driving toward a competing activity" [2]. This is a very general definition where the diversion of attention can be visual, auditive, physical or cognitive and where the competing activity can be anything from mobile phone usage to getting lost in thought. New advances in remote eye tracking technology provide a means to counteract distracted driving in real-time. Eye movements can be used to gain access to several types of distraction. For example, studies have shown that eye movements are sensitive not only to visual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org. distraction but also to auditory secondary tasks [3-5].

There are numerous performance indicators that are based on longitudinal and lateral vehicle control dynamics which correlate with visual as well as cognitive task demands [6-8]. These include steering wheel reversal rate, average proportion of high frequency steering, brake reaction time, steering entropy, throttle hold, variability in lateral position, number of lane exceedences, time- and distance headway and time to collision. Even though many of these performance indicators seem to be promising secondary task identifiers, we restrict this survey to gaze based distraction detection algorithms. The objective of this paper is thus to summarize available real-time gaze based approaches for measuring visual driver distraction and to describe these algorithms in a unified framework. Since the focus of this review is on real-time assessment of visual distraction, many after-the-fact methods based on reaction times, secondary task performance and corrective manoeuvres are left out. A survey of the effects, in contrast to the prediction, of driver distraction can be found in Young et al. [9].

PRINCIPLES OF DRIVER DISTRACTION DETECTION

A schematic overview summarizing the structure of most driver distraction detection algorithms is illustrated in Figure 1. The basis for all algorithms is measures registered in real-time during driving. They can stem from the driver (driver behaviour), like eye movements or hand movements, or they can be logged from the vehicle (driving behaviour), like speed or lateral position. Furthermore, situational variables like time and position can be used (other data). Certain features of these data, like gaze direction, steering entropy or others are extracted and possibly fused in order to arrive at a continuous measure of the driver's distraction level. This output is then used to classify the driver's state of attention. For most algorithms these states are visually distracted vs. not visually distracted.

Field Relevant for Driving

Common for all eye or head movement based distraction algorithms is that they use off-road glances as the basic source of information. The idea is to define a field relevant



Figure 1. A schematic overview of the common features of most driver distraction detection algorithms and mitigation systems.

for driving, which is basically the area where the driver is looking when he or she is driving. If a world model is not available, the field relevant for driving can, for example, be defined as a circle [10-12] or a rectangle [13], see Figure 2. It is also possible to select different shapes. In Kircher et al. [14], a circle where the lower part was removed was used so that the dashboard would not be included in the field relevant for driving. Since there is no information about where the driver is looking in the real world, the selected field relevant for driving needs to be positioned in the real world based on statistics of where the driver has been looking. This is often done by centering the selected shape around the largest peak in the distribution of recent gazes. When enough gaze data has been acquired, it is also possible to define more than one zone based on the distribution of the data. For example, Kutila et al. [15] uses four zones (road ahead, windscreen and left/right exterior mirror.



Figure 2. Examples of three different approaches to select the field relevant from driving. The leftmost plot show a circular area, the middle plot shows a rectangular area and the rightmost plot shows an area that is defined based on the cockpit of the car.

If the eye tracking systems allows a world model to be used, the field relevant for driving can be defined based on different zones related to the interior of the car. This approach is used by Pohl et al. [16] and Kircher et al. [14], see Figure 2. In the latter of these two, the field relevant for driving is defined as the intersection between a viewing cone of 90 degrees and the vehicle's windows. This means that the circular field relevant for driving concept is expanded with information about the design of the car.

Distraction Estimation

Glances away from the road ahead are usually defined as glances residing outside the field relevant for driving. The duration of these glances away from the road ahead is the basic source of information that all visual distraction detection algorithms to date are based upon. If the driver is looking away from the road too often or for too long, the driver is considered distracted.

The mappings that transform glances away from the road to a continuous distraction estimate are often very similar. For example, Zhang et al. [13] used the average duration of glances away from the road in a 4.3-second wide sliding window, Donmez et al. [17] used a weighted sum of the current glance and the average glance duration in a 3second sliding window and Victor [10] used the percentage of on-road gaze data points in a 60-second sliding window. A slightly different approach is to use a buffer [14] or a counter [11] that changes its value when the driver looks away. Here the counter/buffer reaches a maximum/ minimum value when the driver is judged to be too distracted.

So far, there has been a direct link from the FRD via the glance duration to the estimated distraction level in the sense that all gazes have the same weight, regardless of where the gaze is directed. However, it is possible to make this link fuzzier by changing the weight as a function of where the gaze is directed. One idea is thus to penalize glances that are far away from the road centre. In the SafeTE project [12], this was done by the so-called eccentricity function $E(\alpha) = 6.5758 - 1/(0.001 * \alpha + 0.152)$. This is basically a weighting function that favours glances close to the road centre while penalizing glances with a large gaze direction angle. The equation is based on a study by Lamble et al. [18] and is related to visual behaviour and brake response when a lead vehicle suddenly starts to decelerate. In cases where a world model is available, it is possible to use different weights on different objects [14, 16]. For example, the rear view mirrors and the speedometer could have a higher weight as compared to the field relevant for driving but lower than the middle console or the glove compartment. Higher weights in this context mean that the distraction estimate will increase faster while lower weights have the opposite effect. Other combinations of the distraction estimation functions mentioned above, i.e. glance duration, glance history and eccentricity, has also been suggested [19].

Distraction Decision

The continuous distraction estimate needs to be mapped to a decision whether the driver is distracted or not. Basically, the driver enters the distracted state when a threshold is reached and returns to the attentive state when some criteria are fulfilled. The main difference between different approaches is how to leave the distracted state. One approach is to require that the driver is looking forward for some minimum time before he or she is considered to be attentive [14, 16, 17]. The other approach is that it is enough for the driver to look back at the road to be considered fully attentive [11].

Inhibition Criteria

A distraction detection algorithm determines whether a driver is distracted or not, but when and in which way the driver will be warned for distraction is determined by a warning strategy. Information about different warning strategies is out of the scope of this review. More information can be found in, for example, Donmez et al. [20]. However, there are situations when it is not suitable to give distraction warnings. For instance, if the driver is braking hard he or she is probably aware of the situation and should not be disturbed by a warning. For this reason, certain criteria can be set up to inhibit warnings. Common criteria include [21]:

• Speed: Below 50 km/h gaze behaviour is not very uniform. The gaze is often outside the FRD without the

driver being distracted.

- Direction indicators: Changing lanes and turning can include planned glances outside the FRD.
- Reverse gear: Reverse engaged means that the driver should look over the shoulder.
- Brake pedal: No warning should be given while driver is braking, in order not to interfere with critical driving manoeuvres.
- Steering wheel angle: No warning should be given while the driver is engaged in substantial changes of direction, in order not to interfere with critical driving manoeuvres.
- Lateral acceleration: No warning should be given when the vehicle makes strong movements, in order not to interfere with critical driving manoeuvres.

IMPROVEMENTS AND FUTURE RESEARCH

Available algorithms for eye tracking based driver distraction detection attempt to detect visual distraction. All algorithms can be fitted in a common framework; determine if the driver is looking at the road or not, convert this information into a continuous estimate of (visual) distraction and finally use some rule, often a threshold, to determine if the estimated level of distraction should be considered distracted or attentive. The main limitation of these approaches is that they do not take the current traffic situation into account. This could be done by allowing the field relevant for driving to change dynamically over time. Future research is needed to (a) determine the optimal field relevant for driving for different traffic situations and traffic environments and (b) develop technology to be able to measure the current traffic situation and traffic environment.

Only one of the available algorithms (percent road centre) was prepared in order to detect **internal distraction**. Suggested measures of internal distraction are based on the concentration of gazes towards the road centre area, which is higher when the driver is lost in thought. It has been suggested that other eye movements such as saccades and microsaccades could be indicative of workload or inattention. Future research is needed to (a) investigate eye movement physiology during driving, (b) develop remote eye tracking technology with higher accuracy so that these small and fast eye movements can be measured, and (c) develop algorithms that reliably and accurately detect different types of eye movements like fixations, saccades and smooth pursuit from the continuous data stream.

Other distraction indicators such as lateral and longitudinal control parameters seem to be very task and situation dependent, and it is questionable whether they can be used in a general purpose driver distraction detection algorithm. Future research includes fusion of several data sources, including situational variables, so that the appropriate set of performance indicators is used at exactly the right place at the right time. Even though it might be impossible to replace eye movement related indicators completely with
driving related parameters, it would be very valuable to be able to fall back on this type of data when eye tracking is lost.

- Gordon, C.P., Crash Studies of Driver Distraction, in Driver Distraction: Theory, effect and mitigation, M.A. Regan, J.D. Lee, and K.L. Young, Editors. CRC Press, Taylor & Francis Group, London (2009), 281-304.
- Lee, J.D., K.L. Young, and M.A. Regan, Defining Driver Distraction, in Driver Distraction: Theory, effect and mitigation, M.A. Regan, J.D. Lee, and K.L. Young, Editors. CRC Press, Taylor & Francis Group London (2009), 31-40.
- 3. Recarte, M.A. and L.M. Nunes, Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, *6*, 1 (2000), 31-43.
- Victor, T.W., J.L. Harbluk, and J. Engström, Sensitivity of eye-movement measures to in-vehicle task difficulty. Transportation Research Part F: *Traffic Psychology and Behaviour*, 8, 2 (2005), 167-190.
- Engström, J., E. Johansson, and J. Östlund, Effects of visual and cognitive load in real and simulated motorway driving. Transportation Research Part F: Traffic Psychology and Behaviour, 8, 2 (2005) 97-120.
- 6. Green, P. and R. Shah, Task time and glance measures of the use of telematics: A tabular summary of the literature, in SAfety VEhicles using adaptive Interface Technology (2004), UMTRI: Ann Arbor.
- Zylstra, B., et al., Driving performance for dialing, radio tuning, and destination entry while driving straight roads. The University of Michigan Transportation Research Institute: Ann Arbor, MI (2003).
- Östlund, J., et al., Driving performance assessment methods and metrics, in EU Deliverable, Adaptive Integrated Driver-Vehicle Interface Project (AIDE). (2005).
- Young, L.K., M.A. Regan, and J.D. Lee, Measuring the effects of driver distraction: Direct driving performance methods and measures, in Driver distraction: Theory, effects and mitigation, M.A. Regan, J.D. Lee, and K.L. Young, Editors. CRC PRess London (2009), 85-105.

- Victor, T.W., *Keeping eye and mind on the road*, in *Dept of Psychology*. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 9, Uppsala University: Uppsala (2005).
- Fletcher, L. and A. Zelinsky. Driver state monitoring to mitigate distraction. in International Conference on Driver Distraction. Sydney, Australia (2005).
- 12. Engström, J. and S. Mårdh, SafeTE Final Report (2007).
- Zhang, H., M. Smith, and R. Dufour, A Final Report of SAfety VEhicles using adaptive Interface Technology (Phase II: Task 7C): Visual Distraction (2008).
- Kircher, K. and C. Ahlstrom, Issues related to the driver distraction detection algorithm AttenD, in 1st International Conference on Driver Distraction and Inattention, Gothenburg, Sweden (2009).
- 15. Kutila, M., et al. Driver Distraction Detection with a Camera Vision System. in Proceedings of the IEEE International Conference on Image Processing. San Antonio, Texas, US (2007).
- 16. Pohl, J., W. Birk, and L. Westervall, A driverdistraction-based lane-keeping assistance system. Proceedings of the Institution of Mechanical Engineers Part I-*Journal of Systems and Control Engineering*, 221, I4 (2007), 541-552.
- 17. Donmez, B., L.N. Boyle, and J.D. Lee, Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis and Prevention*, *39*, 3 (2007), 581-590.
- Lamble, D., M. Laakso, and H. Summala, Detection thresholds in car following situations and peripheral vision: implications for positioning of visually demanding in-car displays. *Ergonomics*, 42, 6, (1999), 807-815.
- 19. Liang, Y., *Detecting Driver Distraction*. University of Iowa (2009).
- 20. Donmez, B., L.N. Boyle, and J.D. Lee, *Designing feedback to mitigate distraction*, in *Driver distraction: Theory, effects and mitigation*, M.A. Regan, J.D. Lee, and K.L. Young, Editors. CRC PRess London (2009), 519-531.
- 21. Kircher, K., A. Kircher, and F. Claezon, *Distraction and drowsiness*. *A field study*. VTI: Linköping (2009).

Using Semi-Automated Shadowing for Analysing Stress-Induced Spatio-Temporal Behaviour Patterns of Passengers in Public Transport Infrastructures

Alexandra Millonig alexandra.millonig@ait.ac.at **Gudrun Maierbrugger** gudrun.maierbrugger@ait.ac.at

Dynamic Transportation Systems, Austrian Institute of Technology Giefinggasse 2, 1210 Vienna; Austria

ABSTRACT

In the course of the scientific project IANUS, we examined the impacts of stress on the physical and psychological condition of passengers in train stations in order to identify relevant determinants and to develop recommendations for reducing stress-inducing factors in public transport infrastructures. We used a combination of biometric measuring of heart rate, semi-automated observation of spatio-temporal behaviour, psychological interviews, evetracking and visual field analyses during laboratory and field tests. In this contribution, we focus on the methodology of observing the spatio-temporal behaviour patterns of 35 individuals participating in field tests conducted in Vienna, Austria. The observation datasets have been analysed concerning velocities, route choice, stopping behaviour and related activities in order to identify noticeable patterns and events that can indicate stressrelated behaviour. The outcomes were subsequently consolidated with the results of the complementary methods in order to identify stress-inducing factors in transport infrastructures.

Author Keywords

Shadowing, across-method triangulation, pedestrian spatiotemporal behaviour.

ACM Classification Keywords

G.3 Experimental design, H.1.2 Human factors

INTRODUCTION

In the domain of traffic-related research, the impact of stress has primarily been examined in connection with pilot and driver behaviour. Findings have shown that frequent stressful driving encounters may ultimately lead to a dispositional tendency of drivers to experience all driving encounters in a negative manner [2]. The perspective of public transport passengers and the impact of stress on their physiological and psychological health has however largely been disregarded, although the same impact may be assumed when experiencing stressful situations while travelling with public transport facilities.

The aspect described has become of particular importance as today's public transport infrastructures provide a wide range of different services in addition to their primary function as traffic hubs. Due to this development, a growing amount of information is provided within station buildings, forcing passengers to filter out relevant information within limited time. Information overflow, time pressure and the complexity of infrastructures may cause uncertainties and can expose individuals to increased stress. As a consequence, individuals might try to avoid such situations in the future by not using public transport.

This is a development that has to be countered. To do so, it is firstly necessary to identify the key factors that increase stress of passengers in public transport situations. Secondly, the negative impacts of these factors have to be minimized in order to facilitate the completion of different tasks for passengers and augment the attractiveness of public transport usage.

As part of the scientific project IANUS we used a multidisciplinary approach comprising complementary methods to investigate stress-related behaviour of specific target groups. The aim of the project was to identify stressinducing factors in transport infrastructures and their effects on navigation behaviour of passengers as well as the passengers' strategies of gathering information and coping with stress. In this contribution, we focus on a particular method we used in this multi-disciplinary approach: the observation and analysis of spatio-temporal movement patterns of test subjects. We introduce the method, illustrate selected results and demonstrate how the results were consolidated with data obtained by the complementary methods in the project.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

ACROSS-METHOD TRIANGULATION FOR IDENTIFYING STRESS-INDUCING FACTORS

Recent findings have detected considerable divergences in measured physiological stress indicators and subjectively perceived stress levels. After a stressful test situation, test drivers stated to still experience heightened pressure, although no physiological evidence (e.g. high heart rate) could be measured any more [6]. This indicates that a broad set of indicators has to be used to examine stress comprehensively. In order to also take coping strategies and stress-related movement patterns into account, the phenomenon has to be investigated from several perspectives. We conducted experiments with participants of four different target groups in a laboratory environment and during field tests, applying the following methods:

- Questionnaires and personal interviews for gathering relevant information about attitudes, mood and stress coping strategies,
- Eye-Tracking (laboratory tests) and visual field analysis (field tests) for collecting data (visual focus points when orientating under stress),
- Physiological measurements of heart rate in the field tests and additionally skin response, pulse and body temperature in the laboratory environment for gathering information about physiological reactions on stress,
- Observation (shadowing) of spatio-temporal behaviour: semi-automated annotation of trajectories and activities (only for field tests) for identifying potentially stressinduced behaviour.

The selection of relevant target groups was based on previous findings in stress research. In the context of transport, two main individual-related characteristics show high relevance in connection with stress: age and experience in using transport systems. The age of a person is of relevance as the ability to take in and process information decreases with growing age. Stressful situations can aggravate this problem for elderly people and might as well cause specific reactions and coping strategies [1,7,9]. The importance of the second key attribute – whether people are accustomed to using public transport or not - lies in the fact that new situations and environments are generally perceived as more difficult than familiar situations and are therefore more likely to cause stress [4,10]. Taking these two factors into account, we defined four target groups showing combinations of the following characteristics: age groups younger than 35 and older than 55 years, and individuals who either used public transport frequently (several times per week) or rarely (at most several times a year). In total, 65 participants were tested in the laboratory environment and 35 participants took part in the experiments in the field tests. The test subjects had to pass navigation scenarios and fulfil specific tasks.

SHADOWING SPATIO-TEMPORAL BEHAVIOUR

Observing and analysing the spatio-temporal movement patterns of the test subjects was a main part of the field

tests. In general, observation techniques provide an essential link between different empirical methods in behavioural research. When investigating stress, this link gets of particular importance, as it is an especially complex phenomenon. We connected movement observation and analysis to the results of physiological and psychological measuring (for the importance of observations in stress-related studies see [3,8]).

Several potential indicators for stress-induced behaviour can be identified through observation. For our study, we specifically focused on motion-related indicators such as

- unusual speed levels (high speed hurrying, or very low speed hesitating, indicating uncertainties),
- frequent stops for gathering information, or
- uncertainties in route choice (e.g. changes in direction, turning back)

Collection of Trajectories During Field Tests

To collect the required spatio-temporal data, we used the method of "shadowing" [5]. "Shadowing" is a form of tracking where researchers follow the test subjects and annotate the test subjects' individual trajectories and related activities on a map. In the course of this study, this was done by means of special software installed on a tablet PC, which allowed annotating the information in digital form. During the field tests, participants had to follow instructions describing a specific scenario: they had to find a particular destination in the city by using predetermined modes of public transport (subway, railway), buy a ticket in the connecting station and reach a bus at the final station within limited time. While completing the tasks, they were additionally equipped with physiological measuring instruments and a mobile eye-tracking system. This allowed the annotation of their movement behaviour, heart rates and visual fields. Furthermore, they were interviewed before and after the experiment to collect data concerning general stress coping strategies and the current mood before and after the test. The shadowing data subsequently served as basis for comprehensive evaluation of the whole data set. Figure 1 illustrates an example of a typical trajectory: the line represents the path the participant followed. In addition, two stops are marked (in front of a ticket vending



Figure 1. Example of trajectory (coloured sections of the line represent different velocities).

machine and an information display).

The use of technology in this phase (digital map on a tablet PC, tracking software) offers mainly two major advantages: firstly, a large investigation area can be covered without having to handle a large paper map, and secondly, all points drawn in the map are recorded with time-stamps and map coordinates, which allows calculating average speeds and detecting stops for each trajectory. Additionally, the system allows annotating specific activities carried out by the participants when they stop (e.g. gathering information from a public display) and their duration.

The collected shadowing datasets finally included:

- Trajectories of the path a participant followed (including place and time of stops) and
- Activities the person performed on the way through the stations (e.g. gathering information, buying a ticket, waiting) including time and place of each activity.

Data Analysis and Identification of Noticeable Motion Patterns

After data cleansing, we analysed the collected datasets with respect to three indicators: velocities (differences in individual speeds, velocity histograms), stops (frequency, duration, position of stops and activities carried out during stops) and unusual route choice or significant changes in direction. The aim was to select datasets with noticeable behaviour in one or more categories for a subsequent detailed interdisciplinary analysis.

Velocity Histograms

To detect unusual speed patterns, we compiled speed histograms of each trajectory, showing the proportional amount of time (of the total time a test subject needed for completing the field scenario) an individual walked at a velocity within a specific speed interval. Figure 2 shows all histograms compiled from trajectory datasets collected in the connecting station.



Figure 2. Velocity histograms for all participants (at connecting station; velocity intervals in 0.1 m/s steps between 0.1 and 3 m/s).

Each line shows the histogram of an observed participant with higher intensities (lighter colors) indicating higher percentages of time; the values on the left represent the amount of time a person spent without moving. We used the histograms and additionally the average velocity of each participant to identify unusual patterns.

Stop Detection

The analysis of stopping behaviour included the detection of stops (defined as staying within a radius of 3.25 m for at least 5 s) and the analysis of annotated activities performed during those stops. To identify unusual behaviour, we focused on activities indicating uncertainties (e.g. high amount of time for gathering information) and stress coping activities (e.g. pacing up and down).

Route Choice Analysis

To identify unusual route choice,, we qualitatively compared the routes of all participants and selected examples of differing paths or changes in direction that were obviously due to foregoing incorrect decisions.

Consolidating Shadowing Results and Complementary Empirical Results

Based on this analysis of the shadowing results, we integrated the data gathered with the complementary methods for better interpretation. We employed video footage recorded with the eye-tracking system and used the contained visual fields of the test subjects to specify durations of activities and to include information concerning the situational context (e.g. dense crowds forcing individuals to deviate from the direct path). We furthermore made use of information collected in the interviews to interpret the participants' behaviour in general and with regard to specific situations and movement patterns.

This multidisciplinary view made it possible to select specific datasets showing noticeable behaviour patterns for further analysis. We subsequently examined the heart rates of participants selected based on observation results with particular focus on special registered events and specific information extracted from the interviews. This enabled a comprehensive view on stress-inducing factors and their effects on behaviour patterns of individuals.

The results show in many cases correlations between increasing heart rates and (partly upcoming) events (e.g. during periods of uncertainty or shortly before a train reached a station where the participant had to leave the subway). However, crediting this effect solely with psychological stress may not reflect the real picture: increasing heart rates may also result from e.g. stronger physiological activity (e.g. due to ascending a staircase). Some of the heart rate rises, though, are not connected to any changes in physical activity, as we found out by checking additional data from observation and video



Figure 3. Example of heart rate measurements during the beginning part of a field test.

footage. This shows clearly that it is essential to consider multiple data sources in order to deduce comprehensive interpretations in this context.

Figure 3 shows an example of heart rate values measured during the first sequence of a field test. The background shading marks consecutive phases of the test: the first section represents the baseline at rest which had been measured before the test started; the following section shows the heart rate during the beginning of the test. The next (short) section shows heart rate values measured during the participant passed the subway station to catch a train, and the last section covers the time the participant was riding on the subway train.

Considering the general influence of the test setting on the participants' stress-levels, it has to be stated that most participants reported to have experienced merely minor pressure during the experiments. Due to the artificial setting of the test scenario (participants did not really fear negative consequences in case of not reaching the final bus station in time) it was difficult to induce a realistic level of pressure on the participants. Still, the combined test results confirm interrelations between observed behaviour and measured stress indicators. It can be assumed that in test settings where a higher level of stress can be provoked, even more significant results can be expected.

CONCLUSION AND FUTURE DEVELOPMENTS

Results from this study and similar research settings confirm that observations are a valuable extension for investigating behaviour in especially complex situations. In particular, the semi-automated shadowing approach presented in this paper allows collecting a broad range of location- and time-related data for comprehensively analysing spatio-temporal behaviour patterns of pedestrians in different contexts. However, some limitations have to be accepted. The accuracy and completeness of the data strongly depends on the observer's tracking ability (a lot of training is required). Uncontrollable impacts of the situational context (high crowd densities, limited visibility) will also influence data quality to the worse. Finally, the method is very time-consuming and labour-intensive, and the presence of the observer may influence the participants' behaviour. Still, for tracking individual pedestrians within an infrastructure, semi-automated shadowing provides very good results especially in combination with complementary empirical methods for investigating pedestrian spatio-temporal behaviour.

For future applications of the method, we currently develop additional features such as annotating the duration of activities and annotating activities that are taken while in motion (for this study, this information had to be included based on video footage analysis), or the automated synchronisation of additionally collected data (such as "thinking aloud" voice recordings).

ACKNOWLEDGEMENTS

This work is part of the scientific project IANUS supported by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT). The authors thankfully acknowledge Gregory Telepak for collecting and preparing the tracking data and Helmut Lemmerer (TU Vienna) for providing the heart rate data.

- Cavanaugh, J. C. & Blanchard-Fields, F. Adult Development and Aging. Belmont, CA: Thomson Wadsworth, 2006.
- Gulian, E., Matthews, G., Glendon, A.I., and Davies, D.R. Dimensions of Driver Stress. *Ergonomics* 32 (1989), 585-602.
- Hartig, J. Stress: Begriffe und Definitionen, Stressreaktion und Stressoren, Erfassungsmethoden. http://www.joerghartig.de/Doc_Stresstheorie_Joerg_Hartig.pdf
- Lazarus, R. S., Cohen, J. B. Environmental Stress. In: Altman, I., Wohlwill, J. E. (Eds.): Human Behavior and Environment. Vol. 2. Plenum, New York, 1977.
- Millonig, A., Brändle, N., Ray, M., Bauer, D., and van der Spek, S.C. Pedestrian Behaviour Monitoring: Methods and Experiences. In: Gottfried, B., Aghajan, H. (Eds.): Behaviour Monitoring and Interpretation – BMI. Smart Environments. IOS Press, Amsterdam, 2009.
- Schießl, C. Subjective strain estimation depending on driving manoeuvres and traffic situation. In *Proc. of Europ conf on HCD for ITS*, HUMANIST, Lyon (2008).
- Schlag B. Elderly drivers in Germany fitness and driving behaviour. Accident Analysis & Prevention 25,1 (1993), 47-55.
- 8. Seefeldt, D. Stress. *Verstehen Erkennen Bewältigen*. Dreieich: Edition Wötzel, 2002.
- Whelan, M., Langford, J., Oxley, J., Koppel, S., and Charlton, J. *The elderly and mobility: A review of the literature*. Monach University Accident Research Centre, Australia, 2006.
- Zimbardo, P. G. and Gerrig, R. J. *Psychologie*. Berlin, Heidelberg: Springer, 1999.

Assessment of Aging Effects on Drivers' Perceptual and Behavioral Responses Using Subjective Ratings and Pressure Measures

Gyouhyung Kyung

School of Design & Human Engineering, UNIST 100 Banyeon-ri, Eonyang-eup, Ulju-gun, Ulsan, Korea ghkyung@unist.ac.kr

ABSTRACT

A study was conducted in which a total of 22 younger and older participants completed six short-term driving sessions. These sessions involved a subset of combinations of vehicle class (sedan and SUV), driving venue (lab-based vs. field) and seats (from vehicles ranked high and low by J.D. Power and Associates' Comfort Score). Three subjective ratings (comfort, discomfort, and overall) were obtained, along with 36 driver-seat interface pressure measures. For both age groups, localized comfort ratings were found to be more effective at distinguishing among automotive seats / packages, compared to global ratings or localized discomfort ratings. In addition, older individuals appeared to be less sensitive to discomfort than younger individuals. Several pressure measures indicated different dynamic behaviors or loading patterns (due to postural differences) between the two age groups, and bilateral asymmetry of driving postures in general. These results indicate that, when designing car seats and interior geometries, different pressure requirements should be specified and used separately for each age group and for each seat side.

Author Keywords

Aging, comfort, discomfort, driving experience, interface pressure.

ACM Classification Keywords

H5.2. User Interfaces: Ergonomics, User-centered design.

INTRODUCTION

For most people, aging leads to several decrements including poor eyesight, slow reaction time, lack of muscular strength and dexterity, susceptibility to fatigue (Warnes et al., 1993), and loss of joint flexibility (Haywood et al., 1991), each of which has the potential to adversely affect an individual's driving experience and posture. As a

Maury A. Nussbaum

Industrial & Systems Engineering, Virginia Tech Blacksburg, VA, USA nussbaum@vt.edu

specific example highlighting the importance of understanding aging effects in the context of seat design, Reynolds (1993) showed that different forms of spinal curves (i.e., "flatter and more kyphotic") were observed in the sixth decade of life. Hence, to the contrary to its intention, a contoured seat back could actually make the elderly uncomfortable. Aging effects on driving posture and postural sensitivity have been observed previously in automotive seating (Kyung and Nussbaum, 2009; Kyung et al., 2007), with older drivers preferring to sit closer to the steering wheel. From a safety perspective, Burger et al. (1977) showed that the design of the vehicle interior contributed to at least 7.5% of all accidents. However, with respect to driver workspace and interface design, agerelated differences in the efficacy of perceptual ratings have not been investigated, nor is there any evidence regarding whether a single rating scheme might be effective across a diverse age range.

Perceived (dis)comfort can affect drivers' performance, safety, and even posture. Specifically, visual (dis)comfort induced by vehicle in/exterior design is related to hazard misperception (e.g., gap, distance) as well as hazard nonperception (e.g., vision obstruction, reduced visibility), both of which are regarded as major crash contribution factors (Wierwille and Tijerina, 1996). Anshel's study (2005) on man-computer systems showed that visual information is so dominant that its deficiency was compensated by changing body posture. Most elderly individuals have slower and limited abilities in static and dynamic visual acuity (Eby and Kantowitz, 2006; Nicolle and Abascal, 2001), compared to younger individuals. Hence, the former group is more likely to adopt different driving postures due to the decrement in their vision, as well as due to the difference in their normal posture (e.g., kypothic spine). Postural or physiological differences between age groups likely lead to different behaviors (hence, different loading patterns on the seat) within a driver workspace, which necessitates agespecific requirements in terms of interface pressure between the seat and the driver. Hanson et al. (2006) disclosed that driver sitting experience is related to both comfort and discomfort, similar to sitting experience in office or home chairs (Helander and Zhang, 1997; Zhang et al., 1996).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Though not a necessary condition, nociceptors at the nerve endings are generally responsible for nociception, the perception of pain (Brooks and Tracey, 2005). Pain is a factor of discomfort along with poor biomechanics and tiredness, whereas comfort factors are well-being and plushness (Zhang et al., 1996). Less is known, however, whether there are differences between age groups in the perceptions of comfort and discomfort of sitting experience, in terms of magnitude and dominance of each perception. The goals of this study were to investigate whether aging affects 1) the efficacy of each of three subjective ratings (i.e., comfort, discomfort, and overall) when used for designing and evaluating driver workspace, 2) preference for sedan and SUV settings, 3) the decision processes involved when relating whole-body and localized perceptions of comfort and discomfort, and 4) the associations between subjective ratings and pressure measures.

EXPERIMENTAL METHODS

Overview of Experiment And Participants

Eleven older individuals were newly recruited, whereas 11 vounger individual data were selected from among 27 younger individual data (used in Kyung and Nussbaum, 2008; Kyung et al., 2008). The latter were selected in order to achieve a close match in terms of gender distribution, stature, body mass, and data size between the two age groups (Table 1). No significant differences existed in terms of stature and body mass (t-test, p = 0.68 and 0.23). Each participant completed six driving sessions, and subjective ratings and interface pressure measures were obtained in the same way as described in (Kyung and Nussbaum, 2008; Kyung et al., 2008). Each participant completed an informed consent procedure, approved by the local Institutional Review Board, prior to the first experiment session. Brief descriptions of the experimental procedure and settings, and subjective and objective measures used in this study, are given below. For more information, readers are referred to Kyung et al. (2008) and Kyung and Nussbaum (2008).

Six driving sessions (later called Seat Condition) combined two vehicle classes (sedan [S]; SUV [U]), two driving venues (lab-based, [-L]; field-based, [-F]), and two seats (from vehicles ranked high [1] and low [2] by J. D. Power and Associates (2005)' Comfort Score) per vehicle class.

| | 1 | |
|--------------------------|--------------|--------------|
| | Age Group | |
| | Younger | Older |
| # of Participants (M/F)* | 11 (6/5) | 11 (6/5) |
| Mean (SD) Stature (cm) | 168.9 (11.2) | 168.2 (11.7) |
| Mean (SD) Age (year) | 21.8 (3.2) | 71.4 (8.6) |
| Mean (SD) Body mass (kg) | 67.9 (11.1) | 73.5 (22.0) |

* Number of males and females

 Table 1. Participant characteristics.

Hence, specific Seat Conditions were S1-L, S1-F, S2-L, U1-L, U1-F, and U2-L. Participants completed three sessions (two in the lab and one in the field) for each vehicle class. An adjustable driving rig was used in the lab-based sessions that involved simulated driving, and two cars were used for the field sessions that involved on-the-road driving. In both cases, driving was conducted for 20 minutes. Before and during driving, participants adjusted the seat and steering wheel to best support their preferred driving postures. As in Kyung et al. (2008), a modified method of fitting trials (Jones, 1969) was used for the initial adjustment.

After driving, and while maintaining their preferred postures, participants rated their postures in terms of comfort, discomfort, and a combination of these two (overall rating). Several scales were used by participants, at the end of each driving session, to assess drivers' perceptions. Comfort and discomfort scales were derived as combinations of versions developed by Borg (1990) and Corlett-Bishop (1976), and used for the whole body and six local body parts (bilateral thighs and buttocks, and lower and upper back). In addition, a visual analogue scale (VAS) was used to obtain overall perceptual ratings of the whole body. The Karolinska Sleepiness Scale (KSS) by Horne and Reyner (1995) was used to measure drivers' alertness level. Two Tekscan (South Boston, MA, USA) pressure mats (5330 CONFORMatTM) were used to collect a variety of pressure data over the course of each driving session. The first pressure mat was used on the seat cushion and was divided into four areas corresponding to bilateral thighs and buttocks. The second pressure mat was hung on, and tied to the seat back, and divided into two areas corresponding to the lower and upper back. Six types of variables were measured from each of six divided area: contact area, contact pressure, peak pressure, and ratio of each of these three variables (local to global). Hence, a total of 36 interface variables were used to describe the driver-seat interface.

Data Collection and Processing

Subjective ratings were obtained after each session, in a consistent order to minimize confusion (discomfort, comfort, and then overall ratings). Pressure data was collected continuously during the driving sessions, using two Tekscan (South Boston, MA, USA) pressure mats (5330 CONFORMatTM). Each pressure mat was comprised of 1024 (32 x 32) thin (1.78mm) resistive sensors that could easily conform to the contour of the seat, and measure up to 250 mmHg (5 PSI). Each mat had an active area of 471.4 mm x 471.4 mm, and sensor pitch was 14.73 mm (0.5 sensor / cm2). Pressures were recorded at 0.5 Hz, the maximum possible due to hardware limitations. This sampling rate, however, was considered sufficient, as the frequency of postural changes and resultant pressure changes were not observed to occur within an order of magnitude of the sampling rate.



Figure 1. Division of two pressure mats for six local body parts (left, number of sensors in parentheses; row by column) and exemplar pressure distribution (right, a higher peak pressure on left buttock) - (Kyung and Nussbaum, 2008)

Pressure data from the two mats were divided into six groups (Figure 1). Bolsters on the sides of the car seat play a role in supporting thighs, hips, and back (especially when turning), and can affect sitting comfort and discomfort (Andreoni et al., 2002). To account for this, pressure data corresponding to bolstered areas were also included in the data analysis. Contact area and contact pressure were calculated by including only data from sensors that were pressed (i.e. a positive value) at least once, and average (arithmetic mean) values were determined for the last five minutes of driving. Earlier data were excluded here, as they were 'transient' due to settling into the seat (Reed et al., 1999). Contact area and pressure values were obtained from the Movie Contact Averaging function available in the CONFORMat Research software (version 5.80c).

Data Analysis

Methods for data analysis are similar to those described earlier ((Kyung and Nussbaum, 2008; Kyung et al., 2008), except that an Age factor was assessed rather than Stature. Specifically, a mixed-factor analysis of variance (ANOVA) was used to determine the effects of Age (2 levels, between-subjects) and Seat Condition (6 levels, withinsubject) on each of the three subjective ratings. Tukey's Honestly Significantly Different (HSD) test was used, where relevant, for post-hoc pairwise comparisons. Effects were considered 'significant' when p≤0.05, with potential trends highlighted when 0.05 . Specific pairwisecomparisons were done to determine if there were driving venue (lab vs. field) effects (S1-L vs. S1-F, U1-L vs. U1-F) and/or seat effects within vehicle class (S1-L vs. S2-L, U1-L vs. U2-L). In addition, a linear contrast of vehicle classes (S1-L + S2-L + S1-F vs. U1-L + U2-L + U1-F) was tested to compare preferences between the two classes. Ratings between bilateral body parts (thighs, buttocks) were examined using matched pairs comparisons, for each age group. Bivariate coefficients of correlation (p) were

obtained among the several subjective ratings, at both local and global levels, for each age group. Possible aesthetic effects of four seats on comfort ratings were examined using a univariate repeated-measures ANOVA. Additional bivariate coefficients of correlation (ρ) were obtained between each of three whole-body subjective ratings and alertness level.

With respect to the 36 pressure variables, a Multivariate Analysis of Variance (MANOVA) was conducted to preserve the overall significance level, prior to an additional mixed-factor ANOVA. The later was similar to the one described above, but involved different dependent variables (i.e., 36 pressure variables). The same pairwise comparisons as described above were also used as post-hoc analysis. Among these, the comparison between sedan seats (S1-L vs. S2-L), and the contrast between vehicle classes (sedans vs. SUVs) were of particular interest, as only these two were previously found to be differentiated based on subjective ratings obtained from younger individuals (see Kyung et al., 2008). Driving venue effects were also of interest, as only one driving venue (field) involved exposure to road/vehicle vibration, and which was expected to affect the pressure variables as observed in Kyung and Nussbaum (2008). Comparisons between bilateral pressure measures (i.e. at the thighs and buttocks) were made using matched-pairs t-tests. Additionally, bivariate coefficients of correlation (p) were obtained between each of the three subjective ratings and each of the 36 pressure variables. The presence of the following three statistical results was interpreted as supporting the general use of a pressure variable in seat design and evaluation, regardless of drivers' age: 1) an association between a given pressure variable and any of the subjective ratings; 2) a significant Seat Condition effect on the pressure variable; and 3) a lack of a significant Age effect on the pressure variable.

As in Kyung and Nussbaum (2008), relationships between subjective ratings and pressure variables were further investigated using two steps: 1) a principal component analysis (PCA); and 2) a multiple regression of each subjective rating on the factors from the PCA. Here, the focus was to develop a method that can be applied to improve car seats and packages for both age groups. The number of factors were determined by two criteria : 1) eigenvalue > 1; and 2) the cumulative percentage of variance close to 90% (Lehman et al., 2005). The selected factors were rotated using the varimax method.

RESULTS/DISCUSSION

Major results from the current study are listed below.

- At the whole-body level, none of the three rating schemes was particularly effective at distinguishing car seats.
- At the localized level, Age effects were identified, but, only in terms of comfort.
- Several pressure measures also supported that there were

different pressure loadings (due to postural differences) between two age groups.

- From the correlation and regression analyses, average pressure at the right buttock (avgBTR) should be higher to reduce discomfort and increase comfort.
- There are additional ways of improving seat design using the pressure measures obtained in this study (e.g., the seat cushion should be made softer, especially the area contacting the buttocks).
- Lower back and upper back discomforts had, respectively, the highest correlations (ρ =0.74, 0.54) with whole body discomfort than any other local body part discomfort for the younger and older groups.

CONCLUSION

The current study showed that at local levels, comfort ratings are effective at evaluating driver workspace, and that age did not influence the processes used when determining whole-body perceptions based on localized perceptions. Some age-related differences were also identified; different preferences for vehicle class, different local body part predominantly determining whole-body discomfort, and several pressure measures of significantly different values. In addition, older individuals appeared to be less sensitive to discomfort than younger individuals. These similarities and differences should be carefully considered when designing the driver workspace.

- 1. Andreoni, G., Santambrogio, G.C., Rabuffetti, M., Pedotti, A. Method for the analysis of posture and interface pressure of car drivers. *Applied Ergonomics*, 33, 6 (2002), 511-522.
- Anshel, J. Computer Vision Syndrome. In J. Anshel (Ed.), Visual Ergonomics Handbook, Boca Raton: Taylor & Francis Group (2005), pp. 23-35.
- 3. Borg, G. Psychophysical scaling with applications in physical work and perception of exertion. *Scandinavian Journal of Work, Environment and Health, 16*, 1 (1990), 55-58.
- 4. Brooks, J., Tracey, I. From nociception to pain perception: imaging the spinal and supraspinal pathways. *Journal of Anatomy*, 207, 1 (2005), 19-33.
- Burger, W., Smith, R., Queen, J., Slack, G. Accident and near accident causation: the contribution of automobile design characteristics (No. HS-5-01216): Department of Transportation, 1977).
- 6. Corlett, E.N., Bishop, R.P. A technique for assessing postural discomfort. *Ergonomics*, 19, 2 (1976), 175-182.
- Eby, D.W., Kantowitz, B.H. Human Factors and Ergonomics in Motor Vehicle Transportation. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed.): John Willey & Sons, 2006.

- Hanson, L., Sperling, L., Akselsson, R. Preferred car driving posture using 3-D information. *International Journal of Vehicle Design*, 42, 1-2 (2006), 154-169.
- 9. Haywood, K.M., Williams, K., Vansant, A. Qualitative assessment of the backswing in older adult throwing. *Research Quarterly for Exercise and Sport, 62, 3* (1991), 340-343.
- 10. Helander, M., Zhang, L. Field studies of comfort and discomfort in sitting. *Ergonomics*, 40, 9(1997), 895-915.
- 11. Horne, J.A., Reyner, L.A. Sleep related vehicle accidents. *British Medical Journal*, 310, (1995), 565-567.
- 12. Kyung, G., Nussbaum, M.A. Driver sitting comfort and discomfort (part II): Relationships with and prediction from interface pressure. *International Journal of Industrial Ergonomics*, 38, 5-6 (2008), 526-538.
- 13. Kyung, G., Nussbaum, M.A. Specifying comfortable driving postures for ergonomic design and evaluation of the driver workspace using digital human models. *Ergonomics*, *52*, 8 (2009), 939-953.
- 14. Kyung, G., Nussbaum, M.A., Babski-Reeves, K. Driver sitting comfort and discomfort (part I): Use of subjective ratings in discriminating car seats and correspondence among ratings. *International Journal of Industrial Ergonomics*, 38, 5-6 (2008), 516-525.
- 15. Kyung, G., Nussbaum, M.A., Lee, S., Kim, S., Baek, K. Sensitivity of preferred driving postures and determination of core seat track ranges. *SAE Technical Paper Series (2007-01-2471)*, 2007.
- 16. Nicolle, C., Abascal, J. *Inclusive Design Guidelines for HCI*: Taylor & Francis, 2001.
- 17. Reed, M.P., Manary, M.A., Schneider, L.W. Methods for measuring and representing automobile occupant posture. *SAE Technical Paper Series (1999-01-0959)*, 1999.
- Reynolds, H.M. Automotive seat design for sitting comfort. In B. Peacock (Ed.), *Automotive Ergonomics* London, Washington, D.C.: Taylor & Francis, (1993), 99-116.
- Warnes, A.M., Fraser, D.A., Hawken, R.E., Sievey, V. Elderly drivers and new road transport technology. In A. M. Parkes, S. Franzen (Eds.), *Driving Future Vehicles* London, Washington, D.C.: Taylor & Francis, (1993), 99-117).
- 20. Wierwille, W.W., Tijerina, L. An Analysis of Driving Accident Narratives As a Means of Determining Problems Caused by In-Vehicle Visual Allocation and Visual Workload. In A.G. Gale, I.D. Brown, C.M. Haslegrave, S.P. Taylor (Eds.), *Vision in Vehicles*, Amsterdam, The Netherlands: Elsevier, (1996). 79-86.
- 21. Zhang, L., Helander, M., Drury, C. Identifying factors of comfort and discomfort. *Human Factors*, *38*, *3*, (1996). 377-389.

Applying Nonlinear Dynamics Features for Speech-based **Fatigue Detection**

Jarek Krajewski¹ krajewsk@uni-wuppertal.de

Tom Laufenberg¹ laufenberg@uni-wuppertal.de

David Sommer² d.sommer@fh-sm.de

Christian Heinze² c.heinze@fh-sm.de

Thomas Schnupp² t.schnupp@fh-sm.de

Martin Golz² m.golz@fh-sm.de

¹University Wuppertal **Experimental Industrial Psychology** Gaußstraße 20, 42097 Wuppertal, Germany

ABSTRACT

This paper describes a speech signal processing method to measure fatigue from speech. The advantages of this realtime approach are that obtaining speech data is non obtrusive, free from sensor application and calibration efforts. Applying methods of Non Linear Dynamics (NLD) provides additional information regarding the dynamics and structure of fatigue speech comparing to the commonly applied speech emotion recognition feature set (e.g. fundamental frequency, intensity, pause patterns, formants, cepstral coefficients). We achieved significant correlations between fatigue and NLD features of 0.29. The validity of this approach is briefly discussed by summarizing the empirical results of a sleep deprivation study.

Author Keywords

Speech signal processing, nonlinear dynamics features, speech emotion recognition, human-computer-interaction.

ACM Classification Keywords

H.5.5 Sound and music computing

INTRODUCTION

The prediction and warning of drivers against impending critical fatigue play an important role in preventing accidents and the resulting human and financial costs. Hence, many efforts have been reported in the literature for measuring fatigue related states [1-5]. In contrast to electrode- or video-based instruments, the utilization of voice communication as an indicator for fatigue could match the demands of everyday life measurement. Contact free measurements as voice analysis are non-obtrusive (not interfering with the primary driving task) and favorable for fatigue detection since an application of sensors would

cause annoyance, additional stress and often impairs working capabilities and mobility demands. In addition, speech is easy to record even under extreme environmental conditions (bright light, high humidity and temperature), requires merely cheap, durable, and maintenance free sensors and most importantly, it utilizes already existing communication system hardware. Furthermore, speech data is omnipresent in many professional driver settings. Nevertheless, several sources of noise during driving, as e.g. motor sound, radio, and sidetalk can lead to difficult recording situations. In sum, given these obvious advantages, the renewed interest in computational demanding analyses of vocal expressions has been enabled just recently by the advances in computer processing speed.

²University of Applied Sciences Schmalkalden

Neuroinformatics and Signal Processing

Blechhammer, 98574 Schmalkalden, Germany

An important aspect in the vocal tract during fatigue influenced speech production is the generation of nonlinear aerodynamic phenomena including non-laminar flow, flow separation in various regions, generation and propagation of vortices and formation of jets rather than well-behaved laminar flow [6-8]. The collapse of laminar flow arises at high reynolds number. Due to the relevant length and subsonic speed of air flow in the vocal tract, this number is very large, indicating that the air flow can be expected to be turbulent. The air jet flowing through the vocal tract during speech production includes convoluted paths of rapidly varying velocity, which are highly unstable and oscillate between its walls, attaching or detaching itself, and thereby changing the effective cross-sectional areas and air masses.

Several issues are responsible for the generation of these nonlinear effects: The vocal folds behave as a vibrating valve, disrupting the constant airflow from the lungs and forming it into regular puffs of air. Modeling approaches which have their origin in fluid dynamics coupled with the elastodynamics of a deformable solid understand this phonation process as nonlinear oscillation: dynamical forcing from the lungs provides the energy needed to overcome dissipation in the vocal fold tissue and vocal tract air. The vocal folds themselves are modeled as elastic tissue

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

with nonlinear stress-strain relationship. These nonlinear stretching qualities of the vocal folds are based on larynx muscles and cartilage which produces nonlinear behavior. Furthermore, vocal tract and the vocal folds are coupled when the glottis is open resulting in significant changes in formant characteristics between open and closed glottis cycles. The movement of the vocal folds themselves is modeled by a lumped two mass system connected by springs again with nonlinear coupling. These nonlinear phenomena produce turbulent flow while the air jet may be modulated either by the vibration of the walls or by the generated vortices. Several methods based on chaotic dynamics and fractal theory have been suggested to describe these aerodynamic turbulence related phenomena of the speech production system [9-17] including the modeling of the geometrical structures in turbulence (spatial structure, energy cascade) utilizing fractals and multifractals [15-19], nonlinear oscillator models [20-22], state-space reconstruction. This state-space and reconstruction is done utilizing the embedding theorem which reconstructs a multidimensional attractor by embedding the scalar signal into a phase space. The embedding allows us to reconstruct the geometrical structure of the original attractor of the system which formed the observed speech signal. Moreover, it helps us to discover the degree of determinism of an apparently random signal, e.g. by applying measures like Lyapunov exponents.

However, no empirical research has been done to examine the turbulence effects in speech signals, which might be induced by fatigue related change of heat conduction within the vocal tract. Previous work associating changes in voice with fatigue [23-24] has generally focused only on features derived from speech emotion recognition [25] whereas nonlinear dynamics based speech features [26] have received no attention. Thus, the aim of this study is to apply nonlinear dynamics (NLD) based features within the field of speech acoustics in order to improve the prediction of fatigue.

METHODS

Seventeen students, recruited from the University of Wuppertal (Germany), took part in this study voluntarily. The participants were instructed to maintain their normal sleep pattern and behaviour. We conducted a within-subject sleep deprivation design (8.00 p.m to 4.00 a.m). During the night of sleep deprivation a well-established, standardized self-report fatigue measure, the Stanford Sleepiness Scale (SSS), was used by the subjects and two experimental assistants just before the recordings to determine the fatigue state. On this scale, a score of 1 point indicates "feeling active and vital, alert, wide awake" and a score of 7 points indicates "almost in reverie, sleep onset soon, losing struggle to remain awake" (M = 5.06; SD = 2.02). During the night, the subjects were confined to the laboratory and supervised throughout the whole period. Between sessions,

they remained in a room, watched DVD, and talked. Non caffeinated beverages and snacks were available ad libitum.

Speech Material and Recording. The recording took place in a laboratory room with dampened acoustics using a highquality, clip-on microphone (sampling rate: 44.1 kHz, 16 bit). The input level of the sound recording was kept constant throughout the recordings. Furthermore the subjects were given sufficient prior practice so that they were not uncomfortable with this procedure. The verbal material consisted of a long vowel [o:] extracted from a German phrase: "Rufen Sie den N[o:]tdienst" ("Please call the ambulance"). The sentence was taken from simulated communication with a driver assistance system. The participants recorded other verbal material at the same session, but in this article we focus on the material described above (17 subjects x 4 sentences = 68 speech samples).

Seventeen students, recruited from the University of Wuppertal (Germany), took part in this study voluntarily. The participants were instructed to maintain their normal sleep pattern and behaviour. We conducted a within-subject sleep deprivation design (8.00 p.m to 4.00 a.m). During the night of sleep deprivation a well-established, standardized self-report fatigue measure, the Stanford Sleepiness Scale (SSS), was used by the subjects and two experimental assistants just before the recordings to determine the fatigue state. On this scale, a score of 1 point indicates "feeling active and vital, alert, wide awake" and a score of 7 points indicates "almost in reverie, sleep onset soon, losing struggle to remain awake" (M = 5.06; SD = 2.02). During the night, the subjects were confined to the laboratory and supervised throughout the whole period. Between sessions, they remained in a room, watched DVD, and talked. Non caffeinated beverages and snacks were available ad libitum.

Speech Material and Recording. The recording took place in a laboratory room with dampened acoustics using a highquality, clip-on microphone (sampling rate: 44.1 kHz, 16 bit). The input level of the sound recording was kept constant throughout the recordings. Furthermore the subjects were given sufficient prior practice so that they were not uncomfortable with this procedure. The verbal material consisted of a long vowel [o:] extracted from a German phrase: "Rufen Sie den N[o:]tdienst" ("Please call the ambulance"). The sentence was taken from simulated communication with a driver assistance system. The participants recorded other verbal material at the same session, but in this article we focus on the material described above (17 subjects x 4 sentences = 68 speech samples).

Feature Extraction. The following feature families within nonlinear time series analysis are computed: (a) *State space features (375#).* To extract the nonlinear properties of the speech angle signal, we computed a three-dimensional state space (sound pressure, Δ sound pressure, $\Delta\Delta$ sound pressure), and a reconstructed phase space (d = 3,4; $\tau = 1$;

sound pressure t_0 , sound pressure t_1 , sound pressure t_2). The geometrical properties of the resulting attractor figures were described by trajectory based descriptor contours (angle between consecutive trajectory parts, distance to centroid of attractor, length of trajectory leg). The temporal information of the contours was captured by computing functionals; (b) Fractal features (110#). They try to quantify self-affinity and underlying complexity of the speech signal, e.g. box counting dimension, Cao's minimum embedding dimensions, correlation dimension, fractal dimension, information dimension; (c) Entropy features (5#). They assess the regularity/irregularity or randomness of speech signal fluctuations, as e.g. the Lyapunov exponents, which characterize the system's "degree of chaos" by estimating the exponential rate of divergence or convergence of nearby orbits on its phasespace. Positive Lyapunov exponents indicate divergence of nearby orbits and thus long-term unpredictability.



Figure 1. Reconstructed phase spaces (d = 3 ; τ = 10; sound pressure t₀, sound pressure t₁, sound pressure t₂) for an alert (left) and fatigue (right) speech sample [a:]. The geometrical properties of the resulting attractor figures were described e.g. by trajectory based descriptor contours (angle between consecutive trajectory parts, distance to centroid of attractor, length of trajectory leg) or entropy features as Lyapunov exponents, which characterize the system's "degree of chaos" by estimating the exponential rate of divergence or convergence of nearby orbits.

RESULTS

In order to quantify the association between fatigue and (a) commonly applied SER speech features and (b) NLD features, we computed Pearson's correlation coefficients The significant correlations are displayed in Table 1 and 2.

| SER Features | r |
|--|------|
| standard deviation of intensity | 29* |
| range of fundamental frequency | 24* |
| jitter (small fundamental frequency perturbations) | 33* |
| mean of delta-mel-frequency cepstrum coefficient 7 | .34* |
| mean of formant 3 position (third maximum of vocal tract transfer function; resonance frequency) | .29* |

Table 1. Correlation of SER Features and fatigue; * p < .05.

| NLD Features | r |
|---|------|
| skewness of vector length within state space = (sound pressure, Δ sound pressure, $\Delta\Delta$ sound pressure) | 21* |
| skewness of vector length within reconstructed phase space = $(\tau = 1, d = 3)$ | 22* |
| mean of Cao's minimum embedding dimensions = (τ = 1, d _{max} = 3, number nearest neighbors = 3) | .25* |
| mean of Cao's minimum embedding dimensions = (τ = 8, d _{max} = 4, number nearest neighbors = 3); | .25* |
| mean of Cao's minimum embedding dimensions = (τ = 10, d _{max} = 4, number nearest neighbors = 3) | .26* |
| mean of Cao's minimum embedding dimensions ($\tau = 8$, $d_{max} = 5$, number nearest neighbors = 3); | .27* |
| mean of Cao's minimum embedding dimensions ($\tau = 10$, $d_{max} = 5$, number nearest neighbors = 3). | .29* |

Table 2. Correlation of NLD Features and fatigue; * p < .05.

DISCUSSION

Due to several nonlinear phenomena producing turbulent air flow applying NLD speech feature might provide additional information regarding the dynamics and structure of sleepy speech comparing to the commonly applied SER feature set. We achieved significant correlations between SSS and the NLD feature 'Cao's minimum embedding dimension' of .29, which indicates a higher complexity of fatigue speech samples. Explanations for this effect might be found in the following effect chain: fatigue induced decreased body temperature \rightarrow reduced heat conduction within the vocal tract, changed friction between vocal tract walls and air, changed laminar flows, jet streams, and turbulences \rightarrow higher complexity of a more turbulent speech signal.

Several factors might have influenced the results obtained by NLD methods and consequently have to be considered, e.g. recording duration, degree of stationarity, and superimposed noise. Furthermore, it would seem advisable that future studies address the main topics of enriching the NLD feature set with further fractal (multifractal analysis, power-law correlation, detrended fluctuation analysis), entropy (approximate entropy/sample entropy, multiscale entropy, compression entropy), symbolic dynamics measures, and delay-vector-variance [1,27]. Additionally, it would seem beneficial that future studies address the main topics of enriching the steering feature set with easy accessible driving related informations as e.g. pedal movement behaviour. Finally, it has to be emphasized that up to now a reference measure of fatigue (gold standard of fatigue) is lacking. But approximating fatigue by concordance of self and several observer ratings might act as an intermediate solution, that follows the general procedure of finding ground truth values within the speech emotion recognition research community. In conclusion, methods derived from NLD could offer promising insights into fatigue induced speech changes. They supply

additional information and complement traditional timeand frequency-domain analyses of speech.

- Sommer, D. Analyse des Mikroschlafs mit Methoden der Computergestützten Intelligenz. Aachen: Shaker, 2009.
- 2. Sommer D, Golz M, Schupp T, Krajewski J, Trutschel U, & Edwards D A measure of strong driver fatigue. *Proceeding International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 4* (2009) 9-15.
- 3. Schnupp T, Schenka A, Edwards D, Krajewski J, & Golz M Is posturography a candidate for a vigilance test? World Congress on Medical Physics and Biomedical Engineering, *IFMBE Proceedings*, 25, 9 (2009), 388-392.
- 4. Krajewski J, Sommer D, Trutschel U, Edwards D, & Golz M. Steering wheel behavior based estimating of fatigue. *Proceeding International Driving Symposium* on Human Factors in Driver Assessment, Training and Vehicle Design, 4 (2009), 118-124.
- 5. Altmüller T Driver Monitoring and Drowsiness Detection by Steering Signal Analysis. Aachen: Shaker, 2007.
- Teager, HM & Teager, SM Evidence for nonlinear sound production mechanisms in the vocal tract. Speech Production and Speech Modelling, W. J. Hardcastle and A. Marchal, Eds., vol. 55 of NATO Advanced Study Institute Series D, Bonas, France (1989), 241–261.
- Thomas, TJ A finite element model of fluid flow in the vocal tract. *Comput. Speech & Language*, 1, (1986), 131–151.
- Kaiser, J.F. Some observations on vocal tract operation from a fluid flow point of view. In I. R. Titze and R. C. Scherer, editors, Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control, Denver Center for Performing Arts, Denver, CO (1983), 358–386.
- Banbrook, M., McLaughlin, S, & Mann, I Speech characterization and synthesis by nonlinear methods. IEEE Trans. Speech Audio Processing, 7 (1999), 1–17.
- Dimitriadis, D. & Maragos, P. Robust energy demodulation based on continuous models with application to speech recognition. In Proc. Eurospeech-03, Geneva, Switzerland, 2003.
- 11. Dimitriadis, D, Katsamanis, N, Maragos, P, Papandreou, G & Pitsikalis, V Towards automatic speech recognition in adverse environments. Proceedings of 7th Hellenic European Conference on Research on Computer Mathematics and Its Applications (HERCMA '05), Athens, Greece, 2005.
- Maragos, P, Kaiser, JF, & Quatieri, TF Energy separation in signal modulations with application to speech analysis. *IEEE Trans.on Signal Processing*, 41, 10 (1993), 3024–3051.

- Narayanan, S. & Alwan, A. A nonlinear dynamical systems analysis of fricative consonants. J. Acoust. Soc. Am., 97, 4 (1995), 2511–2524.
- McLaughlin, S & Maragos, P Nonlinear methods for speech analysis and synthesis. In Marshall S, Sicuranza G. Hindawi (Eds.) Advances in nonlinear signal and image processing, Volume 6, 2007.
- Maragos, P & Potamianos, A Fractal dimensions of speech sounds: Computation and application to automatic speech recognition. J. Acoust. Soc. Am., 105, 3 (1999), 1925–1932.
- Maragos, P. Fractal aspects of speech signals: Dimension and interpolation. Proc. IEEE ICASSP-91, 1991.
- 17. Ashkenazy, Y The use of generalized information dimension in measuring fractal dimension of time series. *Physica A*, 271 3-4 (1999), 427–447.
- Adeyemi, O & Boudreaux-Bartels, FG Improved accuracy in the singularity spectrum of multifractal chaotic time series. In Proc. IEEE, ICASSP-97, Munich, Germany, 1997.
- 19. Ashkenazy, Y The use of generalized information dimension in measuring fractal dimension of time series. *Physica A*, 271, 3-4 (1999), 427–447.
- 20. Kubin, G Synthesis and Coding of Continuous Speech with the Nonlinear Oscillator Model. In Proc. ICASSP'96, vol. 1, Atlanta, USA, 1996, 267.
- Quatieri, TF, & Hofstetter, EM Short-time signal representation by nonlinear difference equations. In Proc. IEEE, ICASSP'90, Albuquerque, 1990.
- 22. Townshend, B Nonlinear prediction of speech signals. IEEE Trans. Acoust., Speech, Signal Processing, 1990.
- Krajewski J, Batliner A, & Golz M Acoustic sleepiness detection – Framework and validation of a speech adapted pattern recognition approach. *Behavior Research Methods*, 41 (2009), 795-804.
- 24. Krajewski J, Batliner A, & Wieland R Multiple classifier applied on predicting microsleep from speech. In ICPR (Eds.), 19th Conference on Pattern Recognition (4 pages, no pagination). Tampa/ Florida: IEEE Computer Society Press, 2009.
- 25. Batliner, A, Seppi, D, Steidl, S & Schuller, B Segmenting into adequate units for automatic recognition of emotion-related episodes: a speech-based approach. *Advances in Human-Computer Interaction* (in press).
- Banbrook, M, McLaughlin, S. & Mann, I Speech characterization and synthesis by nonlinear methods. *IEEE Trans. Speech Audio Processing*, 7, (1999), 1–17.
- 27. Voss A, Schulz S, Schroeder R, Baumert M, & Caminal P Methods derived from nonlinear dynamics for analysing heart rate variability. *Philos Transact A Math Phys Eng Sci.*, 367, 1887 (2009), 277-296.

Detecting Fatigue from Steering Behaviour Applying Continuous Wavelet Transform

Jarek Krajewski¹ krajewsk@uni-wuppertal.de Martin Golz² m.golz@fh-sm.de Sebastian Schnieder¹ s.schnieder@uni-wuppertal.de

Thomas Schnupp² t.schnupp@fh-sm.de **Christian Heinze**² c.heinze@fh-sm.de

David Sommer² d.sommer@fh-sm.de

¹University Wuppertal Experimental Industrial Psychology Gaußstraße 20, 42097 Wuppertal, Germany

ABSTRACT

The aim of this paper is to develop signal processing based method to measure fatigue from motor behaviour. The advantages of this steering wheel movement approach are that obtaining steering data within driving is robust, non obtrusive, free from sensor application and calibration efforts. Applying methods of continuous wavelet transform (CWT) provides additional information regarding the dynamics and structure of steering behavior comparing to the commonly applied spectral Fourier transform features.

Author Keywords

Motor behavior, fatigue detection, wavelet transformation, pattern recognition.

ACM Classification Keywords

H.5.m Human-Computer-Interaction

INTRODUCTION

Some biosignals contain more information about fatigue than others. Among them, EEG (reflecting cortical activity) and EOG (reflecting eye and eyelid movement dynamics) have the potential to be used as a laboratory reference standard and have been utilized e.g. in evaluating fatigue monitoring technologies. Even so, the requirements for these electrode based instruments are not conducive for being used out in the field. Developing fatigue monitoring devices, which are cheap, non-intrusive, and robust even under extreme demanding environmental conditions (e.g. high background noise, temperature, or humidity) still remains a challenging task [1]. In contrast to these electrode-based instruments, using steering wheel

 ²University of Applied Sciences Schmalkalden Neuroinformatics and Signal Processing
 Blechhammer, 98574 Schmalkalden, Germany

movement as an indicator for fatigue is more robust under these same operating conditions. Collecting this data is favourable for fatigue detection since it is non-obtrusive and uses cheap, durable, and maintenance free sensors that are already integrated into the steering wheel system [2-3]. However, little empirical research has been done to examine thoroughly the benefits of more recently feature extraction methods as e.g. the wavelet analysis [4].

The traditional technique in signal analysis for tracking frequencies as they change over time is the Short-time Fourier Transform (STFT). Using sliding frame segmentation and a subsequently windowing of the signal can provide an acceptable time-frequency representation (Weighted Overlapped Segment Averaging; WOSA). In contrast to the Fourier transform (FT), the continuous wavelet transform (CWT) possesses the ability to construct a time-frequency representation of a signal that offers very good time and frequency localization. A CWT is used to divide a continuous-time function into wavelets. Since CWT is very resistant to noisy signals and has the ability to decompose complex information and patterns into elementary forms, it is commonly used in acoustics processing, biosignal analysis, and business information analysis. Hence, the aim of this study is to compare multiple state-of-the-art pattern recognition methods [5-7] on CWT vs FT based steering features to detect fatigue.

METHOD

Twelve healthy young adults completed 7 overnight driving sessions (1 - 8 a.m.) in our real car driving simulation lab. The combinations of observed and self-rated fatigue (Karolinska Sleepiness Scale; KSS) measured verbally every 2 minutes were considered as ground truth of fatigue. The fixed-based driving simulator consisted of a normal passenger city car (GM Opel Corsa) with original controls. The driving task involved a 40-minute night drive on a monotonous two-lane motorway course with simulated effects of headlights, i.e. involving a restricted range of sight. One round on the course with a speed of about 100

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Monitoring and control units of the fixed-based driving simulator consisting of a passenger city car (GM Opel Corsa) with original controls.

km/h took about 15 minutes. The course was mainly straight with sustained long curves. There were no obstacles or oncoming traffic.

Steering angle as well as other driving performance signal and biosignals were recorded. Here we analyzed only the raw steering angle time series, i.e. without any further event specific selection as e.g. steering away from obstacles or steering back to a suggested driving path.

Feature Extraction. To extract relevant features from the steering behavior data we split the data into 714 nonoverlapping 4 min segments (4800 measurement points). We computed two spectral feature sets using Fouriertransform (FT) and continuous wavelet transform (CWT) to capture fatigue impaired steering patterns. To measure fine temporal changes of spectral descriptors we performed a signal segmentation (sampling rate = 20 Hz, segment length = 25.6 s (512 data points), segment shift = 12.3 s (256 data points), hanning window), and computed the power spectral density per segment (Welch-Method; Weighted Overlapped Segment Averaging; WOSA). This procedure results in 975 frequency domain features as relative and absolute power spectral densities (PSD) of raw and first derivate contours in 30 spectral bands (e.g. minimum of relative PSD of first derivate in the 0.6-0.7Hz spectral band), band energy ratios, (e.g. PSD_{1-5Hz} / PSD_{0-1Hz}), spectral flux (e.g. max of spectral flux = Euclidean distance of PSD between consecutive segments), and long term average spectrum descriptors (e.g. skewness of PSD distribution). Within the wavelet domain we computed the equivalent of these features applying the continuous wavelet transform (CWT; 'db4' wavelet, scaling factor a = [1:1:512]) resulting again in 975 features.

| Classifier | Specification | FT | CWT | Δ |
|------------|--------------------------------------|------|------|-------|
| SVM | 'JmySVM', linear kernel, C=0 | 1.06 | 0.99 | -0.07 |
| | 'JmySVM', polyn. kernel, d=2, C=0 | 1.40 | 1.25 | -0.15 |
| | 'JmySVM', radial kernel, C=0 | 1.19 | 0.90 | -0.29 |
| | 'SMO-Reg', linear kernel, C=0 | 1.10 | 0.95 | -0.15 |
| NN | k=1 | 1.37 | 1.15 | -0.22 |
| | k=3 | 1.13 | 0.97 | -0.16 |
| | k=5 | 1.10 | 0.94 | -0.16 |
| | k=7 | 1.09 | 0.93 | -0.16 |
| | k=9 | 1.08 | 0.90 | -0.18 |
| | k=11 | 1.09 | 0.90 | -0.19 |
| MLP | 'NeuralNetImproved', 2 x | | | |
| | 5 nodes | 1.65 | 0.99 | -0.66 |
| | 'W-RBF-Network' | 1.15 | 1.05 | -0.1 |
| LR | 'W-LinearRegression' | 1.12 | 0.99 | -0.13 |
| | 'IsotonicRegression' | 1.07 | 0.83 | -0.24 |
| GPL | 'GaussianProcess' | 1.16 | 0.98 | -0.18 |
| Average | | | | |
| Learner | | 1.18 | 0.98 | -0.20 |

Table 1. Mean absolute prediction error (in KSS units) of several learning methods on the test set using driverdependent validation schemes. FT: Fourier Transform feature set; CWT: Continous Wavelet Transform feature set.

Machine Learning. We investigated validation of the classification algorithms to examine which automatically trained model can be used to recognize the fatigue of participants best. We applied the following static classifiers - known to be successful within many biosignal based classification tasks - of the popular 4.5 RapidMiner [8] software using standard parameter settings: Support Vector Machines ('JMySVM', linear, radial, or dot kernel function; 'SMO-Reg', Sequential Minimal Optimization), Multilayer Perceptrons ('NeuralNetImproved', 2 hidden sigmoid layer, 5 nodes each), k-Nearest Neighbors ('NearestNeighbors'; k = 1,3,5), Logistic Base ('W-LogisticBase'), Linear Regression ('LinRegression', 'W-LinRegression'), Polynomial Regression ('PolynomialRegression'), and Gaussian Process Learner ('GPLeaner'). In a driverdependent validation protocol, we applied 10-fold cross validation. The final prediction errors were calculated averaging over all 10 validations.



Figure 2. Spectrogram of alert (left) vs. fatigue (right) steering sample based on FT. High Power spectral densities are indicated by the red colour.

RESULTS

The mean absolute prediction error (MAP) of the KSS fatigue value was computed (see Table 1). Within the applied learning algorithms the Isotonic Regression using the CWT feature set reached the lowest MAP of 0.83. For the FT feature set, the Isotonic Regression learner achieved the second best prediction. A Support Vector Machine (linear kernel) reached the lowest prediction error of MAP = 1.06. Averaging all applied learning schemes the CWT feature set achieved 0.98, the FT feature set 1.18 MAP, which document a noteable and statistically significant benefit of 0.20 MAP (-16.9% error reduction) when using the CWT vs. FT feature set.

DISCUSSION

The aim of this paper is to the test the added value of a wavelet based feature set for (automatic) fatigue detection. Wavelet based features have not been used before for driver fatigue detection. Using the CWT based steering behavior features we achieved over all applied regression learner an mean MAP of 0.98 in terms of the KSS clearly outperforming the FT feature set (MAP = 1.18), which corresponds to an error reduction of 16.9%.

A secondary aim of this study was to compare the performance of different pattern recognition methods for learning task with nearly the same amount of instances and features. Within the FT feature set the best learner yielded a MAP of 1.06 (Support Vector Machine, linear kernel), which was again clearly outperformed by the best learner

within the CWT feature set (MAP = 0.83; Isotonic Regression). Nevertheless, the question whether this size of error is acceptable for real life application still remains open.

Our results are limited by several facts. We did not primarily aim at finding the best single regression learner and thus relinquish to optimize the performance of SVM by a fine grained hyperparameter optimization. Therefore, the performance has to be recognized as lower border of the SVM capabilities. Moreover, the present results should be replicated using laboratory gold-standard fatigue measures [9], and enlarged simulator and real-life databases. A further performance gain might be realized by advanced preprocessing steps as the decomposition of the steering signal using Independent Component Analysis (ICA) [10], Empirical Mode Decomposition (EMD) [11] or waveletbased decomposition in approximation and detail signals [4]. Referring to the fatigue reference KSS, it has to be mentioned that up to now a gold standard measure of fatigue is lacking. But approximating fatigue by concordance of self and observer ratings might act as an intermediate solution. Furthermore, it would seem advisable that future studies address the main topics of enriching the steering feature set. Figure 2 provides a first insight in further possible fatigue sensitive spectral features. Finally, easy accessible driving related informations as e.g pedal movement behaviour should be considered as source for fatigue monitoring.

REFERENCES

- 1. Barr L, Howarth H, Popkin S, Carroll RJ An Evaluation of Emerging Driver Fatigue Detection Measures and Technologies: Final Report, 2010.
- Sommer D, Golz M, Schupp T, Krajewski J, Trutschel U, & Edwards D A measure of strong driver fatigue. Proceeding International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 4 (2009), 9-15.
- 3. Schnupp T, Schenka A, Edwards D, Krajewski J, & Golz M Is posturography a candidate for a vigilance test? World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings, *25*, 9 (2009), 388-392.
- 4. Krajewski J, Sommer D, Trutschel U, Edwards D, & Golz M Steering wheel behavior based estimating of fatigue. Proceeding International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, *4* (2009), 118-124.
- 5. Altmüller T Driver Monitoring and Drowsiness Detection by Steering Signal Analysis. Aachen: Shaker, 2007.
- Teager, HM & Teager, SM Evidence for nonlinear sound production mechanisms in the vocal tract. Speech Production and Speech Modelling, W. J. Hardcastle and A. Marchal, Eds., vol. 55 of NATO Advanced Study Institute Series D, Bonas, France (1989), 241–261.
- Thomas, TJ A finite element model of fluid flow in the vocal tract. *Comput. Speech & Language*, 1 (1986), 131–151.
- Kaiser, J.F. Some observations on vocal tract operation from a fluid flow point of view. In I. R. Titze and R. C. Scherer, editors, Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control, Denver Center for Performing Arts, Denver, CO (1983), 358–386.
- Banbrook, M., McLaughlin, S, & Mann, I Speech characterization and synthesis by nonlinear methods. *IEEE Trans. Speech Audio Processing*, 7 (1999), 1–17.
- Dimitriadis, D. & Maragos, P. Robust energy demodulation based on continuous models with application to speech recognition. In Proc. Eurospeech-03, Geneva, Switzerland, 2003.
- 11. Dimitriadis, D, Katsamanis, N, Maragos, P, Papandreou, G & Pitsikalis, V Towards automatic speech recognition in adverse environments. Proceedings of 7th Hellenic European Conference on Research on Computer Mathematics and Its Applications (HERCMA '05), Athens, Greece, 2005.
- Maragos, P, Kaiser, JF, & Quatieri, TF Energy separation in signal modulations with application to speech analysis. *IEEE Trans.on Signal Processing*, 41, 10 (1993), 3024–3051.
- 13. Narayanan, S. & Alwan, A. A nonlinear dynamical

systems analysis of fricative consonants. J. Acoust. Soc. Am., 97, 4 (1995), 2511–2524.

- McLaughlin, S & Maragos, P Nonlinear methods for speech analysis and synthesis. In Marshall S, Sicuranza G. Hindawi (Eds.) Advances in nonlinear signal and image processing, Volume 6, 2007.
- Maragos, P & Potamianos, A Fractal dimensions of speech sounds: Computation and application to automatic speech recognition. J. Acoust. Soc. Am., 105, 3 (1999), 1925–1932.
- 16. Maragos, P. Fractal aspects of speech signals: Dimension and interpolation. Proc. IEEE, ICASSP-91, 1991.
- 17. Ashkenazy, Y The use of generalized information dimension in measuring fractal dimension of time series. *Physica A*, 271, 3-4 (1999), 427–447.
- Adeyemi, O & Boudreaux-Bartels, FG Improved accuracy in the singularity spectrum of multifractal chaotic time series. In Proc. IEEE, ICASSP-97, Munich, Germany, 1997.
- 19. Ashkenazy, Y The use of generalized information dimension in measuring fractal dimension of time series. *Physica A*, 271, 3-4 (1999), 427–447.
- 20. Kubin, G Synthesis and Coding of Continuous Speech with the Nonlinear Oscillator Model. In Proc. ICASSP'96, vol. 1, Atlanta, USA (1996), 267.
- 21. Quatieri, TF, & Hofstetter, EM Short-time signal representation by nonlinear difference equations. In Proc. IEEE, ICASSP'90, Albuquerque, 1990.
- 22. Townshend, B Nonlinear prediction of speech signals. IEEE Trans. Acoust., Speech, Signal Processing, 1990.
- 23. Krajewski J, Batliner A, & Golz M Acoustic sleepiness detection Framework and validation of a speech adapted pattern recognition approach. *Behavior Research Methods*, 41 (2009), 795-804.
- 24. Krajewski J, Batliner A, & Wieland R Multiple classifier applied on predicting microsleep from speech. In ICPR (Eds.), 19th Conference on Pattern Recognition (4 pages, no pagination). Tampa/ Florida: IEEE Computer Society Press, 2009.
- 25. Batliner, A, Seppi, D, Steidl, S & Schuller, B Segmenting into adequate units for automatic recognition of emotion-related episodes: a speech-based approach. Advances in Human-Computer Interaction (in press).
- 26. Banbrook, M, McLaughlin, S. & Mann, I Speech characterization and synthesis by nonlinear methods. *IEEE Trans. Speech Audio Processing*, 7 (1999), 1–17.
- 27. Voss A, Schulz S, Schroeder R, Baumert M, & Caminal P Methods derived from nonlinear dynamics for analysing heart rate variability. *Philos Transact A Math Phys Eng Sci.*, *367*, 1887 (2009), 277-296.

Creating a Richer Data Source for 3D Pedestrian Flow Simulations in Public Transport

Veronika Egger

is-design GmbH Westbahnstrasse 56-58, A-1070 Wien, Austria veronika.egger@isdesign.at

Lisa Ehrenstrasser

inklusiv design, Kajetan Schellmann-Gasse 3, A-2352 Gumpoldskirchen, Austria design@lisaehren.net

ABSTRACT

The distribution of passenger flows in public transport environments can be predicted through simulations of flow characteristics. The flow characteristics are usually based on average values, which do not include behavioural characteristics of people with mobility impairments such as wheelchair users, individuals with prams and small children, elderly people or people with sensory impairments. In the project at hand we gather data on orientation behaviour of these passenger groups through experience and task-related research. In a combination of methods - video and audio recording, user centered scenario and task development, questionnaires, interviews and time-motion-event (TME) recording - qualitative and quantitative data are generated. By linking qualitative information to quantitative measures a multi-dimensional description of the behaviour of individuals and user groups in a transport interchange emerges. It is expected that by comparing and contrasting the resulting behavioural measures to the average values currently used to represent passenger flows, the quality of simulations can be considerably enhanced.

Author Keywords

Usability of environments, behaviour in transport systems, orientation, navigation, wayfinding, pedestrian simulation, environmental design, interaction, decision making, sensory impairment, mobility impairment.

ACM Classification Keywords

G.3 Probability and Statistics: Experimental design. H.1.1 Systems and Information Theory: Value of Information, I.2.0 General: Cognitive simulation

Helmut Schrom-Feiertag

Austrian Institute of Technology Giefinggasse 2, A-1210 Wien, Austria helmut.schrom-feiertag@ait.ac.at

Gregory Telepak

Austrian Institute of Technology Giefinggasse 2, A-1210 Wien, Austria Gregory.Telepak.fl@ait.ac.at

INTRODUCTION

In the planning and adaptation of public transport infrastructures, the distribution of passenger flows can be inferred through flow simulations and, if needed, improved. The flow characteristics are usually based on walking speed, individual space requirement and purpose/intention of travel. However, these average values as incorporated in available simulation tools do not reflect behavioural characteristics of people with mobility impairments. Usually these groups are simply represented by reduced maximum speed and increased space requirement. These simulations also assume that all pedestrians know the infrastructure perfectly and consequently choose the shortest path to reach their goal. There are several phenomena and types of pedestrian behaviour that have a major influence on the overall performance of a transport infrastructure but they cannot be reproduced within the simplifications of traditional flow simulations. Ongoing research is looking to enhance existing simulation models with individual behaviour patterns [2,12,13] and also incorporates persons with reduced mobility [11]. In our approach [3] the cognition of guidance systems is modelled. The main challenge is the realistic adaptation of existing models by adequately representing and measuring group specific behaviour of mobility impaired passenger. We assume distributed attention/cognition as given in any task, when people are not dominated by one single task or focused on tasks involving only local information, but react to their surroundings and to the current context [6].

This extended abstract describes:

- Our general approach to gaining knowledge about behaviour patterns in public transport interchanges,
- The introduction of new combinations of methods for gathering real life user data.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

COMBINING METHODS OF MEASUREMENT

The scientific core challenge comprises the evaluation of group specific walking, orientation and navigation behaviour in relation to features in the built environment such as transport-related information, advertising, other wayfinding guidance, architectural features or contextaware information. A variety of empirical methods were combined in order to gain relevant qualitative and quantitative data on passenger behaviour:

- Video recordings at crucial points in the infrastructure (escalators, stairs), data collection through manual annotation.
- Online questionnaire on individual experience of pedestrian journeys between modes of transport.
- Questionnaire-based interviews with respondents on individual behaviour and experiences in public transport.
- Real-world scenario and task development [4].
- Controlled scenario-based experiments in real transport environments with defined tasks to be carried out by the respondents (user and task analysis).
- Documentation of the experiments employing three different techniques: "Thinking Aloud" as used in usability research, "Shadowing" of each respondent with time-motion tracking and observation/annotation of the performed task.
- Time-motion-event (TME) recording of interaction with the environment, based on multi-modal analysis of interaction [9].

SELECTION OF RESPONDENTS

Eight groups of people were identified who may demonstrate mobility patterns clearly distinguished from the general population: 70+ age group [7], people with pram/toddler, visually impaired, blind [5], wheelchair users, mobility impaired, hearing impaired and deaf [8]. All of them were capable of completing the tasks without assistance. Six people from each group were recruited and divided between two different scenarios (see below). The completion of the task took, on average, 20 minutes, depending on the respondent.

Based on extensive experience with a qualitative usability testing technique called "diagnostic testing" we know that it takes just a few respondents to highlight specific issues connected with the use of a product, information or the environment [14]. The resulting indicators provide the basis for research into particular group behaviour on a wider scale.

Expectations

We expect the behaviour differences to relate to walking speed, patterns of gaining information from the environment, orientation and navigation. The relative differences in behaviour patterns within the groups, between the groups and further the given values for the general population provide the required data.



Figure 1. Summary of methods and types of data generated in the process.

DESCRIPTION OF METHODS AND PROCESS

Figure 1 gives an overview of the methods employed, which are briefly described below.

Video Recordings

Video recordings are made at critical points inside the infrastructure in order to quantify effects on pedestrian flow. Each individual person is manually annotated in the crowd, to obtain speeds, trajectories, collision avoidance behaviour, queuing and counts. This data is used to calibrate the operational part of the microscopic simulation model.

Questionnaires

An online questionnaire is set up with open questions on individual experiences as pedestrians in transport interchanges, namely their reactions to obstructions and not hearing/seeing information. Once the respondents have been defined and recruited, they have to answer a set of questions before experiments can commence. Firstly, we need to determine how their disability influences their use of public transport in order to then place them in an appropriate scenario (see below).

Scenarios and Tasks

Two different scenarios with different levels of complexity reflect typical use patterns in the transport interchange selected for this project. The scenarios contain specific user tasks such as buying a drink for the journey, locating timetable information, or using the restrooms in the station.

Experiments

At a pre-defined meeting point the respondent is met by two researchers who "shadow" the respondent: one records the time-motion information using a special application on a tablet PC [1]. The other observer takes qualitative notes throughout the experiment. The respondent is given an audio recording device which requires no operation on the part of the respondent. The researchers follow the respondent without interfering in the process until the task is complete. They only intervene in exceptional circumstances, for instance if someone completely loses their way or forgets what to do next.

As soon as the task is completed, both audio recording and time-motion tracker are stopped. Further observations and comments are recorded in a concluding interview.

TYPES OF DATA GENERATED

In our research we distinguish quantitative and qualitative data. As described above (see also Figure 1), the data is derived from video and audio recordings, notes from observation and interviews, trajectories, speeds, detected stops and events recorded as part of the shadowing as well as transport and demographic statistics from questionnaires.

Categories of Analysis

In preparation for analysis two sets of categories are defined: quantitative and qualitative.

Quantitative categories are objectively measurable data like walking speed, route choice, time of hesitation, decision point (position and context), event relative to position, and preferences (position and context).

Qualitative categories are applied to each objective measure. They are: conscious awareness of guidance information, signage and the built environment; reasons for choices and hesitations; distributed attention in context.

ANALYSIS

The aim is to understand how people move in space and make use of their surroundings. Our analysis is based upon the framework of multimodal analysis of [9,10,15] which state that next to the modes of language (spoken language, written language etc), analyses of complex situations have to include gaze, head and arm movement, body configuration, the interplay of people and spatial environment. We will describe patterns of action and journeys we observed for each category, analyzing selected scenes, with a view to the interplay between the different modes or semiotic resources.

Time-Motion-Event (TME) Analysis

In order to easily compare behaviour during tasks between different respondents and groups of respondents, the scenarios are sliced into "scenes", which can be condensed to one page per scene per respondent (Figure 2). Based on the TME recording, a visual representation is created which Scenario 1, Scene 1 (U2 -> escalator/stairs/lift)



Figure 2. Template for comparative analysis of time-motion event recording showing one person's results within one "scene".



Figure 3. Single result layers of time-motion-event recording.

combines five of six data streams (Figure 3): The floor plan of the environment, the trajectory, detected stops, annotated events and thinking aloud. The TME datasets will be analyzed to extract speed and stop histograms of each trajectory. These histograms will subsequently be used to find clusters of individuals showing the same spatiotemporal behaviour. The spatial analysis of time-motionevent data will reveal the main routes, strong deviations from these routes and clusters where people mainly stop.

Connecting Types of Data

Transcripts of thinking aloud are applied to defined events within the trajectories. Observers' comments are added as a narrative. Demographic and transport statistics are used to define the user groups and bring results into context.

INITIAL OBSERVATIONS FROM EXPERIMENTS

After the first set of experiments we can observe several issues relating to the methodology:

As may be expected, some respondents do not behave naturally in the test situation.

Most respondents do not comment on any awkwardness of being observed and state that after getting involved in the scenario tasks, they had forgotten about the observers.

The "thinking aloud" method may have greater impact: Even though the recording instrument itself is barely noticeable, the fact that people have to talk (and think about what they say) distracts them. In some cases the distraction is such that these cases are of limited use for the calibration of simulation models. Nevertheless, they provide qualitative insight.

Most severe loss of orientation or lack of basic understanding of the scenario tasks occurs with respondents over the age of 70.

One assumption that can be discarded already is that people with sensory impairments would have greater problems carrying out the scenario tasks. Many completed the tasks in very little time or with minimal deviation from the most direct route, sometimes in contrast to test persons without sensory impairments but facing other challenges (parents with prams, elderly).

RESEARCH OUTLOOK

So far the experiments have revealed on the one hand surprising similarities between very different groups of people, on the other hand extreme deviations within a single group, indicating that in some cases determinants other than disability play a more significant role in navigation/ orientation behaviour. First analysis of some field tests suggests that the link between thinking aloud data and timemotion tracking will help to identify elements of the orientation system and the built environment that respondents use to navigate and thus point out typical areas for orientation in the infrastructure. This explorative research will lead to first steps in describing the behavioural patterns and variations within and between the described groups and may point towards further research necessary to enhance the understanding of environmental parameters and their influence on navigation. In addition it may be possible to generate hypotheses on why certain behavioural patterns emerge recurrently in determined areas. The insights gained will provide a basis for simulation model calibration on a tactical level.

- Bauer, D., Brändle, N., Seer, S., Ray, M. and Kitazawa, K. Measurement of pedestrian movements – a comparative study on various existing systems. In *Pedestrian behaviour* (2009), 301-320.
- 2. Braun, A., Musse, S.R., Oliveira, L.P.L.D. and Bodmann, B.E.J. *Modeling individual behaviors in crowd simulation*. In *CASA*, IEEE Computer Society, 2003.
- 3. Brunnhuber, M., Schrom-Feiertag, H., Hesina, G., Bauer, D. and Purgathofer, W. Simulation and visualization of the behavior of handicapped people in virtually reconstructed public buildings. In REAL CORP, 2010.
- 4. Carroll, M. Scenario-Based Design: *Envisioning Work* and *Technology in System Development*. John Wiley & Sons, 1995.
- 5. Corcoran, C., Douglas, G., McCall, S., McLinden, M. and Pavey, S. Network 1000: surveying the changing needs and lifestyles of 1000 visually impaired people – indicative results from generative interviews. *International Congress Series* (2005), 370-374.
- Hollan, J., Hutchins E. and Kirsh D. Distributed cognition: toward a new foundation for human-computer interaction research; TOCHI Volume 7, Issue 2, ACM 2000, 174-196.
- 7. Mollenkopf, H. Ageing and Outdoor Mobility. IOS Press, 2004.
- 8. Neuhold, H. Hörbehinderte und gehörlose Fahrgäste im öffentlichen Verkehr, 2004.
- 9. Norris, S. Analyzing Multimodal Interaction A methodological framework. Routledge, Taylor & Francis (2000), Chapter 2.
- O'Halloran, K.L. Systemic functional-multimodal discourse analysis (SF-MDA): constructing ideational meaning using language and visual imagery. Visual Communication 7, 4 (2008), 443-475.
- 11. Pearce, L., Powell, P., Duff, R., and Kerr, A. *Passengers* with reduced mobility: modelling their behaviour in underground stations. In European Transport Conference, 2008.
- 12. Pelechano, N., Allbeck, J.M., and Badler, N.I. Controlling individual agents in high-density crowd simulation. In SIGGRAPH/Eurographics, 2007, 99-108.
- 13. Shao, W. and Terzopoulos, D. Autonomous pedestrians. Graphical Models, 69, 5-6 (2007), 246-274.
- 14.Sless, D., Wiseman, R., Writing About Medicines for People: Usability Guidelines for Consumer Medicine Information, Communication Research Press, 1996
- 15. Van Leeuwen, T. and Jewitt, C. *Handbook of Visual Analysis*, Sage Publications, 2002.

Identifying Operators Monitoring Approviately

Dietrich Grasshoff dietrich.grasshoff@dlr.de **Catrin Hasse** catrin.hasse@dlr.de **Hinnerk Eißfeldt** hinnerk.eissfeldt@dlr.de Carmen Bruder Carmen.Bruder@dlr.de

DLR - German Aerospace Center, Hamburg, Germany

ABSTRACT

The objective of the presented study was to identify monitoring behaviour parameters for the purpose of prediction of manual control performance in the area of aviation. In other words: we want to distinguish between good and bad aviation operators. For this reason we developed an air traffic flow simulation tool (SSAS: Self Separation Air Space Simulation) that represents future tasks of pilots and controllers. 90 applicants of the Deutsche Lufthansa AG (DLH) und the Deutsche Flugsicherung GmbH (DFS) performed scenarios of different complexity. Preliminary to the manual control of each scenario, the applicants monitored the same scenario whereas the simulation was operating automatically. Eye movement data were collected while monitoring in the automatic part of the simulation. These data were connected to manual performance data. Results of our Study show, that in tasks of moderate complexity and difficulty, appropriate monitoring behaviour is related to efficient manual control afterwards.

Author Keywords

Monitoring behavior, air traffic flow simulation, prediction of manual control.

INTRODUCTION

Research project Aviator 2030 (see Fig. 1) focuses on an optimal fit between air traffic management (ATM) system design and human operators in future aviation. This will be carried out by adapting selection profiles to future ability requirements. In the first project phase, workshops with experienced pilots and air traffic controllers were conducted in order to develop a concept of future ATM. They were asked to tell their expectations regarding future tasks, roles and responsibilities. Summing up these workshop results, monitoring and teamwork in a highly automated workplace pose a challenge to future aircraft operators (Bruder, Jörn & Eißfeldt, 2008). Thus, research should focus on the ability of monitoring as one major topic. The second project

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

phase comprised the development of simulation tools that represent future workplaces in aviation. Experiments with humans operating in these simulated future workplaces serve as basis for identifying potential changes in ability requirements for pilots and air traffic controllers. Results allow for a timely adjustment of selection profiles and, thereby, for the development of future ability tests.

METHOD

Based on the empirical background (for a detailed explanation see Hasse et al., 2009) we devised a normative model, which describes the monitoring behaviour of operators monitoring appropriately (OMA). According to models of adequate and efficient monitoring behaviour (Niessen & Eyferth, 2001) as well as differences between experts and novices (Underwood et al., 2003) it can be stated that OMA show target oriented attention allocation in general as well as during monitoring phases (orientation – anticipation - operation - debriefing). Whereas the first assumption requires the operator to adapt attention allocation to the specific requirements of a given situation in general, the second assumption focuses the allocation of attention in phases. The operator is required to demonstrate flexibility in:

- anticipating system operations (during anticipation phases)
- detecting relevant system operations (during operating phases)
- controlling system performance afterwards (during debriefing phases)

We further assume that "good monitoring" is associated with an accurate manual system handling in case of automation failure, and therefore aim to connect monitoring behaviour with manual control behaviour. We assume that



Figure 1. Flowchart of the Aviator 2030 Project.

this link reflects differences in understanding of the underlying principles of the automatic system. The question we need to answer is: which monitoring criteria are most important with regard to identifying OMA that have the ability to manually control the system? In order to answer this we aim to test the following hypotheses:

Hypothesis 1: Keeping an overview of the overall automated system during an entire monitoring run is related to an accurate manual control of the system in case of necessity.

Hypothesis 2: Anticipating and detecting important automated system operations in time as well as controlling them is related to an accurate manual control in case of necessity.

Simulation Tool

With the objective of identifying monitoring parameters that have an impact on manual system handling, we developed a simulation tool that allows for the assessment of monitoring performance. The SSAS is a simplified air traffic flow simulation, where two tasks have to be performed – a traffic flow simulation and a simple flight control simulation. SSAS software is running on any WIN XP (minimum requirement: SP2) or WIN Vista pc configuration equipped with a DualCore processor. The graphical user interface (GUI) is designed for any screen sizes with resolutions varying from 800x600 to 1600x1200. Handling of the simulation will be carried out by mouse or touch screen.

Concerning the first task, operators have to manage the traffic flow between two airports – east and west - each consisting of an inbound and an outbound area (see Figure 2). At these areas values of the traffic load are depicted. The *target* value shows, how many aircraft should stay at this area. In the course of the simulation the target values remain unchanged. The *actual* value displays the actual number of aircraft staying in the corresponding area. The task of the operator is to bring all actual values into agreement with the target values of the corresponding areas as soon as possible. Additionally, there are exit and entry areas. These areas simulate extern air spaces which can't be controlled by the operator.

The outbound and inbound areas are connected by four different types of routes: air routes (between inbound and outbound of different air ports), exit routes (between inbound and an external air space), access routes (between an external air space and outbound) and service routes (between outbound and inbound of the same airport). Every route is a one-way road, the traffic flow is unidirectional, e.g. on air routes aircraft only move from the outbound area to the inbound area. Between the inbound and the outbound area of different airports there are two routes of different capacities and velocities. The upper route is capable of twice as much aircraft as the lower one. On the other hand,



Figure 2. Air traffic flow display of the SSAS.

aircraft moving on the lower route are two times faster than aircraft on the upper route.

Operators can handle and control the system via input devices (white buttons). Left-click on an input device will increase the number of announced aircraft (maximum: five aircraft), which will use the corresponding route in the next time units. A right click on the input device will diminish the number of annunciated aircraft (minimum: zero aircraft). Each time unit (configurable, in this simulation one time unit equals two seconds) announced aircraft will be released on the routes, at a frequency of one per route and time unit. When an aircraft is released on a route (green arrow), the number of announced aircraft at the corresponding input device is reduced by one. Some routes can't be controlled by the operator: the two access routes are feeding traffic into the simulation automatically. This feature assures, that an overall system work load can be maintained. Cleared aircraft move to the next segment of the route each time unit. When an aircraft arrives at the designated area on the other side of the route, the actual value of the corresponding area will be increased by one. As mentioned before, goal for the operator is to equate all actual values to the corresponding target values.

Sometimes aircraft are critical, i.e. they do not flight optimally in the airway (green frame of the aircraft symbol changes to a red frame). In this case, operators have to navigate the critical aircraft back on the optimal pathway as soon as possible. To perform this task, the operator must switch to the flight control screen (see Figure 3) by pressing the switch button.

On this screen a detail of the critical air route is displayed, consisting of some segments of the air route and the critical aircraft. Furthermore, performance indicators and adjustment devices for the vertical and horizontal



Figure 3. Flight control display of the SSAS.

performance of the aircraft are depicted. The actual performance indicators, displayed by small red and black bars, show the actual performance of the aircraft. A red bar indicates a need for performance improvement, a black bar symbolizes an acceptable performance. The target performance indicators, depicted as large green bars, mark the desired performance, whereas the target performance indicators are fixed. If a red bar appears, it has to be moved into the green bar. This can be done by simply clicking on the relevant adjustment devices on the screen. Every click on the device will correct the aircrafts pathway towards the target performance. When the target performance is reached, the operator can switch back to traffic flow simulation. In short, the operators' task is to control the traffic flow between two airports. The operator either monitors an automatic process or controls the traffic manually, allowing us to collect data on their performance of both types of task separately. With the objective of varying complexity and dynamics of the automatic system, we developed four scenarios reflecting different degrees of difficulty. The four scenarios were presented in a fixed order for every subject, beginning with the easiest (scenario 1), finishing with the most complex (scenario 4).

Measurements

As for dependent variables, we focus on the establishment and maintenance of system understanding during the monitoring phase. We use eye movement parameters, which act as indicators for the perceptual and cognitive operations involved. Monitoring performance was measured by recording eye movements. We used relative fixation counts and mean fixation durations based on predefined areas of interest. Fixation counts can be used as a measure of expectations and assumptions of the person (Rötting, 2001), where important objects are likely to be fixated upon more often than less important ones (Göbel, 1999). Fixation durations can be used as measure of information processing duration (Inhoff & Radach, 1998, S. 37, cit. in Rötting, 2001). Accordingly, processing difficulty as well as personal strain is reflected in the fixation duration (Rayner, 1982, Balota et al., 1985).

Our normative model postulates that OMA keep an overview of system operations during an entire monitoring run; in this experiment during the automatic mode of one scenario (hypothesis 1). Moreover, OMA are expected to anticipate, detect and control automated operations in time; in this experiment reflected by different operations performed by automation within the automatic mode of one scenario (hypothesis 2). As for testing the first hypothesis, we defined scenario specific areas of interest (AOIs), that is areas on the simulation screen that we expect to be preconditional for keeping an overview of system behaviour. As for testing the second hypothesis, we determined AOIs that help to anticipate and detect system operations as well as to debrief them. As anticipation, detection and debriefing of system operations are only possible within definite time frames within a scenario, we cut every scenario into time frames. Every time frame stands for a monitoring phase and is characterised by AOIs being conditional for monitoring adequately in this phase, e.g. anticipating a system operation adequately. Hence, this model leads us to expect eve movements on areas of interest that are generally relevant for a specific scenario as well as for monitoring phases within specific time frames. As we assume the understanding of the system to be conditional for manual system handling in case of system failure, we combine both, eye movement parameters and performance data, as measurements. Regarding the performance of a test subject during the manual phase of each scenario, we used the mean deviation of actual values from target values of inand outbounds. To avoid the possible impact of a general ability on manual performance when controlling a system, we corrected manual performance by deducting the baseline measurement when both parameters are significantly correlated. Eye Movements are recorded by Eyegaze Analysis System manufactured by L. C. T.. Managing of raw data was conducted by NYAN software, developed by Interactive Minds. Subjects were seated in front of a 19inch LCD computer display with a distance of approximately 60 cm.

Procedure

90 participants from DLH and DFS were tested individually. First, they were given a questionnaire measuring trust in automation, and the instruction for the following experiment. Participants were told that they would work on four scenarios, all consisting of two phases starting with an automation phase followed by a manual phase. Referring to the automation phase they were instructed to monitor the automation with the objective of understanding the rulebased dynamics of the given scenario. Referring to the hand control phase (manual condition), participants were assigned to manually control the system in continuation of the automation. That is, participants should control the system in terms of the rules and dynamics that they have learned from monitoring the scenario in automation. After a short (15 s) calibration phase that ensures adjustment of Eyegaze Analysis System to individual gazes of the participants, the persons were then presented the four scenarios, each taking 5 minutes. There was a smooth transition between the automatic mode and manual mode within each scenario but pauses were placed between each scenario. The four scenarios were presented in a fixed order for every subject beginning with the easiest, scenario 1, finishing with the most complex, scenario 4.

RESULTS

Our results show, that high performers look frequently at relevant areas to keep an overview as well as to detect and to recheck tasks in time. In addition, high performers look relatively long at relevant areas during orientation towards the scenario.

Overall, relative fixation counts on scenario and phase specific areas seemed to be an adequate parameter in order to identify OMA. The better subjects keep an overview of system operations (while monitoring automated processes), the better their ability to resume manual control. Furthermore, the better they are shifting their attention to timely relevant actions, the better their ability to resume control manually.

During orientation phases, gaze durations instead of fixations counts seem to be the appropriate parameter to predict the ability to resume manual control. The longer test subjects look at relevant areas, the better their manual performance in the manual phase.

During orientation phases the simulator screen is "frozen", so it would make sense to look persistently at relevant areas (while events remain static). On the other hand, it could be shown that frequently looking at relevant areas while events are changing dynamically is an appropriate monitoring behaviour, too.

Overall, it could be shown with SSAS that monitoring parameters have predictive power for system understanding and performance on the task. Results are dependent on difficulty of scenario and phase, with scenario 1 being too easy and scenario 4 being too difficult.

- 1. Balota, D. A., Pollatsek, A., & Rayner, K. The interaction of contextual constraints and parafoveal information in reading. *Cognitive Psychology*, *17* (1985), 364-390.
- Bruder, C., Jörn, L., & Eißfeldt, H. When pilots and airtraffic controllers discuss their future. In *Proceedings of the 28th EAAAP Conference: The future of aviation psychology*, Valencia: EAAP (2008), 354-358.
- Eißfeldt, H., Grasshoff, D., Hasse, C., Hörmann, H.-J., Schulze-Kissing, D., Stern, C., Wenzel, J., & Zierke, O. Aviator 2030: Ability Requirements in Future ATM Systems II: Simulations and Experiments. (Rep. No. 2009-29), Köln: DLR (2009).
- Göbel, M. Blickbewegungsregistrierung bei der Gestaltung von Fahrerarbeitsplätzen des ÖNVP. In M. Rötting & K. Seifert (Eds.), *Blickbewegung in der Mensch-Maschine-Systemtechnik*. Sinzheim: Pro Universitate Verlag (1999), 118-132.
- Inhoff, A.W., & Radach, R. Definition and computation od occumulotor measures in the study of cognition. In G. Underwood (Ed.), *Eye guidance in Reading and Scene Perception*. Amsterdam: Elsevier (1998), 29-53.
- 6. Niessen, C., & Eyferth, K. A model of the air -traffic controllers' picture. *Safety Science*, 73 (2001), 187-202.
- Rayner , K. Visual selection in reading, picture perception and visual search – A tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X*. Hillsdale: Lawrence Erlbaum (1982), 67-96.
- 8. Rötting, M.. Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen. Aachen: shaker (2001).
- 9. Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. Visuall attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, 46 (2003), 629-646.

A Detailed Analysis of Eating Behaviour

Ioannis Ioakeimidis Ioannis.Ioakimidis@ki.se Per Södersten Per.Sodersten@ki.se **Modjtaba Zandian** Modjtaba.zandian@ki.se

Karolinska Institute, NVS, Section of Applied Neuroendocrinology. Hälsovägen 7, Novum (Mandolean clinic), 141 57 Huddinge, Sweden

ABSTRACT

The eating behaviour of human female volunteers is expressed as bursts of intake and intervening periods of rests and the cumulative intake curve can be modeled by a quadratic equation. The curve is negatively accelerated in some women and linear in others. Hence, women can be divided into decelerated and linear eaters. Here, we have studied the eating behaviour of decelerated and linear female eaters in further detail through analysis of video recordings of the maxillary-mandibular region and recordings of the plates from which the women ate and related these to weight loss data from the plates during meals. While the amount of food eaten is similar, decelerated eaters take fewer and larger mouthfuls per meal than linear eaters and the frequency of mouthful uptake decreases towards the end of the meal. Both types of eaters chew with stable rhythm throughout the meal, with the decelerated eaters chewing faster.

Author Keywords

Eating behaviour, eating styles, mouthful characteristics, chewing rate.

INTRODUCTION

Rat Drinking and Human Eating Styles

Early studies provided a detailed description of drinking in rats as a function of prior water deprivation [1,2]. The pattern of drinking is the result of successive "intake burst and rests". Since the rate of water intake is stable during the bursts, the cumulative intake of water simply equals the time the rat spends drinking, minus the duration of the "rests" [1]. As the length of the periods of rest increases by the end of an intake period, the cumulative intake curve is decelerated; hence rats are decelerated drinkers. Similar to drinking in rats, eating in human subjects consists of bursts and rests [3,4].



Figure 1. The Mandometer[©], a personal computer connected with a scale recording the meal in real-time. Feedback to the user can be provided through the touch screen.

In the past we have developed an advanced mobile mealrecording device, the Mandometer[©], with the ability to provide live feedback to the user during the meal (see Figure 1) [5]. By analyzing repeated evaluation meals, we have indentified two patterns of eating behaviour among healthy human female volunteers, based on real time weight data, collected from a scale placed under the plate [5]. By fitting such data to a quadratic equation, $y=kx^2 + lx$, the cumulative food intake emerges and subjects can be divided into two types of eater; those who start eating with a high rate (initial l-value) which is reduced towards the end of the meal (k<0) (decelerated eaters), and those who maintain a stable rate of eating throughout the mea0) (klinear eaters) [5].

While decelerated eaters eat less food when the rate of eating is experimentally increased, linear eaters eat more food and while decelerated eaters maintain their intake at the normal level when eating rate is experimentally

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

decreased, linear eaters eat less food [5]. In responding to these challenges, linear eaters thus conform to the eating pattern of patients diagnosed with Binge Eating Disorder (BED) and Anorexia Nervosa (AN), respectively [6]. We have demonstrated that practicing eating at the proper rate by use of real time visual feedback during the meal is a useful intervention in treating anorexic [7] and obese [8] patients and suggested that linear, but not decelerated eaters are at risk of developing such disorders [5,6].

In the present study, we present an innovative, semiautomatic way to analyze meals in depth, combining data from video and weighting scales. We apply our improved methods to further analyze the eating behaviour differences between decelerated and linear eaters. Hence, we analyzed the chewing pattern and mouthful management in the two types of eaters, keeping in mind possible future improvements to our clinical practice [7].

SUBJECTS AND METHODS

Subjects and Meal Sessions

Sixteen healthy female volunteers, with no history of eating disorders, participated in the study. They were served the same kind of food (chicken and vegetable pieces; 426kJ, 10.67g protein, 8.03g carbohydrate and 2.5g fat/100g) in three lunch sessions at 12.00h separated by about a week. The meals were served in a secluded room, without windows. During the meals, the external stimuli were minimized, as reading and listening to music was not allowed. These were "free" meals, with *ad libitum* food and time, without any kind of feedback. The effect of hunger variations was minimized by asking the participants to stick to a specific breakfast type and quantity during the test days.

After preliminary testing, subjects were divided into linear and decelerated eaters, who differed significantly in k coefficient (-0.14 and -0,48 respectively, p < 0.02, from an expected range of 0 to -0.65). Individuals in the two groups were similar in age (23.7 vs 23.3 years) and body mass index (BMI, 21.7 vs 21.4 kg/m2) respectively.

Data Collection

Data obtained from the weight loss of a plate during the meal [7] were combined with video recordings of mouth movements during the meal. A video camera (DigitalCam, Samsung, Korea) was directed at the plate and another one at the subject's maxillary-mandibular region during the meal. The video recordings were manually time-stamped for the occurrence of food removal from the plate (spoonful), and for the occurrence of food entering the mouth (mouthful) and chewing action (any obvious lateral or vertical jaw movement). Spoonful and mouthful time stamping series were synchronized, using a custom made algorithm, eliminating discrepancies between the series, due to rare behaviors, e.g., double food uptake from the same spoonful and the final mouthful sequence of the meal was calculated.



Figure 2. Schematic representation of meal characteristics. A. Cumulative intake over time based on weight reduction on scale, B. Mouthful sequence timestamped from video feedback, C. Detail of chewing occurrences as timestamped from video feedback.

Data Meta-Analysis

We developed a semi-automated procedure for processing a big number of meals in a short period of time.

Initially, the start and the end of the meal and the points of food additions to the plate were manually marked in relation to the weight loss data series.

Errors were filtered automatically by a heuristic algorithm, developed to correct possible errors using a "best-guess" procedure. The multi-step method involved: Characterizing a measurement as correct or mistaken, deciding about the error type (force on plate, utensil on plate or delayed weight registration) and finally correcting the measurement depending on the type of error. Numerous correction options were used: using a null-change value, subtracting the weight of the utensil or generating different unique values, based upon changes in the weight up to 16 sec before and/or after the measurement.

The weight loss data were then synchronized with mouthful sequences in relation to the starting point of the meal and fitted to a quadratic equation. Hence, the detailed eating style of each individual had been determined (see Figure 2).

This method allows calculation of the cumulative meal characteristics, i.e., food intake, meal duration and number of mouthfuls and chewing incidents. The precise load of each mouthful, as well as the chewing frequency during each inter-mouthful interval, were then calculated and averaged over meal thirds. Results are reported as the mean values of the three meals, excluding measures of variability for simplicity. Group differences were evaluated by using t-tests and analyses of variance, followed by post hoc tests, as needed.

RESULTS

Food Intake, Meal Duration, Mouthfuls and Chewing.

Decelerated and linear eaters ate about the same amount of food (295 vs 289 g respectively), but decelerated eaters took their meal in a shorter period of time (8.4 min vs 11.3 min, p< 0.04) and had fewer mouthfuls (36.3 vs 48.9, p< 0.02). The number of chewing incidents didn't differ significantly between the groups (644 vs 740, p> 0.3).

Time Course of Mouthfuls and Chewing During the Meal.

While mouthfuls/min was similar among decelerated (4.9) and linear (4.8) eaters (p > 0.9) in the beginning of the meal, it decreased among decelerated eaters (3.9) during the last third of the meal and retained a stable rate among linear eaters (4.8) (p > 0.3). Additionally, decelerated eaters had significantly more food/mouthful than the linear eaters throughout the meal (group effect, p < 0.05). Still, both groups reduced their mouthful loads toward the end of the meal (p < 0.0001). Finally, decelerated eaters had a higher rate of chewing (chewing incidents/min) throughout the meal (group effect, p < 0.02), and both groups maintained their chewing rate over time (time effect, p > 0.3).

DISCUSSION

Eaters differ mainly in the rate at which food intake decelerates during the meal. It has been proposed that the lack of deceleration is a risk factor for the development of abnormal eating, which might lead to the development of long term eating disorders [5,6]. The method used here shows that this change reflects a reduction of the number of mouthfuls by the end of the meal. The load to mouthful ratio also decreased during the course of the meal, but this change was independent of the rate of deceleration of the meal. The rate of chewing remained constant during the meal, but was higher among decelerated than linear eaters, despite the fact that the load/mouthful decreased with time in both groups. The complete data analysis will include more detailed information about the course of chewing inside each individual mouthful, factoring in the duration of chewing pauses (evident in Figure 2C) between consecutive mouthfuls, together with the mouthful-specific load.

In our clinical practice, patients use decelerated cumulative intake curves, displayed on a computer monitor, as feedback when practicing eating (see Figure 1) [6,7]; the results presented here signify the existence of differences in the mouthful and chewing patterns of the two eater types, pointing to the need of additional research in that area together with the possible clinical significance of those characteristics.

- 1. Hill, J.H. and Stellar, E. An electronic drinkometer. *Science*, 2950 (1951), 43-44.
- Stellar, E. and Hill, J.H. The rat's rate of drinking as a function of water deprivation. J. Comp. Physiol. Psychol., 45 (1952), 96-102.
- 3. Jordan, A.H., Wieland, F.W., Zebley, P.S., Stellar, E. and Stunkard, J.A. Direct measurement of food intake in man: a method for the objective study of eating behavior. *Psychosom. Med.*, 28 (1966), 836-842.
- 4. Stellar, E. Hunger in man. American Psychologist, 22 (1967), 105-117.
- Zandian, M., Ioakimidis, I., Bergh, C., Brodin, U. and Södersten P. Decelerated and linear eaters: Effect on eating rate on food intake and satiety. *Physiol. Behav.*, 96 (2008), 270-275
- 6. Ioakimidis, I., Zandian, M., Bergh, C. and Södersten, P. A method for the control of eating rate: a potential intervention in eating disorders. *Behav. Res. Methods*, *41* (2009), 755-760.
- Bergh, C., Brodin, U., Lindberg, G. and Södersten, P. Randomized controlled trial of a treatment for anorexia and bulimia nervosa. *Proc. Natl. Acad. Sci. USA*, 99 (2002), 9486-9491.
- 8. Ford, A.L., Bergh, C., Södersten, P., Sabin, M.A., Hollinghurst, S., Hunt, L.P. and Shield, J.P. Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial. *BMJ*, *340* (2009), b5388.

Field Measurement Protocol for Team Communication: A Study of Medical Rehabilitation Team Interaction

Balázs Péter Hámornik hamornik@erg.bme.hu

Márta Juhász juhaszm@erg.bme.hu Ildikó Vén

Budapest University of Technology and Economics, Department of Ergonomics and Psychology Egry J. utca 1, Budapest H-1111, Hungary

National Institute for Medical Rehabilitation, Ward for Rehabilitation of Traumatic Injuries Szanatórium utca 19. 1528 Budapest, i.ven@rehabint.hu

ABSTRACT

The paper is introducing an online communication behavior measurement protocol, developed for medical team work, especially for studying the medical rehabilitation teams on interdisciplinary meetings. The purpose of the research is to analyze the development and usage of shared mental models (SMM) through the communication in medical practice of interdisciplinary rehabilitation teams in order to identify the factors determining the patient safety, and the effectiveness of rehabilitation. The online quantitative rating method uses two parallel raters to count and record the communication utterances of each participant directed to another. The quantitative and interactional data is analyzed, and visualized by AGNA network analysis software. Based on the results (network indices, and communication network diagram) the strong and weak points of communication can be recognized, as the important areas of SMM development. This can be a basis of a development process to enhance team effectiveness, and patient safety.

Author Keywords

Team communication, rehabilitation team, online rating, network analysis, patient safety, shared mental model.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Organizational design, H5.m. Miscellaneous

INTRODUCTION

The medical field is not as important subject of research as it should be, especially those aspects that consider a hospital, or ward as an organizational entity [4,6,7,10]. This is why our research project is aiming to study and explore the fields of development in interdisciplinary medical

rehabilitation teams with the tools of work psychology. The aim is to ameliorate the intra-team information flow about the patients and practices, to help the development of the shared knowledge and mental model in the team, and to contribute the faster and successful rehabilitation. The shared mental model is a shared knowledge and reference between the team mates. It develops by the team communication, and contains shared representations about the current situation, about roles and responsibilities, and about the competencies [3]. Thus the goal is not to change the rehabilitation practice, but to give an organizational aid to communication, knowledge and information sharing about the patient, the patients' conditions in order to make the development of the shared mental model faster and make the model itself more effective [9].

THE MEDICAL REHABILITATION TEAM

The medical rehabilitation field is an interdisciplinary field, because the injuries of patient usually exceed the competency of a single specialist, so thus several specialists, and different knowledge fields should cooperate in the rehabilitation team [6]. The medical rehabilitation team consists of (rehabilitation) physicians, clerk physicians, physiotherapists, nurses, occupational therapist, ergotherapist, social worker, clinical psychologist, prosthetist, and speech/language therapist, sport therapist, and psychiatrist [9].

Certain specialists are assigned to a specific ward (like the physicians, the nurses and the physiotherapists), and several other team members are centrally employed by the institute, and serving all the different wards (e.g. occupational therapist, ergotherapist, sport therapist) thus they participate in more than one rehabilitations teams in the same time. This causes that the members of the rehabilitation team are not always in the same location, and not always together, but the information is needed to be updated as fast as possible to guarantee the adequate treatment for the patients by all the specialists. To maintain the information sharing. there are oral and written occasions and tools, both on the formal and informal level. They can be shown in Table 1, organized by the modality, and the formal or informal manner of them.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, contact the Measuring please Behavior secretariat: info@measuringbehavior.org.

| Occasions of information sharing | | | |
|---|---|--|---|
| Written occasions | | Oral occasions | |
| Formal | Informal | Formal | Informal |
| Treatment documentation (e.g. medications, nursing, physiotherapy) Documentation of previous treatments in the past | Shift change notes of nurses Nurses' notes ordered treatments by the physician on a ward round (they are accessible for all team members) | Daily ward round Weekly ward round Interdisciplinary team meeting Prosthetist's ward round | Conversations on the corridors or in the rest rooms of the ward |

Table 1. The occasions of information sharing in the medical rehabilitation team [8].

From the perspective described above, the interdisciplinary team meeting is the most important field for our research, because on this meeting, the specialists of the ward and the specialist serving several wards in the institute are also participating, and discussing, sharing information with each other. The patients are not participating the team meeting usually, thus the communication is more open, and less formal between the specialists. This is why we choose the interdisciplinary team meeting to be the field of our research of communication presented in this paper.

THE METHOD

Our field research took place in a ward of the National Institute for Medical Rehabilitation, near to the capital of Hungary. We participated, and recorded four interdisciplinary team meetings (online) for quantitative communication analysis in January, 2010. The rehabilitation team consisted of fifteen specialists: two physicians, a clerk physician, two nurses, six physiotherapists, an occupational therapist, a social worker, a sport therapist, and an ergotherapist. The physicians, the nurses and the physiotherapists are linked to the ward studied by us, the other specialist are serving more than one wards. The method based on online rating using a rating table containing the sender and the receiver of a communication utterance during the team meeting. The important point in this strategy is the definition of the element of measurement, what has to be recorded as one utterance. Based on the literature of different subjects, but similar measurement method, we jointly defined a standard utterance by intonation contour and by the presence of a discernible pause between it and surrounding utterances [5]. This standard utterance mostly consists of a single phrase or sentence expressing a complete thought [2]. The two raters who participated the interdisciplinary team meetings have recorded the number of utterances spoken by the team mates including the sender and the receiver in order to build a transactional data to build a communicational network. In order to secure the reliability of our measurement of communication behavior, we used two parallel and independent raters. Both raters have been trained together to identify and record the utterances according to the

definition described above. Before the research started, they had the opportunity to practice, and compare recordings on previously taped team meetings from the same institution. In the analysis, we correlated their ratings (the recorded amount of utterances between the team mates), and to justify the inter-rater reliability, we used intra-class correlations also. After the control of the reliability, we summarized the utterances by team meetings (four summarized communication matrices), than we computed an average communication matrix from the four matrices. For the network analysis we summarized the utterances in the specialists' groups, thus we used the physicians, the nurses, the physiotherapists, the occupational therapist, the social worker, the sport therapist, and the ergotherapist as measured senders and receivers in the communication matrix. We used the AGNA social network analysis software [1] to visualize the average communication matrix of the recorded four interdisciplinary team meetings in a network diagram. And also we used the network measures computed by AGNA to interpret the network diagram. The AGNA computed density, closeness, and betweenness measures. The density of a network is the total number of edges divided by the number of all possible edges in that network. The closeness measure as a measure of centrality, which represents the average distance from a node to the other nodes in the network (the higher value means that the node is in a central position). The betweenness measure represents a position in the network, when the node is in a transmitter role between two subgroups.

Ethical Statement

The field research has been authorized by the Ethical Committee of the National Institute for Medical Rehabilitation as a part of a research project "The Development of Knowledge Sharing in Expert Medical Teams". All the participants and the patients are not identifiable in the data, the analyses and in the results. The members of the rehabilitation team agreed to audiotape the meetings for further analyses, and for the training of the raters.



Figure 1. The network diagram of the average communication of the interdisciplinary team meeting. The more shaded edges representing higher amount of utterances in communication.

RESULTS

We performed the statistical analyses with SPSS Statistics 18[®] predictive analytic software. The inter-rater reliability was assessed by the correlation of the two raters' records, the average Spearman's rank correlation coefficient $\rho = 0.838$ (p<0.05). The average measure intra-class correlation is 0.859, and it is higher than 0.547 for all dimensions (specialists). By this, we can appoint a strong inter-rater reliability.

The results are drawing a picture of the interdisciplinary rehabilitation team by its communication pattern during the team meetings. As it can be seen on the network diagram (Figure 1), that the physicians are in center of the communication network: sending and receiving are the most utterances. The physiotherapists are in close and frequent interaction with the physicians, and thus their conversation dominates the team meeting. An important result is that the nurses' role in communication is not as salient as it would have been expected by the importance of their work in rehabilitation. And also the specialists who are not working for the specified ward are less active in the team meeting (e.g. ergotherapist). These results are represented by the network measures too (Table 2.). The density of the communication network is 0.619 (62,9% of the possible connections appeared). The leading position of the physicians is underpinned by the highest closeness measure = 0.167, and also by the highest betweenness measure = 11.0 in the team. The ergotherapist, the social worker and the sport therapist have the lowest three closeness and betwenness measures, what is congruent with

| Specialist groups | Closeness | Betweenness |
|------------------------|-----------|-------------|
| Physician | 0.167 | 11.00 |
| Nurse | 0.125 | 0.667 |
| Physiotherapist | 0.125 | 3.00 |
| Occupational Therapist | 0.125 | 1.334 |
| Ergotherapist | 0.100 | 0.0 |
| Sport Therapist | 0.111 | 0.0 |
| Social Worker | 0.111 | 0.0 |

Table 2. Network measures computed by AGNA.

their almost peripheral position in the communication network.

CONCLUSION

The results are showing the leader position of the physician in the rehabilitation team, which is a formal role as the coordinator of the rehabilitation process [6], and in our study it is underpinned by the communication data. The possible two weak points of the interdisciplinary team are (1) the weaker than expected communication link between the nurses and the other professionals (especially between the nurses and the physiotherapists), and (2) the nearly marginal role in the meeting of the specialists who are not employees of the specific ward. The possible changes toward correcting the weak communication links may lead to a faster developing, and better functioning shared mental models is the team, that may provide more effectiveness, and patient safety. These results can also be explained by the fact, that the interdisciplinary team meeting and the philosophy of the interdisciplinary team work is a new development in the institute [8]. The method for the quantitative communication analyses of medical teams have been developed and successfully used in this field research of medical rehabilitation team. This method, as a communication behavior measurement protocol could be applied in different domains of teamwork, and also it is going to be completed in the near future with a quantitative and qualitative analysis method of the content of discussions in interdisciplinary team meetings.

- AGNA Social Network Analysis Software, developed by Benta, I. M. (2003). Free to download, at <u>http://www.tucows.com/preview/295919</u>
- 2. Bokhour, B.G. Communication in interdisciplinary team meetings: What are we talking about? *Journal of Interprofessional Care*, 20, 4, (2006), 349-363.
- Cooke, N.J. Salas, E., Cannon-Bovers, J.A., Stout, R. Measuring Team Knowledge. *Human Factors*, 42 (2000), 151-173.

- Devenport, T.H., Glaser, J. Just-in Time Delivery comes to Knowledge Management. *Harvard Business Review*, (2002), 5-9.
- Eigisti, I.-M., Bennetto, L., Dadlani, M. Beyond pragmatics: Morphosyntactic Development in Autism. *Journal of Autism and Developmental Disorders*, 37 (2007), 1007-1023.
- Eldar, R., Marincek, C., Kullmann L. Need for Rehabilitation Teamwork Training in Europe. *Croatian Medical Journal*, 49 (2008), 352-357.
- Eppich, W.J., Brannen, M., Hunt, E.A. Team training: implications for emergency and critical care pediatrics. *Current Options in Pediatrics*, 20 (2008), 255-260.
- 8. Hámornik, B.P., Juhász, M., Soós J. Az orvosi team tudás, interakció, és kommunikáció vizsgálata a

rehabilitáció területén. [The analysis of medical team knowledge, interaction, and communication in rehabilitation] In *10th RODOSZ Conference Issue*, Cluj-Napoca, Romania, (2009), 136-147.

- Hámornik, B.P., Juhász M. A team-munka szerepe a rehabilitáció orvosi munkában [The role of teamwork in medical practice] In Juhász, M. (ed.) A foglalkozási rehabilitáció támogatása pszichológiai eszközökkel [The Support of Vocational Rehabilitation by Psychological Tools] Typotex, Budapest, (2010). 61-78.
- Nemeth, C.P., Kowalsy, J., Brandwijk, M., Kahana, M., Klock, P.A., Cook, R.I. Between shifts: Heathcare communication is PICU. In Nemeth, C. P. (ed.) *Improving healthcare team communication*, University of Chicago (2008), 135-153.

Measuring Professional Teams' Information Sharing Behaviour

Juliánna Katalin Soós

soosj@erg.bme.hu

Márta Juhász juhaszm@erg.bme.hu Balázs Péter Hámornik hamornik@erg.bme.hu

Budapest University of Technology and Economics Department of Ergonomics and Psychology Egry J. utca 1, Budapest, H-1111 Hungary

ABSTRACT

Professional teams are highly differentiated from other teams through exclusive membership of expert specialists, where the team members have to integrate their specific knowledge in order to operate the technologically complex system on team level. In the present study, we aimed to describe Nuclear Power Plant (NPP) team members' communicative behaviour under different levels of task load and to focus on the predictability of team performance on the bases of communication. We had observed and analysed 16 NPP operator teams' information sharing behavior in a simulator environment. Video recorders of operators' activity during a selected scenario have been used for collecting and analyzing data. According to our analyses, several specific communication dimensions were related to performance.

Author Keywords

Team work, team cognition, communicative behaviour analyses, and different level of task load.

ACM Classification Keywords

H5.3. Group and Organization Interfaces: Synchronous interaction.

INTRODUCTION

In recent years the technological developments has risen significantly, which led to the spread of complex operations in the field of work. Additionally, the increased complexity of workflow and tasks necessitates a multioperator environment. No wonder that professional teams have started to play a critical role in complex operations, where two or more members having specific roles, responsibilities will be able to accomplish tasks that are too complex for individuals. Professional teams execute brief operations or missions repeatedly under technologically complex conditions, which requires extended professional, vocational training and preparation from the organization as well as from the individuals [4,1].

We aim to examine Nuclear Power Plant (NPP) operator teams work in high risk environment, where it is expected to avoid irreversible failures that endanger human life or cause material damage. High risk environments are environments in which there is a more than normal chance of damaging one's own life, the life of others or material property [4]. The operator team is exposed to face with several environmental task load factors, such as noise, prolonged work, work shifting, complexity, novelty, uncertainty and time pressure, which may have significant effect on the team performance [7]. As long as operator teams have to work under various, changing condition of task load it is important to develop mutual coping strategies, which constitute efficient information sharing coordination, cooperation, activity, terminally communication.

The main question in studying operator teamwork is how the team members representing different special fields are able to operate and manage a technically complex system, in a high risk environment. According to theoretical approaches of team cognition each individual has two different models: team mental model, referring to the collective task and team relevant knowledge (roles and responsibilities, knowledge of teammates, skills, abilities, beliefs), and team situation model, describing team collective understanding of the specific situation. These team level knowledge constructs guide the team in assessing and interpreting cues and patterns of the current situation [1,2,3,5]. Furthermore team level knowledge structures allow team members to reliably predict each other's needs, actions and to act on these without explicit communication. In this way the team process, specially communication is the most appropriate means to prepare for a coordinated cooperation in emergencies, non routine problems creating a shared understanding of the situation [8]. The information sharing activity is essential for the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

coordination within the team, and for the way how a team handles and manages difficult, high task load situations, where the prompt, accurate information flow is crucial especially in a technically complex environment such as NPP. In this sense appropriate information sharing activity, communication can reduce and manage both the causes and the consequences of high task load.

Our study focused on the NPP operator teams' information sharing activity, namely communication, in order to identify and understand those key communicative behaviours that support the joint assessment of the current situation and help to develop adequate team strategies to face unpredictable emergencies.

METHODS

The data collection was based on operator team interactions analysis in the Simulator Centre of a Hungarian Nuclear Power Plant. Since communication is the central factor of our research, the empirical studies of a "lively" interaction can best be carried out by analyses of carefully chosen simulator sessions. The Hungarian NPP Simulator Centre may be considered as a realistic, high-fidelity tool that is widely used in training and examinations creating the required level of face-validity to be relevant for real life situations. The Nuclear Power Plant's operator teams consist of four professional fields requiring the interaction of six members: Unit Shift Supervisor, Reactor Operator, Turbine Operator, Field Operator, Unit Electrician, and Shift Leader. Data from 16 operator teams' interaction have been collected. Each team had to follow the same scenario: however, the operator's reaction may have led to some slight differences. Choosing the simulation, we took into consideration that the scenario had to be oriented toward communication: in this way, all team members had to be involved in solving the control task. Possessing complementary knowledge they had to share information with each other to manage the problems occurring during the simulated malfunctions.

In order to provide a complete picture of simulation the scenario will be described briefly ("Failure of one turbine unit"): according to the annual schedule used by instructors, a live switchover test needs to be performed, while an unjustified operation of the turbine protection occurs resulting in the failure of one turbine unit. The failure of the equipment is followed by the malfunction of the primary circuit pressure control, creating a condition that also needs to be managed. The mean duration of scenario is about 35 minutes. The scenario was divided by the instructors into 3, according to the level of task load (see Table 1).

Performance Evaluation

The performance scores were made by the instructors' evaluation, both at individual and collective levels. The individual performance was based on the evaluation how the role related tasks were accomplished, using the 3-point Likert scale (1 - poor, 2 - medium, 3 - excellent). The team

| Phases of scenario | Estimated level of task load |
|--------------------|------------------------------|
| Phase 1 | Moderate level of task load |
| Phase 2 | High level of task load |
| Phase 3 | Moderate level of task load |

Table 1. Phases of scenario according to the estimated level of task load.

performance was assessed by the instructors' impression about the teams' efficiency under the different phases of the scenario using the same 3-point Likert scale. Eliciting data from performance assessments we developed four team performance categories:

- Excellent team: the whole team performance was evaluated excellent, through all the phases of the scenario.
- Medium team: the team performance is medium continuously through all the phases of the scenario.
- Unbalanced team: the team performance was varying from excellent to poor through the scenario.
- Poor team: the team performance was evaluated steadily as poor through the complete scenario.

Video records of operators' activity during the selected scenario have been also used for collecting and analyzing data. In order to keep the operators' real life behaviour the instructor informed them at the beginning of the simulator study about video recordings during the ongoing training session, but they did not know exactly which of the programmed scenarios would be videotaped. Video recordings were made with the operators' joint consent.

All the recorded conversation of the operators was transcribed in chronological order, identifying the operators' verbal utterances. Difficulties occurred in transcribing videotapes due to communication density during some period of the interaction, much simultaneous conversation flow between members, additionally we had to face with a noisy control room environment. For all these reasons we have few blind points in the transcribed videotapes, where the speaker of some utterances cannot be identified properly.

Communication Analyses

In our study we aimed to capture some relevant task specific aspects of NPP team's information gathering and processing behaviour. In order to capture the most relevant task-oriented aspect of teams' information sharing activity some task specific communication dimension have been developed expanding and specifying the communication dimensions used in similar environments (Conversation Analysis; Speech Act Type-inventory for the Analyses of Cockpit Communication, STACK [6]). The final major task-oriented communication dimensions were the following groups:

Information Collecting Question

The aim of the question is information acquisition, for example asking about certain indicators or resources. This can be formulated in the following two ways:

- *Open Question Information*: The question is addressed in order to complete the proposition with information; therefore it is likely to receive a long answer. It usually starts with words like *what*, *when*, *who*, etc.
- *Closed Question Information*: The aim of this question is verification, to judge the truth of a position; therefore the answer is expressed with either a single word (yes, or no) or a short phrase. For example "*Can we start the program*?"

Information Providing

The team members inform each other about some relevant aspect of the mission related to human or technical indicators. This may be grouped into three categories according to the time focus:

- *Information Providing Past*: The speaker informs the addressee about technological information, certain indicators that happened in the past, or about the crew's past status, personnel resources in the past.
- *Information Providing Present*: The speaker informs the addressee about some actual, present technological information, certain indicators, or about the crew's present status, personnel resources.
- *Information Providing Future*: The speaker informs the addressee about some technological information that may change in the future, foretells about certain indicators, or about his intentions and future actions.

Affirmation

It is the manifestation of two-way communications. It may be formulated in two ways:

- *Simple Affirmation*: Answers to yes/no question or commands. For example affirmations, acknowledgements, acceptances, answer such as 'yes', 'no', 'ok', 'good'.
- *Affirmation with Information*: A feedback, reinforcement on a status report or information, or command completed with additional information.

Coherence Analyses

The anchored point of the *coherence analyses* was the new thought (that can be a question, information, etc.) initiated by one of the team members. The main condition of the coherent conversation is *the turn-taking*, taking up this thought, the interlocutor develops a new question, information or command related to the previous information. Otherwise, if an initiated thought is not taken

up by any of the team members, it will be considered as a *thought without turn taking*.

Two independent evaluators rated the teams' transcribed information flow, based on the communication dimensions described above.

Ethical Statement

The research has been authorised by the Management of the Hungarian Nuclear Power Plant. All the research participants were informed about the research including the videotape recordings and evaluation measurements.

RESULTS

Appropriate information collection and distribution allows the team to understand better the situation, helping to build a shared conceptualization of the faced problems. According to our analyses, several specific communication dimensions were related to performance both on individual and team level. Particular forms of questions proved to be the best way to dispel uncertainties, and to realize safe communication. Our results revealed that the excellent performing teams use fewer open information collecting questions than the lower performing teams (F= 4.690, p<0.05). The frequent use of open questions suggests that when lower performing teams express their questions, they have less information, knowledge about the environmental cues, so they formulate the question in a less complete form, being unable to face the challenging of situation. The open questions are incomplete and force the addressee to use the cognitive resources to complete the proposition.

For the efficient information flow between team members it is also important to answer the supposed question, to provide the information in timely manner. The unbalanced teams show the highest communication density, regardless of the time orientation. These teams use the more frequent information providing activity about the present (F=7.109 p=0.005), past (F=4.779 p<0.05) and future (F=1.337, p>0.05) as well compared to excellent, poor and average performing teams. High density of communication without any special focus indicates the failure to capture the most relevant aspect of the ongoing episode. The results indicate that excellent performing teams focus on the present during their information providing activity, orient least of all about the past. The high performing team's communication is focused on the status, attributes, and dynamics of relevant actual elements in the environment, involving the processes of monitoring, cue detection, perception which leads to the awareness of their current status. It is also important to use the relevant information to project future situations in accordance with the team's goal.

The indicators of efficient communication include also confirming the received information. The use of simple affirmation helps the team to clarify and acknowledge the received information, in this way to establish an accurate shared understanding of the situation. Conversely the affirmation completed with additional information will overload the cognitive resources of both the information provider and receiver. Although the differences are not significant, the results can be regarded as a tendency that describes excellent performing teams using more simple affirmations and fewer affirmations with information, conversely with the low performing teams, where team members exchange more affirmations with information. The result indicates the need for a clear information change that helps to establish accurate team knowledge, instead of creating an interference with additional, not so relevant, information.

When analysing the communication under different level of task load, we observed the tendency of decreasing communication density during high task load, furthermore the increasing communication after high task load. Generally it can be concluded that as the task load increases the frequency of communication dimensions decreased. During high task load the communication is severely impeded, which can be explained by the operators' overloaded cognitive resources. The unexpected problems, failures intensively load the team members' cognitive capacity, being unable to share their attention between the accomplishment of the task and communication. Furthermore as the allocated resources disengaged, the collective need to process the causes and the consequences of unexpected event resulted in more frequent communication. It has been also explored that the poor teams' conversations indicate an incomplete flow of information between team members. Comparing the coherence indicators of excellently and poorly performing teams' dialogue, it has been explored that the poor teams' conversations include more thoughts without turn-taking (t=5,506, p<0,05), and less thoughts with turn-taking (t=4,069, p=0,05). The coherent information flow between team members proved to be an efficient communication attain high performance. strategy to Coherent communication means that the team members are aware of the information distributed by others, and react to the received information (either with a simple affirmation, or with a question, or with additional information transfer), creating a semantic connection in the information sharing activity. In this way coherent communication is one of the key elements of the effective establishment or modification, fine tuning of accurate and complete team knowledge.

CONCLUSION

The study considers some specific aspects of information sharing behaviour that could be directly linked to establishing team knowledge, such as using open and close information questions, affirmations, information provision, and coherence of information flow. The use of effectively formulated information collection utterances, the development of a well established effective communication strategy that focuses on the ongoing events and projecting the environmental cues to future situations, affirming the received information could all help the team to build, modify accurate team knowledge and to improve team performance. Future research should more thoroughly investigate the characteristics of such information sharing behaviour across different high risk environment teams (such as medical team, cockpit team), where the teams have to operate in highly standardised settings under various task load. The future work should go also beyond communication, studying other team processes, such as coordination, decision making, and also capturing the professional knowledge structures, mental model, or situational awareness at both individual and team levels.

- Cooke, N.J.-Salas, E.-Cannon-Bowers, J.A., Stout, R. Measuring team knowledge. *Human Factors*, 42, (2000), 151-173.
- Cooke, N.J.-Salas, E.-Kiekel, P.A.-Bell, B. Advances in measuring team cognition. IN E. Salas and S. M. Fiore (Eds.), *Team Cognition: Understanding the Factors that Drive Process and Performance*, Washington, DC: Am. Psych. Assoc. (2004), 83-106.
- Cooke, N.J.-Gorman, J.C. Assessment of team cognition. In P. Karwowski (Ed.), 2nd EDITION-International Encyclopedia of Ergonomics and Human Factors, UK: Taylor & Francis Ltd., (2006), 270-275.
- 4. Dietrich, R.-Childress, T.M. *Group interaction in high risk environments*. The GIHRE Project. ASHGATE, Burlington, 2004.
- Klimoski, R.J., Mohamed, S. Team Mental Model: Construct or Metaphor. *Journal of Management*, 20, 2 (1994), 403-437.
- 6. Krifka, M. Structural Features of Language and Language use, In: Dietrich, R., Childress, T.M. *Group interaction in high risk environments*, The GIHRE Project. ASHGATE, Burlington, 2004.
- Mumaw, R. J. The effects of stress on NPP operational decision making and training approaches to reduce stress effects, (NUREG/CR-6127). US Nuclear Regulatory Commission, 1994.
- 8. Sexton J.B., Helmreich R.L. Analyzing cockpit communications: The links between language, performance, error, and workload. In *Proceedings of the Tenth International Symposium on Aviation Psychology*, (1999) 689- 695.
Child-Activity Recognition from Multi-Sensor Data

Sabri Boughorbel sabri.boughorbe@philips.com Jeroen Breebaart jeroen.breebaart@philips.com Fons Bruekers fons.bruekers@philips.com

Ingrid Flinsenberg ingrid.flinsenberg@philips.com Warner ten Kate warner.ten.kate@philips.com

Philips Research Laboratories Eindhoven High Tech Campus, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

ABSTRACT

The automatic recognition of child activity using multisensor data enables various applications such as childdevelopment monitoring, energy-expenditure estimation, child-obesity prevention, child safety in and around the home, etc. We formulate the activity recognition task as a classification problem based on multiple sensors embedded in a wearable device. The approach we propose in this paper is to apply spectral analysis techniques of multiple sensor data for activity recognition. Quadratic Discriminant Analysis (QDA) classifier is then trained using manually annotated data and applied for activity recognition. The obtained experimental results for the recognition of 7 activities based on a limited data set are promising and show the potential of the proposed method.

Author Keywords

Activity recognition, feature extraction, activity classification.

INTRODUCTION

The automatic recognition of activities by using sensors such as tri-axial accelerometers can be used in a variety of applications. A first category entails context-aware applications. It consists in adapting, in real-time, the environment according to the recognized activity, such as changing the light condition according to the posture of the person (lying-down, sitting, etc). Other examples are giving an alarm if an elderly person falls, or when a child demonstrates potentially hazardous activities (such as climbing the stairs). The second category of applications is behavior monitoring. It consists in analyzing the activity on the longer term and the identification of trends or certain lifestyle properties. As an example, the amount of physical activity can be monitored to quantify a person's energy expenditure, possibly combined with coaching activities to monitor progress during physical training. Moreover, as a sedentary lifestyle is becoming a commodity in many developed countries, monitoring and stimulation of physical activity can be used to prevent obesity, both with adults as well as for children [1].

In this work we are interested in the automatic recognition of child activities. We distinguish two categories of targeted applications: *real-time* automatic recognition relevant for acute child safety (such as fall detection and stairclimbing), and *long-term* activity recognition and logging to track child development and to support preventing child obesity. The recognition and quantification of sedentary activities such as lying-down, sitting, watching television, etc are relevant to assess any required changes in lifestyle. This can be achieved by stimulating the children by mean of games or rewards. Additionally, activity recognition can help to improve energy-expenditure estimation (e.g., the activity *level*) since it has been shown that the energy expenditure is dependent on the type of activity [2].

The field of automatic activity recognition has been extensively researched. Accelerometers have been employed for many years in the analysis of body posture and activity especially in a clinical setting [3]. In [4] small biaxial accelerometers have been worn simultaneously on different parts of the body. Features such as mean, standard deviation, energy, correlation and frequency-domain entropy are used. The study in [5, 6] investigated the use of multiple sensors (bi-axial accelerometers, light and temperature sensors, microphones) placed at different body locations to recognize human locomotion.

ACTIVITY RECOGNITION

The automatic activity recognition of children is formulated as a classification problem where two stages are involved: a feature extraction stage and a classification stage.

Feature Extraction

The multi-sensor device we used in this research provides tri-axial acceleration data, air pressure data, and tri-axial

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

gyroscope data. The use of multimodal analysis is expected to enrich the data and enable the recognition of wider number of activities. The feature extraction is obtained through two cascaded steps: first-order and second-order feature extraction. The first-order features are chosen to be robust to noise and sensor orientation. For the second-order feature extraction, we employ a spectral variance analysis as previously used for the classification of audio and music signals [7].

First-Order Features

In order to be robust against noise and invariant to a certain extend to the sensor orientation we considered the vector magnitude of acceleration (AccMag) and vector magnitude of gyroscope data (GyrMag). Additionally, the normalized z-component of the accelerometer (AccZ/AccMag) and the measured pressure (Pres) are used. Thus our first-order feature vector consisted of 4 components [AccMag, GyrMag, AccZ/AccMag, Elev].

Second-Order Features

To allow analysis of the dynamic behavior of first-order features, a second-order feature extraction stage was employed. Each of the 4 first-order features were processed as follows:

- **Moving average** by applying a first-order low-pass filter with a cut-off frequency of 0.1 Hz;
- **Moving variance** by first filtering the signal with a 2nd order band-pass filter (0.15-20 Hz), subsequently computing the square of the individual samples, and averaging the result with a low-pass filter with a cut-off frequency of 0.5 Hz;
- Moving RMS 0.1-2 Hz by filtering the signal with a 2nd order band-pass filter (0.1-2 Hz), computing the square of the individual samples, averaging the result with a low-pass filter (0.5 Hz cutoff frequency), and computing the square-root of the result;
- Moving RMS 2-4 Hz same procedure as the previous one except for the first band-pass filter having a pass-band of 2 to 4 Hz;
- Moving average slope by first applying a 4-th order lowpass filter (0.1 Hz), computing the difference of subsequent samples, and applying a smoothing filter (0.5 Hz cutoff frequency).

Thus we extracted 5 second-order features for each firstorder feature, hence in total 4 x 5 = 20 features are obtained. This feature vector is then used as input for the classification stage.

Activity Classification

Generic classification approaches can be applied for activity recognition based on the extracted features. In this work we applied three classifiers namely Linear (LDA), Quadratic (QDA) and Adaboost classifiers [8]. LDA is computationally efficient but cannot deal with non-linearity that is usually present in data. Adaboost is a good classifier in term of generalization however it is known for its sensitivity to data size and noise. QDA provided a good trade-off between classification and computation efficiency. Manually annotated data were used to train and test the classifiers. The train and test set was obtained by recording sensor data while the sensor device was carried by a child. The considered activities are listed in Table 1 with the corresponding labels.

| Label | Posture | Label | Posture |
|-------|-----------------|-------|-------------|
| 1 | Walking | 5 | Falling |
| 2 | Lying-down | 6 | Standing-up |
| 3 | Running | 7 | Other |
| 4 | Climbing stairs | | |
| | | | |

Table 1. List of considered activities and corresponding labels.

EXPERIMENTAL PROCEDURE AND RESULTS

The experiment, which was approved by Philips Research ethics committee, was conducted during a normal activity of a 2-year old child in an indoor setting. As shown in Figure 1, the wireless sensor device was placed in the back pocket of the trouser. The sensor device is based on the Aquisgrain wireless sensor platform [9]. The sensor signals were sampled with a frequency of 50 Hz. Seven types of activities have been considered during the experiment. Approximately 30 minutes of sensor data has been recorded with a synchronized video material for ground-truth creation. The sensor device contains a tri-axial accelerometer with maximum acceleration of 2g, a tri-axial gyroscope for the determination of the sensor orientation and a pressure sensor for the determination of elevation. The data obtained from the pressure sensor needs to be normalized when both indoor and outdoor activity are analyzed. However, we limited the experiment to indoor activities. The elevation data obtained from the pressure sensor is useful to detect climbing and falling events.

From the acquired data, train and test sets were obtained by randomly splitting the data segments into training and test sets. A 3-fold validation was employed to obtain mean



Figure 1. A 2-year old child having the sensor device in the back-pocket of the trouser.

classification performance numbers including 95% confidence intervals.

The mean classification performance obtained using only first-order features amounted to $38.2\%\pm1.5\%$. However the second-order features resulted in a good performance of $97.8\%\pm0.2\%$. Second-order features computed from only accelerometer data i.e., [AccMag, AccZ/AccMag], provided $79.9\%\pm1.5\%$. This result shows clearly the benefit of using a multi-sensor approach.

Figure 2 shows the normalized confusion matrix of the classification results obtained for the second-order features. The activities of *climbing stairs*(4), *falling*(5) and *standing*up(6) have been recognized with more than 99% of accuracy. Such high accuracy is required for building safety applications based for instance on fall and stair-climbing detection. Most of the errors are obtained from the confusion between walking(1) and running(3). This can be explained by the fact that sometimes it is difficult to distinguish between walking quickly and running. To solve this issue, additional classes can be introduced to better describe the speed and the intensity of walking and running. Figure 3 shows an example of a sensor signal (AccMag) and the obtained classification result as a function of time. The output of the classification (in red) is according to the labels in Table 1.

CONCLUSIONS

Child-activity recognition is formulated as a classification problem where two stages are involved, namely feature extraction and classification. Second-order features by means of spectral analysis have been applied on a multisensor data where additional sensors beyond accelerometers have been used. The obtained results for a limited data set indicate the high potential for the proposed approach. As future work, both the extension towards more activities, as well as the validation of the results on a larger data set are required to better quantify the benefits of the proposed approach compared to state-of-the-art methods.



Figure 2. Normalized confusion matrix for 7 activities obtained with the proposed features.



Figure 3. The magnitude of the acceleration is plotted together with obtained classification results. The magnitude was scaled to be able to plot both classification as well as sensor data.

- 1. Livingstone, M. B. E. Childhood obesity in Europe: a growing concern. *Publich health nutrition*, *4*, 1a (2001), 109-116.
- Arvidsson, D., Slinde, F., Larsson, S., hulthen, L. Energy cost in children assessed by multisensory activity monitors. *Medicine and science in sports and exercise*, 41, 3 (2009), 603-611.
- K. Aminian and Najafi, B. Capturing Human Motion Using Body-Fixed Sensors: Outdoor Measurement and Clinical Applications, *CAVW*, 15 (2004), 79-94.
- 4. L. Bao and Intille, S. Activity recognition from user annotated acceleration data, *In Proceedings of PerCom*, 2004.
- 5. N. Ravi, *et al.* Activity Recognition from Accelerometer Data, *In Proceedings of AAAI* (2005), 113–116.
- 6. Ward J. et al. Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers, *IEEE Trans. PAMI*, 28, 10 (2006).
- Maurer U. *et al.* Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions, *Proceedings of the International Workshop on WIBSN*, (2006), 113-116.
- McKinney, M. F., Breebaart, J. Features for audio and music classification. In *Proc. 4th Int. Conf. on Music Information Retrieval*, Baltimore, Maryland, 2003.
- 9. R. O. Duda, Hart, P. E., and Stork, D. G. *Pattern Classification*, Wiley-Interscience Publication, 2000.
- N. Loew *et al.* Medical Sensors of the BASUMA Body Sensor Network. *In IFMBE Proceedings*, 2007.

Studying Media Usage in the Living Room by Measuring Infrared Signals

Thomas Mirlacher IRIT - ICS 118 Route de Narbonne 31062 Toulouse, Cedex 9, France thomas.mirlacher@irit.fr Thomas Fischer thomas.fischer@ruwido.com Regina Bernhaupt regina.bernhaupt@ruwido.com

ruwido research Köstendorfer Str. 8 5202 Neumarkt, Austria

ABSTRACT

While HCI field studies are traditionally undertaken by using tools from social sciences, this paper describes a complementary method for studying media usage in the living room. It provides a first step for blurring the boundary between lab and field-test studies for media usage in living room.

We constructed a device that measures and records the distinct infrared (IR) patterns of the remote controls used in the living room to interact with media devices. These recordings are later used to deduct usage patterns across media devices.

Author Keywords

User behavior, media usage, living room, Infrared recording.

ACM Classification Keywords

H.5.m Information interfaces and presentation: Miscellaneous

INTRODUCTION

Traditionally, HCI field studies are conducted by using a set of tools from a large repository of social sciences. These tools allow insights into a user's daily life and often include cultural probing or a variation like playful probing (Bernhaupt et al. [1]).

For a finer temporal description of user experiences, methods like the experience sampling method (ESM) and the day reconstruction method (DRM) are getting more and more common.

ESM was originally referring to a technique involving random signaling of users during their daily lives (Larson and Csikszentmihalyi [5]). Today, the term ESM is used more broadly, and describes procedures that assess in-situ experiences at the time (or near the time) they are occurring (Christensen et. al [2]). Experience Samples can be taken in response to a random signal, at predetermined times, or following a particular event (or events).

While ESM requires the user to report nearly in real-time, DRM (Kahneman et al. [4]) is designed to provide detailed descriptions of daily activities and experiences by the mental reconstruction of a day – Participants provide demographic information and rate items on satisfaction and mood, then divide the day into episodes starting at arising and ending with retiring in the evening. Each episode is marked for activity, location, and social contact, and rated on positive and negative affective states.

However, for studying the behavior of users in the living room we needed a complementary measurement, providing objective and quantitative data as well and allowing us to monitor users behavior with an even finer temporal resolution. We need a tool that would allow us to observe users at home, with minimal interference of the users daily life and without intrusion into their privacy. Essentially, we want to have lab-like measurements in the field, to eliminate lab-based influences on the user's behavior.

Lab-based studies have the advantage of a higher degree control, and the use of techniques such as video monitoring of user interaction, eye-tracking as well as other physiological measurements, while field studies have the advantage of being closer to the way users are likely to interact with technology.

The approach described in this paper should be a first step for blurring the boundary between lab and field-test studies for media usage in living room.

APPROACH

Since a majority of interaction in the living room is performed via a remote control (or as we can find in most households an accumulation of remote controls), we can achieve lab-like measurements of user interactions and therefore usage behavior through measuring and recording infrared (IR) signals.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. IR measuring dongle with USB cable.

Each remote control is producing a distinct IR pattern for each key pressed, and this is what we are measuring and recording. We record the raw demodulated IR signal for post-processing in the lab. This eliminates the need to update the device for each new remote control, and keeps the hardware and its firmware as simple as possible. For being able to find out when exactly a button is pressed we are also storing a timestamp derived from built-in real time clock, allowing us temporal reconstruction of the remote control usage as recorded in the living room. But placing a device into the users home usually requires some extra issues to be considered: The device has to seamlessly blend into the users' already installed infrastructure. Users are usually reluctant if a device has to replace one of their own devices, or the researcher needs to rewire the whole home theatre equipment. This rewiring would take quite some amount of time, is error prone and even less likely to be welcome by a user. So the only viable approach, which we have chosen, is the "side-car" approach, to have a device sitting next to the users own devices.

IMPLEMENTATION

Figure 2 depicts the block diagram of the IR measurement hardware, which is placed in the households to record IR frames sent by all remote controls present in a room.

Most used IR protocols that are based on carrier-modulated pulses, exhibit a carrier frequency of 35-39kHz or 56-58kHz respectively.

IR receiver modules are available in a wide variety of carrier frequencies, although the passband of the bandwidth filters is usually big enough to successfully receive IR frames of carrier frequencies deviating from the center frequency.

Therefore, two IR receiver modules have been selected for the IR measurement dongle that exhibit a center frequency of 37kHz and 56kHz. These modules allow the detection of all carrier based IR protocols using these two frequency bands.

Protocols that are not carrier based cannot be detected by the IR dongle, but these protocols sum up to less than 5% of all still used IR protocols.

The IR modules filter the carrier frequency of the modulated IR signal and output a demodulated signal (see Figure 3). A microprocessor measures the timing for high and low states of the IR signal and stores those on a flash based memory card along with a timestamp of microsecond resolution.

Since we are very interested in the correct timestamp – which is configured once when the device is set-up – we wanted to make sure that under no circumstances this valuable datum is lost. Therefore, an external real time clock and calendar (RTCC) chip is used, that is additionally backed up by a large capacitor (SuperCap) and is able to keep the time and date for up to two weeks without external power supply by the USB interface.

For a rapid deployment of the IR dongle, we have developed software, which allows us to set the clock and read the contents of the memory card via USB.

This software further offers learning of the IR remote controls to match the IR pattern to the key pressed on remote control (Figure 4).

This information is used in a post-processing step after the IR dongles return from the households. The files with the recorded IR patterns are decoded using the data retrieved during deployment. This allows the analysis of multiple IR protocols without prior knowledge of the type of remotes used in a household.



Figure 2. Block Diagram of the IR recording hardware.



Figure 3. Modulated IR signal as received by the IR dongle. Demodulated signal as measured after the IR modules.



Figure 4. IR Logger Setup Tool used to set the current time on the IR dongle, as well as use it in USB mode to send the received signal to the computer for displaying it.

CONCLUSIONS

With this small recording device we can reveal media usage patterns within a household, as well as general patterns common to all the households in a region.

We can find out if a user belongs to the group of Zappers, Loyals and Casuals (as characterized by Jenkins [3]) Zappers constantly shift channels and essentially only watch snippets of shows. For them the fast transition from one channel to another or from one type of content to another is essential. Loyals cherry pick content and spend more time socializing about their shows. They are the series watchers and are more likely to record shows on DVRs. Hence, the capability to easily record and navigate through recorded content may have high value. The Casuals have elements of both: they wander away from boring shows and will have a tendency to multitask until they find a show that attracts their attention.

This IR recorder enables us to optimize remote controls, since most of them are just there to fulfill one or two functions. Imagine a typical system with TV set, Hi-Fi system and set top box (STB): usually all tree devices are powered on, but then only the volume is changed with the Hi-Fi remote control, while the TV channel is switched with the STB remote.

FUTURE WORK

The first field study using this IR recorder is currently conducted, but we have already identified some optional enhancements to the current recorder.

For future field studies we plan on recording additional signals and optionally add sensors. These sensors can measure bio-physiological values like heart rate or skin conductance.

Another improvement is to identify users interacting in the living room, by adding a camera and storing a video stream. But placing a camera into a living room causes privacy concerns, and so another approach is inevitable. A possible solution would integrate in the already working system: users can wear simple IR badges, which are repeatedly sending a unique pattern, which in turn can be recorded by the existing IR measuring device, and users within a household can be distinguished, while still guaranteeing anonymity.

- Bernhaupt, R., Weiss, A., Obrist, M., and Tscheligi, M. Playful probing: Making probing more fun. In *INTERACT* (Berlin/Heidelberg, Sept. 2007), M. C. C. Baranauskas, P. A. Palanque, J. Abascal, and S. D. J. Barbosa, Eds., vol. 4662 of Lecture Notes in Computer Science, Springer, 606–619.
- 2. Christensen, T.C., Feldman Barrett, L., Bliss-Moreau, E., Lebo, K., and Kaschub, C. A practical guide to experience-sampling procedures. *Journal of Happiness Studies*, *4*, 1 (2003), 53–78.
- 3. Jenkins, H. Convergence Culture: Where Old and New Media Collide. NYU Press, New York, 2008
- 4. Kahneman, D., Krueger, A.B., Schkade, D. A., Schwarz, N., and Stone, A.A. A survey method for characterizing daily life experience: the day reconstruction method. *Science*, *306*, 5702 (2004), 1776–1780.
- Larson, R., and Csikszentmihalyi, M. The experience sampling method. New Directions for Methodology of Social & Behavioral Science, 15 (1983), 41–56.
- Millen, D.R. Rapid ethnography: time deepening strategies for HCI field research. In DIS '00: Proceedings of the 3rd conference on Designing Interactive Systems (New York, NY, USA, 2000), ACM, 280–286.

Optimization of User Satisfaction on an Augmented Reality Based Guided Tour

Marília Felismino mfelismino@gmail.com **Ricardo Luiz** ricardo.luiz@ydreams.com

Edifício YDreams, Madan Parque - Sul, P-2825-149 Caparica, Portugal

Marta Castela martavcastela@gmail.com **Pedro Marques**

pedro.a.r.marques@gmail.com

Faculdade de Psicologia e de Ciências da Educação, Alameda da Universidade, P-1649-013 Lisboa, Portugal

ABSTRACT

This paper reports on work being conducted with the goal of optimizing users' satisfaction on an augmented reality based guided tour. It shows the applied methodology, the metrics used to assess the quality of the user experience and the results obtained from the statistical analysis of the experimental results. We stereotyped users by considering whether or not they had a background related to information technologies, and compared their behavior patterns. Next, we evaluated how social inhibition would affect the way users' interacted with the applications. We then tested for correlations between several measured metrics and users' satisfaction answers, in order to acquire knowledge such as: the influence that the ease of use had in users' satisfaction; whether the time that users spend interacting with the Interactive Installations (I.I.) directly influenced their satisfaction; and others. Finally, we analyzed the influence that the relative visit order of each I.I. in the guided tour had on the users' satisfaction, and found an optimal sequence in which to present I.I.s to visitors, so as to provide them with the most pleasant possible experience.

Author Keywords

User modeling, Behavior Metrics, Augmented Reality, Guided Tour Route Planning, Optimization, User Modeling, Affective Computing

INTRODUCTION

This project was developed by YDreams, SA in collaboration with the Faculty of Psychology of the

University of Lisbon. YDreams is a global company that works with the concept of interactivity, with a focus on the field of Augmented Reality technologies. Over the last few years, YDreams has been developing full-scale interactive environments (from stores to exhibitions), products and intellectual property, combining technology and design. This company has a showroom that is frequently visited by customers, schools, and media, among others. These visits are guided by salespeople who present the interactive installations (I.I.), but in an order that changes randomly with each visit. As the users are naturally different from each other, the aim of this project is to study the best way to create an immersive environment in the showroom that adapts the visit according to the different user types, in order to maximize their satisfaction. To achieve that, there are several areas of research that are being studied and applied, such as affective computing [1], storytelling[2], user modeling [3], intelligent agents and smart objects [4], augmented reality [5], among others. This paper presents a component of the user modeling research. The first section describes the problem we studied and the methodology applied is in the second section. The results obtained with the performed experiments are reported on the experiment results section. Finally the last section contains a short summary, conclusions and the future work.

THE PROBLEM

Ydreams' showroom has several I.I. that are frequently shown to visitants, but the users' behavior on those visits had never been studied. This paper aims to shed light on a set of questions about the way users interact with those I.I.. To achieve that, six I.I. representative of users' behavior were selected. These are YWalk, YReal, Tangitable, YWaterfall, YMagic and YFaces. Three of them show the user's image as he/she interacts with it (YReal, YWaterfal, YFaces); two are entertaining and easy to use (Tangitable, YWalk), and two are new and have a considerable learning curve (YReal, YMagic). YFaces also personalizes users'

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

interaction, as the application catch them by a camera and places their image, after some transformations, inside a comic book with personalized text in the balloons.

The research questions considered in this study were:

- 1- If we stereotype users based on whether or not they have an IT background, do we observe any significant difference in their behaviors?
- 2- When users are publicly displayed in the I.I. with which they are interacting (e.g.: in a video projection), is their behavior influenced by any kind of social inhibition?
- 3- Can a user's degree of satisfaction with an I.I. be inferred from the time he/she spends interacting with it?
- 4- How do users evaluate an I.I. they find easy to interact with? Do they consider it a boring experience, or a satisfactory one?
- 5- Are there any I.I.s that more strongly influence users' global satisfaction with the guided visit than the remaining ones?
- 6- At which point in the guided tour should those I.I. that provide a personalized experience to the users be presented?
- 7- How strongly does the relative order in which I.I.s are presented during the guided tour influence the users' satisfaction, and the remaining metrics under consideration?

METHODOLOGY

For this study, a sample of 30 people was selected, with ages between 20 and 30 years old. Part of the sample was composed of university students with a strong technology background (like computer science, aerospatial engineering, among others) and the other part was composed of students from a less technological background (like psychology, nursing, among others). Several visits were scheduled, in groups of 2 people, always with identical academic backgrounds per group. Each visit consisted in presenting those 6 interactive installations and a feedback questionnaire at the end. In order to analyze for each I.I. how much its relative position in the guided tour influenced the users' satisfaction and the other analyzed metrics, 3 different visit orders were tested. Several questions were included on the questionnaire, with the goal of measuring users' satisfaction on this set of tests. In the future the goal is to infer this metric automatically and dynamically during the visit [6]. In order to achieve that, several other metrics were measured with the goal of correlating them (simple or composed) to users' satisfaction answers. These metrics were *latency* (time since the instruction to interact is given until users start to interact), duration (time spent interacting) and user *engagement* (time since user stops to interact until the beginning of the next presentation usually users ask several questions after the I.I. presentation). The variables in analysis for each I.I. were then: user satisfaction, ease of use, latency, duration and

engagement. The global variables analyzed were: boredom of the visit (*boredom*), global satisfaction with the visit (*global satisfaction*), whether the visit met expectations (*expectations*) and the likelihood of recommendation of this visit to other users (*recommendation*).

EXPERIMENTAL RESULTS

1. If we stereotype users based on whether or not they have an IT background, do we observe any significant difference in their behaviors?

In order to answer to the first question, we compare the results from users with an IT background versus users without one. The statistical results obtained considering the users' distinction were: Average of User satisfaction by interactive installation (Fig.1); Average of Ease of use by interactive installation (Fig.2); User satisfaction by interactive installation and by order in sequence presentation (Fig.3). Fig.1 shows that there are differences between the users' preferences. For example, YWaterfall is the most preferred for IT users and the least preferred for the remaining users. The perception of the different interactive installations' ease of use was also different between user types. This fact can be observed in Fig.2 where, for example, YWaterfall was considered the easiest for IT Users and almost the hardest for the remaining users. Analyzing the order in which interactive installations were presented (Fig.3), we see no significant differences between the two types of users. In order to analyze if there are statistically significant differences between the answers given by the two groups, the Friedman Test with Replication [7] was applied. The results obtained were a pvalue of 0.62 for users' satisfaction and 0.47 for ease of use, with a 0.05 significance level. Analyzing the results obtained by the Friedman Test with replication, we see no statistically significant differences between answers given by IT Users and Other Areas Users on these two questions. But, in Fig.1 some I.I. seem to have differences. In order to verify this fact, the Kolmogorov-Smirnov Test was applied to each I.I. [7]. The results showed that, at the $\alpha = 0.05$



Figure 1. Users' Satisfaction by Interactive Installation and by Users' background.



Figure 2. Perception of Ease of Use by Interactive Installation and by Users' background.

level of significance and p-value of 0.004, there exists enough evidence to conclude that there is a difference in user satisfaction for the YWaterfall I.I. between the IT group and Other Areas Group. The users' satisfaction and perception of ease of use for the remaining I.I. don't differ significantly.

There are no major differences between users with an IT background versus users without one. YWaterfall was the only I.I that had presented statistically significant differences between these two users' stereotypes.

2. When users are publicly displayed in the I.I. with which they are interacting (e.g.: in a video projection), is their behavior influenced by any kind of social inhibition?

The I.I. that shows users image are YWaterfall, YReal and YFaces. In order to study the social inhibition, we compared latency, duration of interactions and users' satisfactions of these I.I., with the remaining ones. The results showed that YWaterfall and YFaces had the shortest duration, YFaces had the longest latency and YReal were the one that users liked less. Taking these results together with the observations done during the visits, where we could observe that users didn't feel comfortable seeing their images on the screen, we can affirm that these I.I.s promote social inhibition on users.

3. Can a user's degree of satisfaction with an I.I. be inferred from the time he/she spends interacting with it?

To investigate this question we applied Spearman correlation [7] to users' satisfaction with each I.I. and the corresponding interaction duration. There was no statistically significant correlation between these two metrics for any I.I.s. Moreover, duration hardly appears in correlation results with other metrics. However, some correlations do give us some degree of information related to this topic. For YReal, *Satisfaction vs Ease* had a



Figure 3. Users' Satisfaction by Interactive Installation order in guided visit, grouped by users' background.

correlation of 0.631, *Ease vs Duration* had a correlation of -0.508, *Boring vs Duration* had 0.437 and *Expectations and Duration* had a correlation of -0.46, at a 0.01significance level. These results shows that, for YReal, the easier users considered it, the better their satisfaction and the shorter would be their interaction duration with it. Also, the longer the duration, the more users found this I.I. to be boring, and the worst would users' expectations about the visit be met. So, although we did not have any direct correlation result that relates duration with satisfaction, we could infer that the shorter the duration, the better users' satisfaction is. As the other I.I. didn't present any information about the interaction duration, we could only say that, with the data obtained with this experiment, the only result is that the shorter the duration, the better.

4. How do users evaluate an I.I. they find easy to interact with? Do they consider it a boring experience, or a satisfactory one?

To analyze this fact we applied Spearman correlation [7] to the ease of use, boredom and users' satisfaction metrics. The results showed that for YWaterfall, YMagic, YReal and YFaces ease of use and users' satisfaction had a statistically significant positive correlation (at 0.01 significance level). So, the easier users found these I.I., the higher their satisfaction with it. Also, YWalk, Tangitable and YFaces showed a statistically significant negative correlation (at 0.01 significance level), for *ease of use vs boredom*. So, users like I.I to be easy to use.

- 5. Are there any I.I.s that more strongly influence users' global satisfaction with the guided visit than the remaining ones?
- To achieve this result we applied Spearman correlation [7]



Figure 4. Influence of previous I.I on users' satisfaction for each I.I. (1- very dissatisfied; 6- very satisfied).

to users' satisfaction with each I.I. and their global satisfaction evaluation. The results showed that users' satisfaction with YWalk, YMagic, YReal and Tangitable have influence in their global satisfaction with the whole visit. YWalk had a weak correlation of 0.33 at a 0.05 significance level, YMagic had a 0.433 at a 0.01 significance level, YReal had a correlation of 0.451 and Tangitable had a correlation of 0.514, both at a 0.01 significance level. So, Tangitable is the most influent on global users' satisfaction, followed by YReal and YMagic and, at the end, Ywalk. In future experiments, these I.I. will be given a more central role in the visit, so as to provide users with a more satisfactory experience.

6. At which point in the guided tour should those I.I. that provide a personalized experience to the users be presented?

The I.I. that provides a personalized experience to the user is YFaces. This I.I. was tested on the 2^{nd} place, 4^{th} place and 6^{th} place (end of the visit). Although the differences were small, users' satisfaction with this I.I was better if it was presented at the beginning of the visit. Fig.1 also shows that this I.I is one of the preferred by users and is one of the easiest to interact with, so it is a good one to "break the ice" at the beginning of the visit.

7. How strongly does the relative order in which I.I.s are presented during the guided tour influence the users' satisfaction, and the remaining metrics under consideration?

Fig. 4 depicts a graph where all the considered interactive installations appear as nodes, and the directed edges represent the measured satisfaction levels (value shown as the edge label) of the I.I. on the destination node when it was visited at some point in the visit after the I.I. on the

source node had already been visited. At each node, the thickest of its incoming edges represents the precedence that was observed to give it the highest satisfaction level in these experiments. The short number of tests we were able to perform meant that we could not obtain enough reliable information from which to measure all the possible edges. Instead, we allocated people in the multiple tests over a set of 3 different visit orders that would collectively supply information about an adequate variety of edges. As a result, we can observe that the I.I.s in Fig. 4 all have both in- and out-degrees of 2 (with the exception of YWalk). When the information obtained from these experiments is represented in this way, the problem of identifying the visit order that maximizes the satisfaction levels of all I.I.s is identical to the well known Travelling Salesman Problem [8] in combinatorial optimization. Here however, instead of seeking the shortest tour between cities, we want the tour with maximal satisfaction levels. Optimizing the visit order by taking this goal into account, we come up the following sequence: YWaterfall - YFaces - YWalk - YMagic -YReal - Tangitable. Considering the values measured in our experiments, this would be the optimal visit order for the I.I.s in the company's showroom.

CONCLUSIONS AND FUTURE WORK

The aim of this project is to maximize users' satisfaction in a guided tour environment. To achieve that, some experiments were carried out. The results obtained showed that in general there are no statistically significant differences in behavior between users with an IT background and those without one. Results also showed that users' satisfaction with YWalk, YMagic, YReal and Tangitable have influence in their global satisfaction with the whole visit. Finally, the influence of each I.I.'s relative position in the guided tour on users' satisfaction was also analyzed. The best order in which to visit the multiple I.I.s in the guided tour was inferred, from the feedback data provided by the users in our tests, to be one starting with YWaterfall, followed by YFaces, then YWalk, YMagic, YReal and finally, Tangitable.

The next steps to be done are, at first, to stereotype users based on their information into multiple classes. We will further investigate this topic because we observed, while the test visits were occurring, that users' behavior and interactions with some I.I.s did not follow the same pattern for all users. We then intend to check which metrics could better segment users in order to personalize and adapt the visit to each user type. Afterwards, we will create an optimization algorithm for dynamical selection of the best I.I. sequence to be used in the tour, which takes users' stereotypes into account. Finally, we will implement a virtual agent in the showroom, with the goal of providing users with a more personalized experience. To achieve that, the agent will adapt its behavior by taking into account feedback measured from users' interaction with I.I.s..This agent will function as an interaction unblocker, by showing the user how to interact with each I.I., as well as by

performing other tasks.

- 1. R. W. Picard, Affective Computing, MIT Press, 1997.
- 2. Carvazza, M. et al. Interacting with Virtual Characters in Interactive Storytelling. Proceedings of the first international joint conference on Autonomous agents and multiagent systems, 2002.
- Wang, Y. Aroyo, L., Stash, N., and Rutledge, L. Interactive User Modeling for Personalized Access to Museum Collections: The Rijksmuseum Case Study. In Proc.the 11th International Conference on User Modeling (UM'07), (2007), 385-389.
- 4. M. Wooldridge. *An Introduction to Multiagent Systems*. John Wiley and Sons, 2002.
- 5. R. T. Azuma, "A survey of augmented reality," Presence: Tele-operators and Virtual Environments, 6, 4 (1997), 355-385.
- Parmenter, D. Key performance indicators: developing, implementing, and using winning KPIs. NJ: John Wiley & Sons, 2007.
- 7. S. Siegel. Nonparametric Statistics. McGraw-Hill, New York, 1956.
- 8. Applegate, D. L. *et al*, The *Traveling Salesman Problem: A Computational Study*, Princeton University Press, 2006.

Posters

Proceedings of Measuring Behavior 2010 (Eindhoven, The Netherlands, August 24-27, 2010) Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

A New Method of Computer-Supported Measurements of Distances Moved by Animals

Julia Brendle Julia.Brendle@agrar.uni-giessen.de Hoy Steffen Steffen.Hoy@agrar.uni-giessen.de

Department of Animal Breeding and Genetics, Justus-Liebig University Bismarckstr. 16, D-35390 Giessen, Germany +49-641-9937647

ABSTRACT

A newly developed software to measure the distance moved by animals under practical conditions was studied in an experiment using a total of 32 piglets. Each piglet, with an age of nine days, was kept under standardized conditions in an open field (OF) for ten minutes. The behaviour was video-recorded using a camera installed vertically above the OF. During the methodological investigation the distances moved were measured first automatically computersupported by the software EthoVision® (Noldus) and secondly with the special-made software VideoMotionTracker® (Mangold). The analysis of correlation, regression and also the comparison of the means showed a very good correspondence between both methods.

Author Keywords

Locomotion, computer-supported measurement, distances moved, farm animals.

INTRODUCTION

Locomotion defined as the motion of an animal in conjunction with the change of location can allow details about abnormal behaviour, health status and well-being. The existing methods to capture the locomotion of farm animals (e.g. treadmill, pedometer, photoelectric barriers, Local Positioning System – LPS, Global Positioning System GPS, ETHOSYS or video tracking systems) can be used only restricted in conventional housing systems. Automatic tracking systems (EthoVision, Noldus) require high requirements on the identification of the individual animal.

If it is possible to create standardized test conditions these

programs offer the opportunity to measure locomotion of individual animals computerized [1]. In practice with unequal conditions (light – dark, different materials and backgrounds) it is awkward if not impossible to realize reliable measurements.

The aim of this study was to develop and test a new tech solution allowing computerized measurement of distances moved by farm animals related to group size and other factors of housing.

MATERIAL AND METHODS

The basic requirement to use the new developed software VideoMotionTracker® is the generation of digitalized videos which are recorded by a camera installed vertically above the pen or cage (observation of the whole analysis-area).

Before the measurement can be started the dimension of the pen (length / width) has to be defined in an x-/y-coordinate system. To begin the measurement the examiner has to position the mouse-pointer or alternatively the stylus at a defined point of the test animal.



Figure 1. Camera view of analysis-area in a pig housing unit.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 2. View of a marked test animal.





Figure 3. Distance moved by a pig (example).

Also it is possible to visualize so-called "activity hot spots". With "heat maps" the program can announce the frequency of using different areas in the pen (e.g. feeder or watering place) by the animal.



Figure 4. Heat map with activity hot spots.



Figure 5. Comparison of the distances measured with EthoVison® and VideoMotionTracker®.

RESULTS

In methodological investigations, the distances moved by a total number of 32 piglets with an age of nine days under the standardized conditions of an open field during ten minutes (based on digitalized video sequences) were measured first automatically computer-supported by the software EthoVision® (Noldus). Referring to the automatic measurement the same 32 piglets were analyzed with the special-made software VideoMotionTracker® (Mangold). The results of the comparison of the means (EthoVision® 91,3 m; VideoMotionTracker 91,4 m) showed a very good correspondence between both methods. Furthermore, the correlation analysis (r = 0,997) showed a highly significant coherence.

CONCLUSION

With the VideoMotionTracker®, a method is available to measure the distance moved by animals during a defined period of time (e.g. 24 hours). The precondition for this computer-supported measurement is that the whole area can be video-observed. In combination with infrared technology the program can also be used to analyze the behaviour of nocturnal animals.

Based on the present results various fields of application in ethological research are conceivable in the future.

REFERENCES

 Lind, N.M., Vinther, M., Hemmingsen, R.P., Hansen, A.K. Validation of a digital video tracking system for recording pig locomotor behavior. J. Neurosci. Methodes, 143 (2005), 123-132.

Optimization of a Contextual Conditioning Protocol for Rats Using Combined Measurements of Startle and Freezing

Laura Luyten

Laboratory of Experimental Functional Neurosurgery, Department of Neurosciences K.U.Leuven, Provisorium I Minderbroedersstraat 19, bus 1033, 3000 Leuven, Belgium laura.luyten@med.kuleuven.be

Kris van Kuyck

kris.vankuyck@med.kuleuven.be

Debora Vansteenwegen

Centre for the Psychology of Learning and Experimental Psychopathology, K.U.Leuven Tiensestraat 102 bus 3712, 3000, Leuven, Belgium deb.vansteenwegent@psy.kuleuven.be

Bart Nuttin¹

bart.nuttin@uz.kuleuven.ac.be

Laboratory of Experimental Functional Neurosurgery Department of Neurosciences, K.U.Leuven, Provisorium I Minderbroedersstraat 19 bus 1033, 3000 Leuven, Belgium

ABSTRACT

Contextual conditioning in rats is typically quantified using startle amplitude or freezing time. Our goal was to create a robust contextual conditioning protocol combining both anxiety measures. Comparison of ten 0.8 mA - 250 ms shocks with an established shock configuration (0.3 mA - 1s) favoured the first parameters. Next, we systematically investigated the effect of shock intensity (0.6 mA, 0.8 mA or 1.0 mA) and concurrently compared two conditioning procedures (shocks alone versus explicitly unpaired shocktone presentations). The 0.8 mA shocks produced significant contextual freezing and startle potentiation, whereas the 0.6 mA and 1.0 mA shocks only led to a significant increase of freezing time. We found no major differences between the two types of conditioning, implying that these procedures might be equivalent. In conclusion, training with ten 0.8 mA - 250 ms shocks produced reliable contextual conditioning as measured with both startle amplitude and freezing time.

Keywords

contextual conditioning, startle amplitude, freezing, shock intensity, conditioning protocol, rat.

INTRODUCTION

Background

In classical cued fear conditioning, a neutral cue (e.g. a tone) is repeatedly paired with an aversive stimulus (usually footshock). The tone acquires aversive properties and consequently produces fear responses. Conditioned emotional responses are also elicited by reintroducing the rat in the context (e.g. the experimental cage) in which it has previously experienced an aversive event. Contextual conditioning is enhanced when using unpredictable shocks, and can thus be obtained by training the animal with shocks alone or with explicitly unpaired shocks and tones [1,3]. In future experiments, this second type of conditioning may form the optimal contrasting condition for a cued fear conditioning group, trained with explicit cue-shock pairings.

The expression of contextual anxiety in rats is typically quantified by measuring startle amplitude (of the wholebody startle reflex elicited by a loud noise) or freezing time (total immobility of the rat except for respiratory movements) [2].

To date, a whole range of protocols has been used to impose contextual conditioning, but systematic investigations of the different aspects of these protocols are largely lacking in the literature. Moreover, in most studies, anxiety is quantified with one single behavioural measure.

Aim

Our goal was to create a robust contextual conditioning protocol for rats, combining both startle amplitude and freezing time as measures of contextual anxiety. Therefore,

 $^{^1} Neurosurgery, UZ Leuven, Herestraat 49 bus 7003, 3000 Leuven, Belgium. Tel.: +32 16 344290.$

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

we compared different shock parameters and training protocols.

MATERIALS & METHODS

Subjects

The experiments were conducted on 80 male Wistar rats (24 rats in experiment 1, 56 rats in experiment 2) weighing 200–250 g at arrival. They were housed in groups of 3 with food and water ad libitum available. The rats were maintained on a 12-h light–dark cycle (lights on at 6:00 a.m.) with a room temperature of $\pm 21^{\circ}$ C. All experiments were carried out in accordance with protocols approved by the animal ethics committee of the Katholieke Universiteit Leuven.

Experimental Setup

To record the startle amplitude, a stabilimeter device was used. The rats were placed into an acrylic cylindrical rat holder (7.6 cm inner diameter, 14.3 cm length) with a grid floor, which was firmly placed on the response platform by four thumb screws. The grid floor consisted of nine 3-mmdiameter stainless steel bars spaced 9 mm apart, through which footshocks could be delivered (ENV-414SA-SR + ENV-262B-GF, Med Associates, St. Albans, VT, USA). The stabilimeter and platform were located inside a ventilated sound-attenuating chamber (65 x 52 x 52 cm) (MED-ASR-PRO1, Med Associates). A red light bulb (3.8 W) in this 'startle box' was continuously on. The behaviour of the animals was recorded by a video camera (DCR-SR55E Super NightShot Plus, Sony, Tokyo, Japan)



noise-alone trial = startle stimulus (50 ms, 100 dB, white noise)

- tone-noise Irial = Ione (10 s. 75 dB, 4000 Hz) co-terminating with startle stimulus (50 ms, 100 cB, while noise)
- in TONES groups: both trial types were presented in a balanced, invigular order, with a maximum of two subsequent noise-stone or tone-noise trials
- ITI = intertrial interval

Figure 1. Experimental designs.

positioned in front of the rat holder. Afterwards, the freezing behaviour during the first 5 min of the test sessions was analysed from videos by a blinded observer.

The startle reaction of the rats generated a pressure on the response platform and analogue signals were amplified, digitized, and processed by software (Startle reflex, version 5.95, Med Associates) provided by the manufacturer of the equipment. The presentation and sequencing of the acoustic stimuli and footshocks were controlled by the same software (which does not allow to administer shocks longer than 250 ms in a protocol with unpaired tones). One of two loudspeakers, both located 7 cm behind the rat holder, was used to deliver a continuous white background noise (55 dB), the other speaker delivered the startle and tone stimuli. The amplitude of the startle response was defined as the first peak accelerometer voltage that occurred during the 200 ms after onset of the startle stimulus. The stabilimeter and loudspeakers were calibrated before each experiment.

Figure 1 shows the designs of experiments 1 and 2. On 4 consecutive days, the rats were placed in the startle box and after 5 min. of acclimation (background noise only), the session started. Freezing during acclimation and startle amplitude on noise-alone trials were measured on pre-test (baseline) and post-test (expression of anxiety). Repeated measures ANOVAs were used to compare the contextual conditioning potential of the different shock parameters and conditioning protocols. Since we found no significant effects of the presence or absence of tones during testing and training, we redid the ANOVAs for experiment 2 and omitted the factor 'presence of tones'. Tukey's post-hoc tests were carried out, with the significance level set at p < 0.05.

RESULTS

In experiment 1, we compared the contextual conditioning potential of 0.8 mA - 250 ms shocks with an established shock configuration (0.3 mA - 1 s) [4] and a non-shocked control group. With the 0.8 mA - 250 ms configuration, significant contextual conditioning was achieved, as measured with both startle amplitude and time of freezing. On

the contrary, the 0.3 mA $-1\,$ s shocks only produced a significant increase of freezing time, not of startle amplitude.

In experiment 2, we wanted to investigate whether we could still improve the protocol, to obtain even more robust results. We therefore investigated the effect of shock intensity, comparing 0.6 mA, 0.8 mA and 1.0 mA shocks with a non-shocked control. Since the 0.8 mA shocks produced both significant contextual fear-potentiated startle and freezing, this was the optimal configuration compared with the lower or higher shock intensities (0.6 mA and 1.0 mA) which only had a significant effect on time of freezing, but not on startle amplitude. In addition, we compared two different conditioning procedures (shocks alone versus explicitly unpaired shock-tone presentations) and found no major differences between them, implying that these procedures might be equivalent.

CONCLUSION

The aim of this study was to produce reliable contextual conditioning as measured with both startle amplitude and freezing. This goal was achieved using a protocol with 10 unsignalled 0.8 mA - 250 ms shocks.

- 1. Fendt, M. and Fanselow, M.S., The neuroanatomical and neurochemical basis of conditioned fear. *Neurosci. Biobehav. Rev.*, 23, 5 (1999), 743-760.
- 2. Grillon, C., Startle reactivity and anxiety disorders: aversive conditioning, context, and neurobiology. *Biol. Psychiatry* 52, 10 (2002), 958-975.
- 3. Phillips, R.G. and LeDoux, J.E., Lesions of the dorsal hippocampal formation interfere with background but not foreground contextual fear conditioning. *Learn. Mem. 1*, 1 (1994), 34-44.
- 4. Santos, J.M., et al., Pharmacological dissociation of moderate and high contextual fear as assessed by freezing behavior and fear-potentiated startle. *Eur. Neuropsychopharmacol.*, *15*, 2 (2005), 239-246.

EEG Data Logging with the NeuroLogger[®] System: Spontaneously Occurring Electrophysiological Correlates of Migraine in Mutant Mice

Alessandra Di Lio

Institute of Pharmacology and Toxicology, University of Zurich Winterthurerstrasse 190, CH-8057 Zurich alessandra.dilio@pharma.uzh.ch

Hans-Peter Lipp

Institute of Anatomy, University of Zurich Winterthurerstrasse 190, CH-8057 Zurich hplipp@anatom.uzh.ch

Dimitrula Arabadzisz

Institute of Pharmacology and Toxicology, University of Zurich Institute of Anatomy, University of Zurich Winterthurerstrasse 190, CH-8057 Zurich arabadzi@anatom.uzh.ch, arabad@pharma.uzh.ch

Hanns Ulrich Zeilhofe

Institute of Pharmacology and Toxicology, University of Zurich Institute of Pharmaceutical Sciences, ETH Zurich Winterthurerstrasse 190, CH-8057 Zurich zeilhofer@pharma.uzh.ch

Migraine is an episodic neurological disorder with spontaneously reoccurring headaches with or without aura. The best described correlates of migraine on the EEG level are neuronal hyper-excitability and cortical spreading depression. These phenomena are investigated in animal models by eliciting them pharmacologically, chemically, mechanically and/or electrically. Studies, investigating these phenomena when they occur spontaneously in freely moving animals are missing. To fill this gap and to investigate spontaneously occurring neurological correlates of migraine in freely moving mice, we used the cable free, non-telemetric data logging system, (NeuroLogger®) in genetically modified mouse lines like the Cacna1a mutants, which carry a point mutation (R192Q) in P/Q-type Ca2+ channel gene (cacna1a) originally described in patients with familial hemiplegic migraine. The NeuroLogger® is a microchip (16 bit MSP430 32 kb Flash, 1kb RAM processor; UART) with 750 kbit/s service interface. It is powered by two 1.4 V hearing aid Zink-air batteries and weighs about 2.8 g. It has 4 EEG channels, each sampling up to 500 Hz, 2 reference channels, 1 passive rolling ball as movement detector, and 1 synchronization channel for event marking and/or synchronizing the EEG record with behavioral records like video monitoring (1). The inputs to the NeuroLogger® are unity gain buffered with the AC input range of +/- 750 μ V, the four EEG channels have 1000x gain and are band-pass filtered 1-70 Hz. ADC resolution is 8 bit. Electrophysiological data is stored in hexadecimal format and can be converted either to the Spike2 (CED, UK) compatible file format (smr) or to the MatLab compatible file format (mat). With the NeuroLogger® System we successfully recorded EEG with 500 Hz sampling rate on 4 channels over 16 hrs and found that freely moving Cacna1a mutant mice have spontaneously occurring paroxysmal neocortical activity consistent with a neocortical hyperexcitability observed in animal models of migraine and in human migraine patients.

Keywords

Freely moving, EEG, synchronized video recording.

REFERENCES

1. Lars E. et al. Seizure logging: A novel approach to combined synchronized cable free EEG-registrations and Video recordings of seizure activity in the mouse. Submitted to *Neurosci Meth*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Ethosearch: A Comprehensive Repository of Ethograms for Use in Animal Behavior Research

Stephen R. Ross sross@LPZoo.org Megan R. Ross sross@LPZoo.org

Lincoln Park Zoo 2001 North Clark Street, Chicago, IL USA

ABSTRACT

Ethograms, as logically compiled catalogues of a species' behavioral repertoire, represent a fundamental underpinning of behavioral research. Well-designed ethograms underlie any rigorous quantitative study of behavior, because they actually specify "units" of data, the acts or events that are recorded and relationships analyzed. As the recognized need for the collection of standardized behavioral information in zoos and other study sites, along with the increase of collaborative, multi-institutional research, the need for a searchable, and easily used database of ethograms becomes critical. Ethosearch: the Ethogram Archive Project (EAP) is designed to address the lack of consistency and develop a tool that will be of critical value to zoological managers, researchers and students.

In 2000, a diverse group of behavioral scientists representing members of the Animal Behavior Society, Association of Zoos and Aquariums, International Society for Behavioral Ecology, and the American Psychological Association founded "EthoSource", envisioned as a centralized web portal through which to find, organize and interpret behavioral data. In 2003, this daunting initiative was divided into several components, one of which was an Ethogram Archive. In 2006, the Institute of Museum and Library Services (IMLS) provided funding to launch the Ethosearch Project. The Ethogram Archive Project (EAP) aspires to develop a central web-based tool, Ethosearch, that will serve a wide range of users: zoological managers, researchers, and students. It will consist of an Ethogram Archive and associated tools for searching, adding and relating ethograms to each other.

The goals of Ethosearch are:

- 1. To create a searchable, expandable database of existing and new ethograms;
- 2. To encourage and enhance the capacity of zoos to engage in collaborative, multi-institutional behavioral

Anne B. Clark

Dept of Biological Sciences Binghamton University, Binghamton, NY 13902 aclark@binghamton.edu

research;

- 3. To develop multi-level user interfaces and tools for specialists to use the database for their own research goals;
- 4. To provide educational outreach and research opportunities for students from K-12 through graduate school, associated with developing, writing, and utilizing ethograms.

The structure of Ethosearch is based upon the classification of specific behaviors and their functional and operational definitions. Every ethogram will have a number of informative "metadata" tags. These will include such information as author, publication and citation information and details on the animal population on which the ethogram is based. Users of the database will be able to use these tags to access specific ethograms or behaviors of interest. For instance, Agonism is a category within the functional hierarchy of the behavioral ontology. This breaks down still further, and allows the submitter to identify head bob as a particular type of agonism (threat). The behavior is also categorized as a body part movement in the operational hierarchy. This form of multi-level and flexible ontogeny will maximize the effectiveness of this tool for researchers and educators.

The applications for an archived database of ethograms are broad. Zoological managers are frequently in need of reliable descriptions of animal behavior to help guide and measure the impact of a variety of environment and management interventions on animal use and welfare. Likewise, students are often in need of a resource with which they can work to begin rudimentary animal behavior investigations. Such "starter projects" can take place at zoological gardens but also in less formal animal environments such as backyards, parks and farms. With potential application to both the scientific and education community, Ethosearch has a proven niche to address and has positioned itself as a critical resource to encourage high quality ethological endeavors. The Ethosearch website is scheduled to launch publically on the world wide web in the fourth quarter of 2010.

Author Keywords

Measurement, behavior, ethogram.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

The Complexity of Reliably Investigating Effects of Novelty Stress in Rats: Dynamics and Correlations of Behavioral, Cardiovascular, Endocrinological and Electroencephalographic Responses Under Simultaneous Measurement

Leen J. Raeymaekers lraeymae@its.jnj.com +32 4 607509 Wieteke Beerling wbeerli1@its.jnj.com W.H.I.M. Drinkenburg wdrinken@its.jnj.com

Neuroscience Research & Early Development, Johnson & Johnson Pharmaceutical Research & Development, a Division of Janssen Pharmaceutica, N.V. Turnhoutseweg 30, 2340 Beerse, Belgium

Jaap M. Koolhaas

Department of Behavioral Physiology, Biology Centre, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands j.m.koolhaas@rug.nl

ABSTRACT

Stress research in animals is generally hampered by limited possibilities to measure various relevant stress parameters simultaneously and by confounding stress effects of sampling procedures. The aim of this study is to analyze correlations between physiological, hormonal, electroencephalographic (EEG) and behavioral stress parameters, and their dynamics, under strictly controlled conditions with minimized confounding factors due to sampling procedures.

Therefore, stress parameters were measured simultaneously in freely-moving rats both under undisturbed, well-habituated home cage conditions and after novel cage exposure, which is considered a mild psychological stressor with good face validity and translational value.

First, male Sprague-Dawley rats were implanted with a telemetry transmitter for blood pressure, heart rate, body temperature, locomotor activity and cortical EEG and with a jugular catheter for blood sampling. Next, using a Williams experimental design, novelty stress was induced for 15 min by placing a rat in a new, clean, empty home

cage. Telemetric read-outs (Data Sciences Instruments: TL11M2-C50-PXT) and home cage behavior (analysis via The Observer XT) were recorded and analysed offline. Simultaneously, blood was sampled automatically every 10 min to determine plasma corticosterone levels.

Novelty exposure induced increases in sniffing behavior $(66.3 \pm 6.3\% \text{ vs. } 1.5 \pm 1.4\%, \text{ p} < 0.001)$, in rearing $(9.2 \pm 1.4\%)$ 1.6% vs. 0.4 \pm 0.3%, p<0.001) and in grooming (18.2 \pm 6.2% vs. 2.6 \pm 1.4%, p<0.02), together with a reduction in immobility $(0.4 \pm 0.4\% \text{ vs. } 84.3 \pm 11.8\%, \text{ p} < 0.001)$. This indicates that the animals were behavioral active during the exposure to novelty. During the first 20 min after this exposure to novelty, an increase in sniffing $(11.7 \pm 1.6\%)$ vs. $0.5 \pm 0.2\%$, p<0.001), rearing $(1.5 \pm 0.7\%$ vs. $0.0 \pm$ 0.0%, p=0.037), grooming (18.0 \pm 2.5% vs. 4.4 \pm 2.0%, p=0.001), immobility (41.0 \pm 10.2 % vs. 93.3 \pm 2.5%, p<0.001) remained and an increase in food uptake (19.0 \pm 9.2% vs. $0 \pm 0\%$, p=0.045) was induced. During novelty exposure, behavioral changes concurred with an increase in blood pressure (max 14.8 ± 2.9 mm Hg), heart rate (max 102.5 \pm 9.3 bpm), body temperature (max 1.0 \pm 0.2 °C), locomotor activity (max 22.2 ± 2.5 counts/min) and plasma corticosterone (max 131.9 ± 24.7 ng/ml). The latency to reach these maximum response values varied across parameters. Analyses of the EEG in terms of vigilance states showed results that were in accordance with these behavioral observations. Values for all parameters remained stable under control conditions, while in the stressed situation values gradually returned to stable baseline levels. This confirms that our approach

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

allows un-confounded measurement of several parameters simultaneously.

In summary, novelty stress causes physiological, hormonal, electroencephalographic and behavioral response that can be reliably measured with different time profiles during and after mild stress exposure. Various interesting correlations between the different read-outs are identified which can now be interpreted in conjunction. It is concluded that the present outcomes confirm the added value of simultaneous assessment of multi-factorial stress effects under well-controlled, un-confounded conditions.

Development and Evaluation of an Operant-Based Reversal Learning Task in the Rat, Relevance of the Visual Cue

Sofie Embrechts sembrech@its.jnj.com

Darrel J. Pemberton dpember1@its.jnj.com

Janssen Pharmaceutica, Department Neurosciences Turnhoutseweg 30, 2340 Beerse, Belgium

Impairments in cognitive flexibility and/or executive functioning are a core symptom of many neurological and psychiatric diseases, for example Alzheimer's disease and schizophrenia. In rodents, the attentional set-shifting task is one of the most widely employed paradigms to tap into this important cognitive domain. However, due to its complexity, considerable study length (~12 weeks) and potential for operator bias, the aim of the present studies was to develop and evaluate an operant-based version on this test initially focusing on discrete discriminations.

Studies were conducted in standard rat operant chambers (MED Associates, USA) utilizing two nose poke response apertures (with 3 internal coloured cue lights), an additional cue light above the response apertures and a food dispenser/hopper situated at the rear of the chamber. Initial experiments aimed to train animals to perform a simple discrimination utilizing a light on/light off rule. Half of the animals were required to track the stimulus light presented pseudo-randomly between the left and right response apertures, whilst the remaining half were required to respond to the aperture where no stimulus light was present. Animals were trained in daily thirty minute session's until they reached a performance criteria of >90% correct responses for three consecutive days after which the rule was reversed. Significant differences were observed during initial acquisition of the light on/light off rule; with light on animals showing significantly improved acquisition demonstrating cue/response learning is dependent on saliency of the cue. However, once both groups had reached

asymptote performance, there were no differences in performance prior to or following each of the reversals.

In a separate cohort of animals, attempts are made to train rats to discriminate between orange and green LED stimuli housed within the response apertures. Although rats were initially thought to be colour blind, recent behavioural experiments have shown that rats can indeed perceive ultraviolet light, and with training can distinguish between ultraviolet and visible light, and between different colours in the blue-green range (Jacobs et al. 2001). However, after ten daily sessions, performance remained static at chance level. This may be a consequence of the strain used (Lister Hooded) or more likely a consequence of the stimulus lights although surprising they are marketed and supplied for use with rodents MED associates (Part No: ENV-114M). This cohort was then transferred to a spatial discrimination protocol comprising of a light in/light above rule. Once again significant differences were observed during initial acquisition of the light in/light above rule; with light in animals showing significantly improved acquisition. However, once both groups had reached asymptote performance, there were no differences in performance prior to or following each of the reversals. In the final series of studies, the performance of three commonly used rat strains (Lister Hooded, Long Evans and Sprague Dawley) and the effects of sub-chronic phencyclidine (5mg/kg; ip; 7 days BID) were assessed on the reversal learning paradigm.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Measurement of Akinesia in Rats: Design and Validation of a Side Effect Paradigm in Freely Moving Animals

Michel Mahieu mmahieu@its.jnj.com

Hansfried Van Craenendonck

Luc Ver Donck

RWILLEM1@its.jnj.com

Lverdon1@its.jnj.com

Roland Willems

HVCRAENE@its.jnj.com

Johnson&Johnson Pharmaceutical Research and Development, a Division of Janssen Pharmaceutica NV, Turnhoutseweg 30, B-2340 Beerse, Belgium

ABSTRACT

In this paper we describe a method to assess akinesia, a side effect of antipsychotic drug treatment, in freely moving rats. Animals were placed on a circular platform and the latency to initiate exploration of the novel environment was recorded using image processing software. This was done by video tracking of the rat to determine time of first entry into a defined area, 15 cm away from the starting position. Vehicle treated rats immediately initiated exploration, while rats treated with antipsychotic drugs showed a dosedependent increase in latency to start exploration. It is concluded that drug-induced akinesia can be accurately assessed in rats using the present setup.

Author Keywords

Akinesia, antipsychotics, rat, behaviour, video tracking.

INTRODUCTION

Apart from their therapeutic benefit, CNS-medications are also known to potentially induce extra pyrimidal sideeffects such as catalepsy, akinesia etc ... [1]. Akinesia is the inability of a subject to smoothly initiate a voluntary movement. Drug-induced akinesia can be assessed in rodents in several ways, but is generally based on qualitative observations of animal behaviour. The purpose of the present study was to develop and validate a noninvasive set-up to obtain quantitative measures of akinesia induced by antipsychotics in freely moving unrestrained rats.

METHODS

Animals

Male Sprague-Dawley rats (Harlan, The Netherlands), weighing 230-260 g at the time of the experiment, were used. Rats were housed in individual ventilated cages (25cm×33cm×18 cm; Tecniplast) in groups of four. The animals were allowed to acclimate prior to the experiment for at least 7 days after receipt from the supplier. All animals were maintained under controlled environmental conditions throughout the study: 22±2 °C ambient temperature, relative humidity at 60%, with a 12/12 h light/dark cycle (6PM lights off, 6AM lights on). Food and water were available ad libitum.

Experimental Setup

The setup was designed and constructed with key principles such as ease of use, multi functionality, low cost and robustness. A stable circular black platform of 120 cm diameter was positioned 75 cm above ground level, in a room with low light intensity (1 lux). The centre of the platform was marked with a 2 cm^2 red spot. A video camera was mounted 165 cm above the platform.

Image Analysis

Video images were captured using Ethovison 3.0 (EV [2]) on a Windows PC. By means of the EV-software, the image of the platform area was divided into 4 circular areas with identical width (Figure 1). The position of a rat on the platform was detected in real time using a manually set grey-threshold in EV: the centre of gravity of the region of interest (=rat) was taken as the rat's actual position in each video frame. At the start of a recording session, a rat was gently placed in the middle of the platform on the central mark in area 1. When an object (=the rat) was detected by EV in this area 1, the recording automatically started after 0.5 sec to allow the experimenter to with draw his arm. Recording was programmed to stop when the rat was first detected in the outer area of the platform (area 4), or after 30 seconds.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Top view of platform area showing the 4 circular sub-areas identified by the green lines, and the threshold image of a rat (red area). A white cross, that marks the centre of gravity in the threshold image, is taken as the actual position of the rat on the field.

Study Design

Each individual rat was evaluated in this setup just prior to administration of a dose of the test compound or its vehicle (t=0h), and after 0.5, 1, 2, 3 and 4 h. Tracking data from EV were stored on disk and analysed off-line: the primary parameter of interest was latency until first occurrence of the rat in area 2, which was considered as an indication that the rat had uninitiated a movement away from its starting position. In addition, also latency to first entry into area 3 and 4 respectively, total distance travelled, and total recording time were calculated.

RESULTS

Vehicle treated rats almost immediately started to explore the novel environment, and were detected in area 2 after a latency time of 0.7 ± 0.1 to 1.1 ± 0.4 sec (mean±SEM, n=25)

during the first hour. A small increase in latency was observed at later time points (up to 9.0 ± 2.6 sec at 4h). Haloperidol (0.01-0.63 mg/kg, n=5) and risperidone (0.08-2.5 mg/kg, n=5) showed a dose-dependent increase in latency time up to 30 sec during the 4h observation period, while MP10, a novel PDE10-inhibitor (0.16-10 mg/kg), showed a maximum of only 16 sec at 2.5 mg/kg (n=5).

CONCLUSIONS

The size of the starting area (area 1) was chosen such that the centre of gravity of the rat as identified by EV was detected in area 2 when the rat had made two or three steps away from the centre of the platform. This occurs almost immediately in control rats, as they will voluntary explore a novel environment [3]. This initiating behaviour was further facilitated by the low light intensity environment. This allowed an accurate assessment of the inability of a rat to initiate a movement as is the case in drug-induced The present results show that drug-induced akinesia. akinesia can be accurately assessed in rats using the present setup, and that novel therapies with a more favourable side effect can be identified. Whether akinesia elicited under these conditions involves motor deficits or impairment of motivation remains to be elucidated.

Ethical Statement

The study was performed in strict accordance with the European Communities Council Directive of 24th November 1986 (86/609/EEC) and was approved by the Institutional Ethical Committee.

- 1. Fowler SC, Ford KE, Gramling SE, Nail, G. Acute and subchronic effects of neuroleptics on quantitative measures of discriminative motor control in rats. *Psychopharmacology*, *84* (1984), 368–373,
- 2. Spink AJ, Tegelenbosch RA, Buma MO, Noldus LP. The EthoVision video tracking system--a tool for behavioral phenotyping of transgenic mice. *Physiology & Behavior*, 73 (2001), 731-44.
- 3. Hughes NR. Neotic preferences in laboratory rodents: Issues, assessment and substrates. *Neuroscience & Biobehavioral Reviews*, *31* (2007), 441-464.

Visualizing Group Dynamics Signals

Erik Kuijpers

Research & Development, GGz Eindhoven hjj.kuijpers@ggze.nl Stef van den Elzen Laboratory for Quality Software, Eindhoven University of Technology s.j.v.d.elzen@student.tue.nl

Henri Hambartsumyan

Laboratory for Quality Software, Eindhoven University of Technology h.hambartsumyan@student.tue.nl

ABSTRACT

The group dynamics visualizer is a solution to the problem healthcare workers are facing when working with social therapeutic groups. It is a well known problem that their reflection ability is affected negatively by the group dynamics. Healthcare workers become part of this groups dynamics and are no longer able to objectively observe and intervene. In this paper we describe the process of measuring group dynamics, a part of the 'group prevention plan'; more specific, giving feedback on the tension level of groups inside a psychiatric ward. A software application, 'group dynamics visualizer', is developed to measure and visualize the data. The visualization methods allow the user to quickly explore the data in order to find relations and abnormalities. This information is then used to prevent high-tension build up. Furthermore the software serves as an objective observation tool.

Author Keywords

Matrix representation grid, visualization, data exploration.

ACM Classification Keywords

H5.m Information interfaces and presentation: Miscellaneous, H.1.2 User/Machine Systems: Human information processing

BACKGROUND

By measuring and using group dynamics signals the healthcare worker is able to predict and prevent aggressive patient behavior within psychiatric units. Experience learns that if tension increases within a group of patients, the staff's reflection abilities are affected. A prevention plan supports the reflection abilities of the health care worker. This makes a collective intervention possible even when tension in the group continues to stay at a high level. Doing so, the autonomy of the staff increases. This prevention plan based on group dynamics signals consists of two parts.

1. A written part called 'group prevention plan' in which signs and signals of tension build up are described, including interventions based on de-escalation of the patient group and support of the staff.

2. An instrument which measures the tension building called 'Grid'. This is an important element because the health care workers themselves are part of or influenced by this tension building. So this Grid is an important objective monitor.

GROUP PREVENTION PLAN

The group prevention plan is based on a crisis development model [3]. This model is based on the solid phases of the development in crisis including the set of possible actions and interventions to de-escalate.

The basic situation is called phase 0 (the normal value; tension level is normal) to indicate the basic level of tension in a certain group. By internal or external influences this tension level can increase. This is described in level 1 to 3, as shown in the following table.

| Tension level | Signs and Signals | Intervention on patient- and staff level |
|------------------------------------|---|--|
| <i>Level 0</i> : "Normal value" | Grid-level 0 Relaxed environment | Normal routine, Self coaching |
| Level 1 | Grid-level 1 Irritated environment or busy group | De-escalate Self coaching Aimed at result |
| Level 2 | Grid-level 2 Group- and staff destructives | Directive acting External coaching |
| Level 3 | Grid-level 3 Aggression out of a group | Physical intervention Management makes decision External support |

Table 1. Four stages of tension.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

The most important characteristics of each phase, each level and the possible interventions are determined. Also agreements are made for daily measuring by filling in the Grid. Action must be taken, if the Grid shows abnormalities.

THE GRID

The Grid is an instrument which visualizes the tension building.



Figure 1. The matrix representation grid.

Many group indicators are hidden in the communication and the relation level between mutual group members and of the group members with the team. The health care worker who cooperates closely with patients is by definition part of this group's dynamics as well. Hence we often see that the reflection ability reduces within the staff on the communication and relation level if the tension in the patient group mounts.

The reflection of staff members, on the increased tension level of the patient group, will decrease. For this reason a prevention plan on noticeable group indicators is not sufficient, there is also a need for an objective monitor. We found such instrument in the Matrix Representation Grid by G.Ahlin [1,2]. The Matrix Representation Grid, hereafter referred to as 'Grid', measures the communication and relation level in the patient group by means of 8 dimensions:

- 1. Communication flow.
- 2. Imagery.
- 3. Emotional climate.
- Self disclosure.
- 5. Acceptance.
- 6. Relating pattern.
- 7. Authority pattern.
- 8. Boundary character.

The 8 dimensions of the group functioning are described on 8 axes. Staff members value which level the group functions on for each of these 8 dimensions on a 5 point scale. The points on the outer circle represent score values of 1, and the points on the innermost circle represent score values of 5.

If enough measurements are available, a typical normal tension level emerges. The score values indicating the tension level are characteristic for this group. If the tension level in the group deviates from typical values this becomes visible in the Grid.

Regular daily measuring is achieved by entering the scores to the automated version of the Grid. The routine scoring reveals increased tension levels. In combination with the group prevention plan it helps the further diagnosis of the tension build up and the adequate action which should be undertaken.

VISUALIZATION

There are four proposals for visualizing group dynamics. The first visualization proposition is the Matrix Representation Grid (see Figure 1). GDV (group dynamics visualizer) allows users to see data in a Grid with additional visual features to gain insight. The interpretation of the Grid will be explained in the next section.

Each dot on this Matrix Representation Grid represents a



Figure 2. The parallel axes plot.

score on one of the eight axes. This visualization is used to see the results of one particular measurement.

If one wants to get a deeper understanding of the relation among multiple aspects of a score the parallel axes plot (see Figure 2) is used.

This visualization presents essentially the same data, however, it also provides the possibility to plot multiple evaluation moments thereby visualizing correlation of different aspects among multiple evaluation moments. Also multiple evaluation moments of different groups can be plotted to explore group-correlations.

The downside of the parallel axes plot is that it does not provide a broad overview of evaluation in time. The next visualization, the stacked line diagram, provides this feature (see Figure 3). It reveals evaluation moments in time, providing the user with ongoing trends and giving the possibility to analyze events that lead to an extreme score.

The next visualization is called the scatter plot matrix (Figure 4). It visualizes the occurrence of a score on two aspects. The main advantage of this visualization is that is provides the possibility to explore a large dataset and discover correlations or patterns specific to a group. Correlation between two group dynamics aspects are visualized by color coding. A brighter color indicates a higher correlation between two aspects. This aids health care workers on the long term to further optimize and steer group behavior.

INTERACTION

Exploration of the data is only effective if the user is provided with an overview, is able to zoom, filter and see details on demand [4]. Interactive exploration of the data is provided by linking the above described visualizations together and providing the user with extensive exploration possibilities. For example, a user can click on a point of



Figure 3. The stacked line diagram.



Figure 4. The scatter plot matrix.

interest in the stacked line diagram and explore the according Grid. Another example is to show the parallel axes plot of all groups for a certain theme. This might reveal interesting relations; for example certain themes may always build up significantly more tension inside a group compared to other themes. All visualizations are linked to provide the user with easy navigational means to allow for extensive data exploration. Figure 5 shows the possible transitions between the visualizations.

INTERPRETATION

Most features in the visualizations, for example color, are group dependent. Each group has its own normal level of tension. For example, the tension in a group of children will be much higher compared to a group consisting of adults. Therefore the group of children will score high on certain aspects, which does not mean there is reason to take action. In order to determine the standard tension level of a group, the median value (x) for each of the aspects is calculated. To be able to detect abnormalities or high-tension build ups, the standard deviation (s) is calculated. If the current



Figure 5. Interaction between visualizations.



Figure 6. Fast tension level insight through use of color coding.

tension of one or more of the group dynamics aspects lies outside the (safe) range [x - s; x + s], this is an objective warning to the responsible health care workers. Tension levels which lie inside the safe range are colored green in the visualizations, whereas tension levels which lie outside the range are encoded by a red color. Furthermore a timeline, again color coded, shows the aggregated tension level of a group. This enables the user to determine very fast whether the tension level of a group is normal or not. An example is shown in Figure 6.

FUTURE WORK

The concept of visualizing group dynamics signals as well as the group dynamics visualizer is not limited to the field of psychiatry. It could easily be applied to other fields of interest in which an objective observation about the groups dynamics is desired. Furthermore the group dynamics visualizer could be coupled with other systems which take action if abnormal group behavior is observed. To conclude we think that the visualization of group dynamics in order to provide an objective observation can be successfully applied in many fields.

- 1. Ahlin G. Exploring psychotherapy group cultures. Stockholm: Kongl Carolinska Medico Chirurgiska Istitutet, 1996.
- 2. Kristina Herlitz, Göran Ahlin, Christer Sandahl and Ingela Brimberg. Developing the matrix representation grid (mrg): A method for observing group processes. findings from time-limited group psychotherapy for alcohol-dependent patients. *Behavioral Science*, (1996), 145–173.
- 3. Kuijpers E. Signaleringsplannen ter voorkoming van agressie. *Psychopraxis*, 7 (2005), 155-161.
- 4. Brock Craft and Paul Cairns. Beyond guidelines: What can we learn from the visual information seeking mantra? In IV '05: Proceedings of the Ninth International Conference on Information Visualization, pages 110-118, Washington, DC, USA, (2005). IEEE Computer Society.

Automated Video Registration in Home Cage for Measuring Maternal Behavior of Wild, Aggressive and Tame Rats

Maria Yu. Solov'eva msol@ngs.ru Irene Z. Plyusnina iplysn@bionet.nsc.ru

Institute of cytology and genetics of the SB RAS Prospekt Lavrentyeva 10, 630090 Novosibirsk, Russian Federation

Igor G. Tarancev egor@sl.iae.nsk.su **Evgeny D. Bulushev** e.d.bulushev@gmail.com

Institute of automation and electrometry of the SB RAS Prospect Koptyuga 1, 630090 Novosibirsk, Russian Federation

ABSTRACT

This work is aimed to study the maternal behavior using visual observation and automated video registration system in tame, aggressive and wild gray rats. It is shown that the frequencies of "outside the nest" and mother licking/grooming do not depend on the used methods in tame dams. The presence of observer is accompanied by decreasing of the frequency of these parameters in aggressive dams. It is supposed that this effect is associated with increased emotionality and stress responsiveness in aggressive females as compared with the tame females. The frequency of maternal arched-back nursing is decreased under visual observation compared with automated video registration in females of both lines. Maternal aggression of both selected lines and wild population is recorded using the same registration system. It is shown that maternal aggression towards unknown male conspecific is lower in wild rats, than in tame and aggressive ones.

Author Keywords

Tame and aggressive rats, maternal aggression, maternal behavior, visual observation, video registration.

INTRODUCTION

Maternal behavior in rodents is associated with long-term programming of individual coping capacities in the offspring. Gray rats selected for elimination and enhancement of aggressiveness towards human differs in emotionality, anxiety and stress response [5, 8]. Crossfostering showed that postnatal maternal environment doesn't affect behavior in tame and aggressive rats but can influence corticosteroid stress response of tame rats [9]. It is still unknown whether there are differences in maternal behavior between tame, aggressive and wild rats. There are a lot of factors that should be considered when evaluating parental behaviors in rodents. The different methods of data collection can provide very different results and may influence on their interpretation [3]. Thus, it is interesting to study the maternal behavior in different rat lines and to compare methods of visual observation and automated video registration.

MATERIALS AND METHODS

Maternal Behavior

The study of maternal behavior was conducted on tame and aggressive female rats. At the age of 4 month each female was placed in cage with male. On the 20th day we put males away. One part of dams was left in their cages for subsequent visual observations. Visual observation of maternal behavior was carried out from 2 to 12 postnatal days. Different behavioral parameters were measured in 5 minutes during one hour twice a day according to standard protocol [4]. Total number of observations for one female was 20 in day. For this experiment 15 tame and 15 aggressive females were used. Another group of pregnant females (8 tame and 8 aggressive females) were placed in special home cages with front wall made of clear plastic for automated video registration. In front of each home cage there was a highly sensitive black-and-white video camera 1/3" CCD with automated diaphragm tuning. Our observation system allows sequentially record image data

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

from different cameras. Switchboard Kramer V016 was used to switch from one camera to another. We supplied each home cage with infrared source for video registration in dark phase of the day. Automated video registration was carried out from 2 to 12 postnatal days during one hour 5 times a day (09.00, 14.00, 18.00, 22.00 and 06.00). Total number of observations for one female was 100 in day. After video recording trained observer watched the video and assessed behavioral parameters. Following parameters of maternal behavior were estimated: mother is in nest, mother licking and grooming pup, and mother nursing pups [4]. Nursing posture was rated as either an arched-back posture when the mother was arched over pups with legs extended, a "blanket" posture in which the mother lays over the pups with no leg extension, or a passive posture in which the mother is lying on her back or side while the pups nurse. Percentage of each behavior from total number of observations was assessed. The percentage of licking and grooming of pups was estimated from number of observations during which the female was inside the nest.

Maternal Aggression Test

Automated video registration system, developed in our laboratory, allows not only prolonged repeated registration during selected time intervals, but also single video registration from selected camera. This is very useful for many behavioral tests that are conducted in home cages. One of such test is maternal aggression test. The study of maternal aggression was made using 10 tame, 10 aggressive rats and 5 rats from unselected wild population. During this test unfamiliar male Wistar rat was placed in home cage of female with pups. The test was held on 6 postnatal days during 5 minutes. The behavior of lactated rats was videotaped, and following behavioral elements [7] were scored using special computer program: aggressive behavior (attack, chasing, kicking, offensive upright, keep down, lateral threat), social explore (moving toward, anogenital sniffing), nonsocial behavior nosing, (locomotion, rearing, self-grooming, sitting, lying). The computer program made in our laboratory allows assessing the latency, time and number of each behavioral pattern. The time of all behavioral elements, the number of aggressive behaviors and the latency of the first attack were measured. The time of maternal care was calculated as total time witch rat spent inside of the nest or carrying pups. The parameters of maternal behavior were analyzed using ANOVA and Kruskal-Wallis ANOVA analysis, a p value of <0.05 was accepted as significant.

RESULTS

Maternal Behavior

Analysis of the results showed that tame females spent less time in nest than aggressive ones. There was no difference in this parameter in tame rats while using automated registration and visual observations. Aggressive females spent less time in their nests in the presence of observer. The same differences were shown for frequency of nursing. It is important that some estimations of maternal care were significantly lower in dark phase of the day than in light phase. Arched-back nursing posture was the main for both tame and aggressive rats. Tame mothers nurse less frequently in arched-back and more frequently in passive posture, than aggressive. Females of both lines had lower frequency of arched-back posture in the presence of human than under automated video registration. These animals demonstrated an increasing of "blanket" posture. The frequency of licking-grooming was lowest in aggressive mothers under visual observations.

Maternal Aggression

Using Kruskal-Wallis ANOVA analysis it was shown that wild females attacked intruders later than tame and aggressive ones. Number of attacks and total time of aggression were significantly lower in wild than in tame and aggressive mothers. The numbers of upright postures, kicking, chasing, pinning and lateral threats were lower in wild rats than in tame. Also the number of upright postures and kicking was higher in aggressive animals than in wild ones. There were no significant differences in behavior of tame and aggressive lactated rats except the number of pinning. Social non-aggressive behavior of wild, tame and aggressive rats didn't differ. The time of "individual behavior" was significantly lower in tame than in wild mothers. Total time of maternal care was higher in wild than in tame and aggressive rats.

DISCUSSION

Our results show that the presence of observer influences the behavior of aggressive females stronger than the behavior of tame ones. Earlier it was shown that aggressive rats are more anxious than tame [8, 9]. Probably, high anxiety of aggressive females can be a reason why they spent more time with pups in the presence of human. The type of nursing and licking-grooming frequency are important factors of maternal care [1, 2]. There are three main nursing postures but the most effective for nutrition of pups is an arched-back posture. Dams of both lines showed increased "blanket" posture during visual observations. Probably, the presence of human could increase the defensive behavior of animals and we suppose that the "blanket" posture was its part. Licking-grooming frequency was lower in the aggressive mothers under visual observations than in other groups. Our results show that the presence of the human affects frequency of lickinggrooming in the aggressive mothers and doesn't influence this parameter in the tame ones. Thus, the presence of observer attenuates maternal care in aggressive rats. It is shown that maternal aggression towards unknown male conspecific is lower in wild rats than in tame and aggressive ones. Low anxiety of tame rats and high aggressiveness of aggressive ones could result in their high maternal aggression. Earlier it was shown that some strains of laboratory rats are characterized by high maternal aggression [6]. It should be mentioned that wild rats demonstrate defensive postures in the presence of human so

automated video registration is necessary for objective study of behavior.

Thus, our results show that genotype-dependent behavioral features of rats can affect manifestations of maternal behavior while using visual observation method. The behavioral patterns of maternal care and maternal aggression are very complicated for automatic identification. Therefore, we used automated video registration of behavior according the table and then analyzed patterns using standard program. But now we are developing special system of automatic monitoring of rat diurnal activity. The method of image analysis is the basis of new approach to this task solution.

- 1. Francis D., Diorio J., Liu D., Meaney M.J. Nongenomic Transmission Across Generations of Maternal Behavior and Stress Responses in the Rat. *Science*, 286 (1999), 1155.
- Liu D., Diorio J., Tannenbaum B., Caldji C., Francis D., Freedman A., Sharma S., Pearson D., Plotsky P.M., Meaney M.J. Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. *Science*, 277 (1997), 1659–1662.111.
- Lonstein J.S., Fleming A.S. Parental Behaviors in Rats and Mice. Current Protocols in Neuroscience UNIT 8.15, 2001.

- 4. Myers M.M., Brunelli S.A., Squire J.M., Shindeldecker R.D., Hofer M.A. Maternal behavior of SHR rats and its relationship to offspring blood pressure. *Dev. Psychobiol.*, *22*, 1 (1989), 29-53.
- Naumenko E.V., Popova N.K., Nikulina E.M., Dygalo N.N., Shishkina G.T., Borodin P.M., Markel A.L. Behavior, adrenocortical activity, and brain monoamines in Norway rats selected for reduced aggressiveness towards man. *Pharmacology biochemistry and behavior*, 33 (1989), 85-91.
- 6. Numan M., Insel T.R. *The Neurobiology of Parental Behavior*, Springer-Verlag New York, Inc, 2003.
- Olivier B., Mos J., van Oorschot R. Maternal aggression in rats: Effects of chlordiazepoxide and fluprazineBerend. *Psychopharmacology*, 86 (1985), 68-76.
- 8. Plyusnina I., Oskina I. Behavioral and adrenocortical responses to open-field test in rats selected for reduced aggressiveness toward humans. *Physiology and behavior*, *61*, 3 (1997), 381-385.
- 9. Plyusnina I.Z., Oskina I.N., Tibeikina M.A., Popova N.K. Cross-fostering effects on weight, exploratory activity, acoustic startle reflex and corticosterone stress response in Norway gray rats selected for elimination and for enhancement of aggressiveness towards human. *Behavior Genetics*, *39*, 2 (2009), 202-212.

Measuring Behavior in Complex Maritime Operations

Heidi Seljestad

John Ferkingstad

Tone Sydnes

Gisle Kleppe

Vidar Frette vidar.frette@hsh.no

Stord/Haugesund University College Bjoernsons gate 45, N-5528 Haugesund, Norway

ABSTRACT

We deal with complexity in maritime operations, through data collection that allows analysis of human behavior and interaction. We have focused on bridge systems, in particular those that incorporate equipment for Dynamic Positioning (DP). Our empirical study has been based on simulator training of professional operators. The training set-up included stress-causing factors, with increased demands on the operators' situation awareness. The imposed situation variables as well as the resulting behavior and performance were recorded by qualitative and quantitative methods.

Author Keywords

Maritime operations, simulator training, offshore vessels, complex tasks and processes, MTO.

INTRODUCTION

On board advanced ships used in fields such as offshore oil exploration, operators are faced with increasingly more complex bridge systems. Most bridge systems today have equipment for Dynamic Positioning (DP).

Dynamic Positioning

DP is a method to keep ships and semi submersible rigs in a fixed position using the propulsion systems instead of anchors. It may also be used for sailing a vessel from one position to another along a predefined route. Like an autopilot on a plane, DP may operate without human interaction. The method relies on accurate determination of position from external reference systems such as the Global Positioning System (GPS).

Some offshore operations involve long standby periods on DP without active participation from the operators. This can

lead to passivity and fatigue, a well-known phenomenon in the sector. Periods of mental underload may reduce the operators' ability to handle demanding situations [3].

Simulator Training

Due to the large risks involved, compulsory simulator training is an integrated part of education for DP operators. We have studied the behavior of participants in regular DP simulator courses. These course participants all had experience from offshore operations. Simulator sessions are what Salas et al. [1] call Synthetic Task Environments, simplified situations that are more accessible to research than the complex real-life cases.

RESEARCH TOPIC

We are interested in the complex interaction between Man, Technology, and Organization (MTO) in demanding situations on a ship's bridge. These situations may arise due to the complexity of the operations, short distance to other vessels or installations, weather and sea conditions, and equipment failures. The information processing perspective, applied to both individual and team behavior, is important in our approach [2].

We have studied how DP operators handle situations with loss of position reference systems during simulator training. The exercise given to the course participants were to take



Figure 1. The figure shows three platforms connected by bridges (in the middle) and an approaching ship (to the right).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 2. Example of maneuvering history.

the vessel from one given position to a new one close to platforms (marked by a cross in Figure 1). At this new position, a subsea survey was carried out. However, the position reference systems were at this point intentionally blocked by the instructor. Simultaneously, the wind speed was increased. The operators needed to keep the vessel in position during the survey. We monitored, qualitatively and quantitatively, how the operators handled this emergency situation.

DATA COLLECTION

The operators' reactions to the position reference fall-out and changed weather conditions were recorded by three methods: qualitative observations, post-session interviews, and maneuvering history recorded by the simulator software. Only course participants that gave their consent, were included in the experiments.

Observations

During each simulator session (approximately 1 hour) we observed continuously the bridge team. We looked for changed behavior as the new, demanding conditions were introduced: In some cases, communication in the team was reduced, in others we observed reduced small-talk and more professional discussions. We also observed changes in body language, for instance, one participant repeatedly lifted his cap.

Interviews

After the simulator session, we carried out interviews with the participants. We asked how they evaluated their own performance and the problem-solving process. Their evaluation of simulator training, as compared to real-life situations, was also of interest. Furthermore, we asked them to describe their subjective experience of stress. In addition, we interviewed the simulator instructors. We asked them, as experts, to evaluate the quality of the job performance of each group.

All interviews were recorded and transcribed, provided that informed consent was given.

Maneuvering History

The simulator software continuously recorded vessel position, heading, instantaneous engine power, speed, rudder direction etc; as well as environmental parameters like speed and direction of wind and current.

An example is shown in Figure 2, where position, engine power and wind speed are shown. Horizontal position curves correspond to the period when the vessel conducted a subsea survey near the platforms, see Figure 1.

DATA PROCESSING

The different kinds of data described above give information about several aspects of the exercise process.

The recorded maneuvering history shows objectively the series of choices taken by the team, but not the strategic thinking behind. The observations, on the other hand, indicate the intentions inherent in the actions taken.

The interview with participants reflects their subjective understanding of the process as well as the performance of the team. A different view may be obtained from the interview with instructors. These interviews form a basis for a more objective expert evaluation of the task performance.

From the data both independent variables (how demanding was the task?) and dependent variables (how well was the task performed?) were extracted. The difficulty of keeping the vessel in position during the subsea survey was varied

for instance through the wind speed increase imposed. Measures for the quality of performance were obtained from expert interviews, duration of the various stages in the operation, and various observations.

CONCLUSION

While human behavior under demanding situations has been studied extensively in the context of air traffic and nuclear power plant operation, less has been done in the maritime sector.

A combination of qualitative and quantitative methods was used to shed light on the man-machine relations in complex maritime operations. Participants at simulator sessions were used as study objects. Video recordings of the sessions will be considered in the future.

- 1. Salas, E., Cooke, N. J., Rosen, M. A. On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors*, 50, 3 (2008), 540-547.
- 2. Wickens, C. D. and Hollands, J. G. Engineering Psychology and Human Performance. Prentice Hall, Upper Saddle River, NJ, 1999.
- 3. Young, M. S. and Stanton, N. A. Attention and Automation: new perspective on mental underload and performance. *Theor. Issues in Ergon. Sci.*, *3*, 2 (2002), 178-194.
Diagnostics of Motivational Conflicts Through Thinking Activity

Anastasia Karpukhina nastasyakarp@bk.ru +7 910 422 6472 Olga Arestova arestova@mail.ru

Moscow State University Russia, 125009, Moscow, Mokhovaya, 11/9

ABSTRACT

This paper is devoted to the battery of projective techniques which allows to investigate tension and inconsistency of motivation structure through particularities of thinking.

Author Keywords

Thinking, motivation, emotional regulation, motivational conflict, generalization, projective techniques, conceptualization, structuring function of motives.

ACM Classification Keywords

F.4.2 Decision problems, H.1.2 Human information proceeding, I.2.3 Deduction and theorem proving, I.2.4 Knowledge representation formalisms and methods, I.2.6 Learning

DIAGNOSTICS OF MOTIVATIONAL CONFLICTS THROUGH THINKING ACTIVITY

Discovery and exploring of the structuring function of motives is one of the main achievements of the O.K. Tikhomirov's psychological school. It is now followed by a range of investigations dealing with the interrelationship between personal and cognitive, intellectual processes. In some cases thinking peculiarities are conditioned by motivation, which may have organizing or destructive influence on thinking process [1].

When the problem, which the subject is solving, is related to the zone of current motivational conflict, appears the phenomenon of rational distortion [2]. To investigate this phenomenon we use a special battery of techniques. Applying this battery allows to make conclusion about the content and intensity of the motivational conflicts, and make prognosis about preferable solution of the subject. The battery includes eight techniques: Sentence Completion Test (SCT), Complicated Analogies Test, Odd Fourth Test (classification test), Pictograms Test, The Family Drawing Test, The Non-Existent Animal Drawing Test, Luscher Color Test and Proverbs Interpretation Technique.

Sentence Completion Test is one of the widespread projective techniques. It is a descendant of Free Association Technique. Verbal form of stimulus is considered to determine subject's responses more than drawings or inkblots. One of the main advantages of SCT is an opportunity to modify it due to the researcher's needs. Different variants of the test are based on different approaches of their authors. In Russia the most popular variant for working with adults is the test made by J.M. Sacks and S. Levy. Unfortunately, the last translation and adaptation of the test was made more than 20 years ago, and some of incomplete sentences lost their applicability. That's why in the current research we use modified variant of SCT, where several sentences are reformulated without change of their topic. The scheme of data retrieval also differs from the "classical" one: not only content is taken into account, but also such parameters as handwriting and grammar particularities, time of the response and nonverbal emotional expression are considered. Such type of analysis combines standard scheme and clinical approach. Applying of SCT helps to elicit areas of personal problems, which may stay unconscious or not completely conscious, and make conclusions about personal traits such as egocentricity, adaptiveness, mental and social maturity and some others.

SCT is conducted at the beginning of the survey because its data allows to develop hypotheses about current motivational conflicts of a person. These hypotheses can be clarified during the clinical conversation. This helps to select cues for Pictograms and Proverbs Interpretation Test.

Complicated Analogies Test is used for checking ability of understanding logical connection, observing distinctions between different types of logical relations and transferring it from one group of elements to another [3]. The test consists of 30 pairs of words and 6 pairs that mark 6 categories of relations. The subject has to classify the pairs

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

of words according to the type of relation between them. This test helps to make a selection of subjects who are able to solve tasks, which require highly developed conceptuality (abstract thinking).

The Odd Fourth Test is designed for checking abilities of generalization and abstraction [3]. The cues of this test are pictures with the groups of four things (e.g. scissors, threads, thimble and pipe). The subject is asked to choose an odd thing and explain his choice. To eliminate "the odd fourth" the subject needs to make a group of three things. Many pictures provide opportunity for two or more answers based on different principles of generalization: conception thinking will lead to other classification than thinking in chain complexes etc.

We should also mention that the result of this test is sometimes influenced by the personal meanings of the pictures. It happens when subject somehow relates the content of the picture with his personal experience. Such phenomenon of rational distortion is a manifestation of current motivational conflict.

Initially Pictograms Test was a method for investigation thinking and memory and only afterwards it developed as a projective technique [4]. This method reflects emotional traits and current state of the person, his meaningful themes and cognitive properties.

The test is introduced to the subject as a memory control. He needs to remember 12-16 words and phrases, which are dictated to him. He is also given with a pencil and a sheet of paper. Instruction says that he should draw pictures which help him to remember given phrases, but he cannot write down anything, only drawing is permitted. Experimenter may vary the complexity of the cues from concrete things to abstract ideas. Also the test may include phrases related to the current problems and motivational conflicts of a person for whom the list of stimuli is made.

When the subject completes the last drawing, the paper is taken away and during the next hour he makes easy tasks not related to the test. After that the experimenter returns him the paper and asks to remember the phrases. The lacunas and mistakes in retrieval are taken into consideration, but the most interesting part is the drawings themselves. In the final part the subject is asked to tell what he has drawn and how it helped him.

Obtaining the data means analyzing content and appropriateness of drawings, whether they are abstract or concrete, and how do they relate to the life events of their author, special attention is paid to the rare and unusual images. The drawings may be processed according to the main principles of drawing methods, such as "Drawing a Man" technique. Another aspect of analysis is the reference of the conclusions to the data obtained from the other techniques. The Family Drawing Test, The Non-Existent Animal Drawing Test and Luscher Color Test can be used to fill up the gap between drawing and retrieval in the Pictograms test.

The applying of Luscher Color Test on the former Soviet area has a long story. In Russia currently most of psychologists use a modified version, which is the second subtest of the original one. This test provides a quick opportunity to get information without disclosing anything to the subject. It is an effective instrument for supporting hypothesis about the current state of a subject and his personal traits.

The Family Drawing Test is focused on the family relationships. Adult subjects make a drawing on the standard sheet of paper using an ordinary pencil. Processing includes comparison between the drawn family and the real amount of family members, the type of activity of the people on the picture and the order of drawing family members. When the drawing is completed, the conversation is held: it is necessary to find out who is present at the drawing, what he is doing etc. Usually the conversation helps to clarify the subject's attitude toward family members. Figures may also be examined according to the "Drawing a Man" principles of analysis.

The Non-Existent Animal Drawing Test was initially made for the work with children, but later spread as an effective psychodiagnostic method for adults. The subject has to think of imaginary animal, draw it and tell a fairy tale about his life. The subject is expected to identify himself with the animal, and metaphorical character of the image helps to display the features which remain blurred in more straightforward tests.

Application of proverbs in psychological research has a long history. In psychology of thinking proverbs are traditionally used to examine thinking, in particular to find out what levels of conceptualization are apprehensible for the subject [5]. Using proverbs in speech allows to convey the meaning without naming the situation, and the same proverb may become both blame and excuse [6]. Other important function of a proverb is regulating: its authority as a "folk wisdom" helps subject to handle with difficulties, to direct his thoughts and activity.

It was investigated that the interpretation, which the subject gives to the proverb, may sometimes reflect the current motivational conflict. Elements of particular interest are cases for rational distortion, such as egocentric interpretation, aggression, evaluative speech and some others. In spite of understanding the proverb correctly, the subject may give the bizarre interpretation, which has no relation with the wide-spread and commonly accepted one.

Applying Proverbs Interpretation Technique for investigating motivational conflicts differs from a traditional way of using it: here we discuss dynamics of meaning and conceptualization processes. Other important advantage of the technique is a variety of proverbs and their content which provides the opportunity to choose proverbs for every subject, according to the hypothesis of the motivation conflicts and contradictions which are currently important for him. Proverbs are relatively seldom used in everyday speech, but, according to the paroemiological studies, it doesn't prevent the subject from understanding the meaning even when they meet an unknown proverb inbuilt into the context, when he can find relevant situations in his own experience.

Further we give an example of a research made with the abovementioned techniques. This research was made to investigate the phenomenon of identification, which manifested in the Proverbs Interpretation Technique. This phenomenon appears when subject switches from solving a given task to the attempt of solving his own conflict, and creates a personage - participant or experiencer of this conflict. Conditions of the task and the way of solving reflect the dynamics of inner contradiction and conflict process. Personage - participant or experiencer of the conflict may vary from well-developed (for example, nonexistent animal on the drawing and in the story) to undetermined "somebody", described with a pronoun: "he", "someone" etc. In the interpretation of proverb or a drawing the subject often uses a hero of this proverb or drawing and enriches him with his own senses and feelings. Such process of identification in TAT was described by H.Murray [7]. Here we see how objective logic is substituted by a subjective process of meaning production. It is externalized in the decrease of the conceptualization level, distortion in understanding the problem conditions, change of emotional expression and other features. Proverbs Interpretation Technique is a perfect material for clarifying this phenomenon due to the dichotomical structure of the proverbs and culture-determined way of using proverbs, which stimulates subject to refer the situation described in the proverb with the events of his own life.

Our research included two parts. The first, preparatory part was held with two purposes: examination of the thinking of test subjects to select the ones who are able to complete tasks with the high level of conceptualization. For that purpose we used Sentence Completion Test, Complicated Analogies Test, Odd Fourth Test and Pictograms Test. The second purpose was to determine the current motivational conflicts of a person. That was made according to the data of Sentence Completion Test, Pictograms Test, drawings and Luscher Color Test.

The second part was conducted with the selected subjects, whose ability to cope with difficult tasks such as interpretation of the proverbs was checked during the preparatory part. Proverbs given for interpretation were especially selected due to the current motivational conflicts of every subject, in order to provoke identification. Another part of the proverbs was common for all subjects which allowed making conclusions about the difference in provocative strength of different proverbs.

In our research we used common Russian proverbs which are included into vocabularies for schools. All the volunteers were raised in Russia and were expected to have the educational level high enough to cope with all the tests. The amount of common proverbs was 10, and 10 other proverbs were selected especially for every subject. The cases of identification (170 cases) were pointed out according to the main criterion: creation of the personage, and compared to the results of the other techniques and the interviews.

Cases of identification were classified according to the intensity and content of interpretation discourse. It was discovered that on different stages of conflict different kinds of identification appeared.

Many authors emphasize that motivational conflict is a complicated phenomenon which sometimes cannot be investigated using self-report or questionnaires, especially before it slips or turns to the critical stage. The choice between two alternatives is painful for the subject, and even after the moment when the decision is made, it can be denied. The proverbs provide the material for the safe embodiment of the decision and for the evaluating of its result. The relationship between the conflict situation and the proverb may vary: proverb may support or belittle one of the alternatives, may embody the struggle of two motives or reinforce the chosen one which looks like a conclusion or a lesson.

The existence of the motivational conflict usually begins from the unconscious contradiction between two motives of approximately equal strength. Sometimes the subject is aware of one motive, while the second remains unclear. At this period the sense of a proverb may be changed to the opposite during interpretation, without clarifying the alternatives. Two main tendencies of this period are aggravation and smoothing of the contradiction: the subject may be ready to develop the conflict or he can try to escape the unpleasant necessity of making a decision. In the case when one of the alternatives remains unconscious and for some reasons cannot be realized, there appears the phenomenon of anti-identification, or denied identification [7]. That means that the subject depicts the personage which looks extremely disagreeable and distasteful from his point of view. This personage embodies the underlying motive which the subject cannot accept. Denied identification is related to the sour grapes mechanism which can be classified as a defense activity.

When the contradiction is aggravated to the conflict, the critical stage begins. At this stage the interpretation of the proverb related to the conflict looks like the embodiment of the struggling motives in the heroes of the proverb or in the personages of the situation which the subject describes. Often the proverb which looks like an impersonal phrase becomes the reason to describe the subject's experience,

and the heroes appear from his own discourse. In any case interpretation at the critical stage contains two opposite tendencies. The subject shows the signs of strong emotional tension, which seems inadequate to the simple, childish task.

At the final stage, when the decision is already made, the subject tries to support the chosen alternative and diminish the importance of the refused one. It is often mentioned then that the proverbs have a reputation for wisdom. The subject may agree or disagree with the proverb, regarding it as an advice which is good or bad for him. Identification at the final stage appears when the fact of the conflict eliminates from the conscious part of the subject's mind. In this case actions and feelings of the personage are described in the past.

Besides the current stage of the motivational conflict, the interpretation is influenced by the characteristics of the proverb itself. Such characteristics are: emotional richness of the images of the proverb; its resemblance to reality or its fantastic content; mention of certain personages (e.g. animals, people, things) or impersonal phrase; description of an attitude or an activity; encouragement or warning against the certain type of activity.

It is important to separate cases of rational distortion from the incapacity of abstract thinking, and it refers not only to the proverbial interpretation, but to many types of tasks which may somehow involve the motives of the subject. For example, cases of identification resemble concrete thinking. To make a difference it is necessary to take into consideration the results of several tasks related to different topics and pay attention to the emotional reaction of the subject.

- Арестова О.Н. Развитие представлений о роли мотивации в мыслительной деятельности. Вестник Московского университета, 2 (2008) [Arestova O.N. Razvitie predstavlenij o roli motivacii v myslitel'noj dejatel'nosti. Vestnik Moskovskogo universiteta, 2 (2008)]
- Арестова О.Н. Аффективные искажения в понимании пословиц. Вопросы психологии, 1 (2005) [Arestova O.N. Affektivnye iskazhenija v ponimanii poslovic. Voprosy psihologii, 1 (2005)]
- Рубинштейн С.Я. Экспериментальные методики патопсихологии. Изд-во Института психотерапии, Москва, Россия, 2007 [Rubinshtejn S. Eksperimental'nye metodiki patopsihologii. Izd-vo Instituta psihoterapii, Moscow, Russia, 2007]
- Херсонский Б.Г. Психодиагностика мышления. Студия «Негоциант», Одесса, Россия, 2003, 96-147 [Hersonskij B.G. Psihodiagnostika myshlenija. Studija «Negociant», Odessa, Ukraine, 2003, 96-147]
- Беломестнова Н.В. Клиническая диагностика интеллекта. Речь, Санкт-Петербург, Россия, 2003, 40-52 [Belomestnova N.V. Klinicheskaja diagnostika intellekta. Rech', Saint-Petersburg, Russia, 2003, 40-52]
- 6. Пермяков Г.Л. Основы структурной паремиологии. Наука, Москва, Россия, 1988 [Permjakov G.L. Osnovy strukturnoj paremiologii. Nauka, Moscow, Russia, 1988]
- 7. Murray H.A. Thematic Apperception Test Manual. Cambridge, USA, 1943.

Use of Ethological Rodent Behavior to Assess Efficacy of Potential Drugs for Alzheimer's Disease

Robin Roof Robin.Roof@pfizer.com

Bradley Snyder Bradley.Snyder@pfizer.com Yoko Yabe Yoko.Yabe@pfizer.com

Margaret Zaleska Margaret.Zaleska@pfizer.com

In vivo neurobiology, Neurodegeneration Unit, Pfizer R&D Eastern Pointe Road, Groton, Connecticut 06340, USA

ABSTRACT

Nest construction is a natural, species-typical behavior in rodents [1,2] that may be considered analogous to the "activities of daily living" disrupted in Alzheimer's disease (AD) [3]. Assessment of species-typical home cage behaviors such as nesting in transgenic mice may aid in identifying drugs with potential for improving functional status in AD patients [4]. This strategy provides an alternative or addition to the standard preclinical assessment of cognitive function via maze learning. Most cognitive deficits in AD related transgenic mice occur later in life and are very time consuming to assess. Use of nesting would be a novel screening method. We used two methods to assess nest building; a rating scale based on subjective assessment of the nest quality and an objective grid counting system. We placed individual mice in cages with measured amounts of corn cob and an additional bedding material (Alpha Dri). Mice were left overnight and cages photographed the following morning for subsequent blind assessment of nest quality. Wildtype mice typically separated out and formed nests from the Alpha Dri material. This behavior was less evident in Tg2576 mice, which model AD-type amyloid beta-protein (AB) neuropathology and in tg4510 mice, which carry a mutated human tau gene implicated in the fronto-temporal dementia. Since deficits in most behaviors do not occur until 6-18 months of age, using this early appearing behavior is advantageous.

Author Keywords

Ethological behavioral assessment, nest building, Alzheimer's disease, transgenic, amyloid, tau.

INTRODUCTION

Several types of non-cognitive behavioral impairments occur in AD, including psychiatric symptoms (delusions, hallucinations, anxiety and depression), dysfunctional social behavior (aggression, lack of social interaction), and other behavioral deficiencies that fall into the category of "activities of daily living". Certain species-typical rodent behaviors, including nest construction, burrowing, grooming and open field behavior, might be considered to be rodent versions of the "activities of daily living" that deteriorate in AD. Nest building in rodents involves active interaction with the environment. It is a species typical behavior that is exhibited by both males and females. Construction of a nest is a complex behavior, requiring step by step organization. Nesting is disrupted in some ADrelevant transgenic models [4] and has shown to be mediated in part by hippocampal activity [5], a brain region central in the neuropathology of AD. Assessment of such behaviors in amyloid or tau-based transgenic mouse models may be a novel and useful tool in preclinical assessment of potential AD treatments.

METHODS

We assessed quality of nest construction in two transgenic mouse lines that each serve as a model for a key component of AD neuropathology. Female 129/tg2576 mice, which express the human ß-amyloid precursor protein with the Swedish mutation (APP695SWE) and their wildtype counterparts were tested at 4.5, 7, 9 and 14 months of age. Male and female tg4510 mice, which express human tau containing the fronto-temporal dementia-associated P301L mutation, were tested at 6 months of age. Wildtype counterparts were tested as well.

We acclimated all animals to our facility and home cages for least two weeks prior to the nesting experiment. For testing, we placed individual mice in micro-isolator cages (18 cm W x 19 cm L x 12 cm H) filled with 300g of Corn Cob and 100g of Alpha Dry (WE Fisher & Sons, Inc) approximately one hour before the dark phase (6:00 pm)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Example of before and after nest construction.

with food and water available, in a quiet testing room. Cages were left untouched throughout the overnight experiment. Individual mice were carefully removed from cages approximately one hour after the light phase (6:00 a.m.). We photographed each cage (aerial view) to allow for later assessment (as shown in Figure 1). We used two methods to quantify nest quality. For the first method we used a subjective rating scale of 1-5 based on visual observation of cage photographs as shown in Figure 2 (right panel), the number of grid segments cleared (=1 pt per segment) or partially cleared (=0.5 pts per segment) of the Alpha Dri nesting material was quantified. In both assessments, photographs were scored by investigator blinded to genotyping.

RESULTS

We observed statistically significant disruption of nest building in mice from both AD-related transgenic models tested. Nests built by transgenic mice were of reduced quality as indicated by significant differences in nesting scores compared to their wildtype counterparts. These deficits were observed in 6 month old Tg4510 vs. WT



Figure 2. Nesting scale and Grid measure. The left panel shows nesting scale applied. Small white squares represent pieces of Alpha Dri material and brown background represents corn cob bedding. Low score is associated with highly distributed pattern, high score when alpha dri is concentrated in one place. The right panel shows grid overlying cage. Score is based on the number of grid boxes cleared (1 pt) or partially cleared (0.5 pts) of alpha dri.



Figure 3. Nest building deficit in 7 month old female tg2576 mice as measured by nesting scale shown in Figure 2.

males: $\underline{t}(37)=3.65$, p < 0.001; 6m old Tg4510 vs. WT females: $\underline{t}(30)=2.42$, p <0.05) and in Tg2576 vs. WT female mice at 4.5 ($\underline{t}(14)=2.65$, p<0.05); 7 ($\underline{t}(24)=4.34$, p<0.001) 9 ($\underline{t}(19)=2.7$, p <0.05) and 14 months of age ($\underline{t}(34)=3.7$, p<0.001). Results for 7 month old female tg2576 and wildtype mice are shown in Figure 3. Male tg2576 mice were not assessed.

CONCLUSIONS

Nest building is a behavior that can be easily measured and quantified. Using the assessment methods described, clear differences between wildtype and both tg4510 and tg2576 mice were detected. In our study, wildtype mice typically produced well formed nests by separating individual Alpha Dri particles from the corn cob bedding, transferring them to one end or corner of the cage and piling them in a mound. Transgenic mice tended to either not build nests or to build nests of lesser quality. Reduced nesting in ADrelated transgenic mice may be due to deficits in organized behavior, or to loss of initiative, both symptoms common to AD patients. Measurement of naturalistic behaviors may provide a way of preclinically assessing the potential effects of new medicines on functional outcome in AD patients.

Ethical Statement

The studies were reviewed and approved by the institutional animal care and use committee of Pfizer Global Research and Development. The principles and procedures were conducted in accordance with the NIH Guidelines for the Care and Use of Laboratory Animals.

REFERENCES

1. Deacon, R.M.J. Assessing nest building in mice. *Nat Prot* 1(3), (2006), 1117-1119.

- Filali, M. and Lalonde, R. Age-related cognitive decline and nesting behavior in an APPswe/PS1 bigenic model of Alzheimer's disease. *Brain Res* 1292, (2009), 93-99.
- Galasko, D., Schmitt, F., Thomas, R., Jin, S, and Bennett, D. Detailed assessment of activities of daily living in moderate to severe Alzheimer's disease. J. Inter. Neuropsych. Soc., 11, 4, (2008), 446-453.
- 4. Deacon, R.M.J. Age-dependent and independent deficits in tg2576 mice. *Behav. Brain Res.*, *189* (2008) 126-138.
- 5. Deacon, R.M.J., Croucher, A., and Rawlins, J.N.P. Hippocampal cytotoxic lesion effects on species-typical behaviors in mice. *Behav. Brain Res.*, *132* (2002), 203-213..

Who Said Pedestrian Means Dull: Teaching Methods in the Measurement of Behaviour Through Observational Studies of Pedestrian Behaviour

Richard E. Brown rebrown@dal.ca **Timothy P. O'Leary**

Stephanie L. Allen

Dalhousie University, Halifax, NS, Canada, B3H 4J1

ABSTRACT

The Measuring Behaviour class (PSYO 3161) at Dalhousie University is designed to teach third year undergraduate students the theoretical and methodological procedures for the measurement of behaviour. For this class, students complete laboratory projects on the behaviour of mice and humans. For one of these projects, students completed observational studies on the behaviour of pedestrians at cross-walks. Students worked in groups to design an experiment to answer a question concerning pedestrian behaviour, and by doing so applied knowledge learned in the class on experimental design, sampling rules, objective definition of behavioural measures and statistical analysis. The students collected data, analyzed the data and then completed written reports describing the results. This poster describes some of the studies completed by students in the Measuring Behaviour class, and demonstrates the value of pedestrian behaviour as an educational tool for classes designed to teach methods for measuring behaviour.

Author Keywords

Teaching, pedestrian behaviour.

INTRODUCTION

The theme of pedestrian behaviour is an effective topic of research in an undergraduate class as (1) students can easily relate to pedestrian behaviour, (2) there is an established field of research and many journals publish articles on pedestrian behaviour (3) data on pedestrian behaviour can be quickly obtained and (4) students become motivated based on the growing need for research on pedestrian behaviour. Indeed, pedestrians represent the largest group of road users today [3] and 20-35% of pedestrians cross illegally at intersections [2]. Moreover, the high rates of illegal pedestrian activity increase the potential for

pedestrian-car accidents, which account for 12% of traffic related injuries in the United States of America [1]. As such, research completed by students may help to determine the potential causes and correlates of risky pedestrian crossing behaviour.

EXPERIMENT 1: DOES A COUNT-DOWN TIMER SIGNAL INFLUENCE CROSSING BEHAVIOUR?

Background

Countdown timers are used to indicate to pedestrians how much time is left to cross the street before the light will change. This experiment was designed to determine if pedestrians used the count-down timer to regulate their walking speed when crossing the street. Students hypothesized that pedestrians would be faster to cross when the countdown timer was present, than when a walk signal was present. Furthermore, students expected that there would be a positive correlation between the time required to cross the street and the time on the count-down timer.

Methods

There were 100 pedestrians observed at an intersection in Halifax, Nova Scotia, Canada from 3:30-5:30 pm. The intersection was monitored by a signal-light that displayed a walk signal followed by a countdown timer. The time on the count-down timer when the pedestrian started crossing the street was recorded, and the time to cross the street was measured.

Results

Pedestrians took significantly less time to cross the street when the count-down timer was present, than when the walk-signal was present (t(98) = 4.30, p < 0.0001, Figure 1A). Also, there was a significant positive correlation between the time exhibited on the count-down timer and time to cross the street (r(41) = .612, p < .001, $r^2 = .374$, Figure 1B). This suggests that pedestrians used the countdown timer to regulate how fast they walked to cross the street, and tended to cross faster when the count-down timer approached zero. Further research is needed to determine if this increase in crossing speed is accompanied by more risk assessment behaviours (looking for on-coming traffic), which may decrease the risk of car-pedestrian accidents.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. (A) The latency (mean, +/- SEM) to cross the street for pedestrians who started crossing when the count-down timer was present was shorter than pedestrians who started crossing when the walk signal was present. (B) The time spent crossing the street was positively correlated with time on the count-down timer.

EXPERIMENT 2: DOES THE USE OF PERSONAL ELECTRONIC DEVICES INFLUENCE TIME TO CROSS THE STREET?

Background

The use of cell-phones and other electronic devices by drivers has been found to increase the risk of car-related accidents [4]. The goal of this study was to determine if the use of electronic devices also influences the behaviour of pedestrians at cross-walks. The students hypothesized that pedestrians using electronic devices (cell phones and portable music devices) would take longer to cross the street than pedestrians not using such devices.

Methods

Pedestrians (N=50) were observed at a signalized intersection from 11:30 am - 2:30 pm. The time required to cross the street and whether the pedestrian was distracted (i.e., texting/talking on a cell-phone or using headphones) or not distracted was recorded.

Results

Pedestrians who were distracted took significantly longer to cross the street (t(24) = 7.14, p < .0001, Figure 2) than pedestrians who were not distracted. This increase in time required to cross the street may be due to impaired attention in pedestrians using electronic devices, as they may not be



Figure 2. The mean (+/- SEM) latency to cross the street for pedestrians who were distracted was longer than pedestrians who were not distracted.

attending to signal lights, traffic or other pedestrians. Furthermore, because distracted pedestrians spend more time crossing the street, they may be at an increased risk for a pedestrian-car accident.

SUMMARY

Results from these observational studies completed by students of the 3161 Measuring Behaviour class provide information regarding factors that influence street crossing behaviour by pedestrians. The data suggest that pedestrians attend to count-down timers on signalized cross-walks, and regulate crossing speed based on the time of the countdown timer. Also, portable electronic devices were found to slow crossing speed. This poster demonstrates that pedestrian behaviour can be used as an effective area of research for classes designed to teach the theory and methodology for the measurement of behaviour.

ACKNOWLEDGEMENTS

Thanks to the students of the 3161 measuring behaviour class, especially those who completed experiment 1 (Jean Pallaprat, Troy Young), and experiment 2 (Abby Poirier, and Emily Nguyen).

- Bungum, T.J., Day, C. and Henry, J.L. The association of distraction and caution displayed by pedestrians at a lighted crosswalk. *Journal of Community Health*, 30 (2005), 269-275.
- Keegan, O. and O' Mahoney, M. Modifying pedestrian behaviour. *Transportation Research Part A*, 37 (2003), 889-901.
- King, M.J., Soole, D. and Ghafourian, A. Illegal pedestrian crossing at signalised intersections: Incidence and relative risk. *Accident Analysis and Prevention*, 41 (2009), 485-490.
- Redelmeier, D.A. and Tibshirani, R.J. Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine*, 336 (1997), 453-458.

Development of Behavioral Measures of Osteoarthritisinduced Pain in Rabbits

Anna Vardanyan

anna.vardanyan@genzyme.com

Bryan Mastis bryan.mastis@genzyme.com Nance Moran nance.moran@genzyme.com

Joseph Serriello joseph.serriello@genzyme.com Gloria Matthews gloria.matthews@genzyme.com

Genzyme Corporation, 49 New York Ave, Framingham, MA, 01702, USA

ABSTRACT

Osteoarthritis (OA) is a painful and disabling disease affecting millions of patients worldwide. The existing drug therapies for OA reduce pain, but are only moderately effective. The broad objective of our work is to develop OA pain models to both better understand the pathogenesis of OA-induced pain and assess potential analgesic agents. The specific objective of the work described is to establish behavioral methods to measure OA induced spontaneous pain in rabbits, species that are not commonly used for pain studies, using weight bearing and rearing assessments. Weight bearing was evaluated by placing rabbits in a chamber, specifically designed so each hind paw was resting on a separate transducer pad, to accurately measure the difference in weight distribution between ipsilateral and contralateral limbs. Rearing assessment was accomplished by placing rabbits in spacious chambers, where animals could freely move, and measuring the number of rearing episodes over a fixed period of time. This allowed utilizing a well characterized surgical model of OA in rabbits anterior cruciate ligament transection (ACLT) that closely resembles post-traumatic OA in humans, as well as taking advantage of relatively large joint volume for assessment of intra-articularly delivered therapies.

Author Keywords

Osteoarthritis, pain, osteoarthritis-induced pain, behavioral measure, ACLT, rabbit, weight bearing, rearing.

INTRODUCTION

Despite exciting progress in the understanding of molecular and cellular mechanism of osteoarthritis, chronic pain which is the most prominent symptom of this disease remains a challenging entity to treat. This in part owes to lack of optimal models and assays for measuring OAinduced pain in animals. To date, most of preclinical studies use rat models and evoked stimuli to measure osteoarthritisinduced pain. Here, we provide an analysis of pain in rabbits – species that are not commonly used for behavioral studies of pain. We have utilized anterior cruciate ligament transection (ACLT) - surgical model of osteoarthritis, which closely mimics pos-traumatic OA in humans. To assess behavior signs of pain we have used weight bearing and rearing assays. These assays allow to measure ongoing, spontaneous pain – a common clinical complaint and an important target for new drug development.

METHODS

To model OA, rabbits were anesthetized and after skin incision, joint was fully exposed, ACL was identified, apprehended with a micro-dissecting hook, and transected. The procedure predictably resulted in chondrocyte loss and lesion formation, characteristic of OA, which was confirmed histologically. A separate group of animals underwent sham surgery. To assess the development of pain-like behavior in rabbits we have utilized weight bearing (WB) and rearing tests. WB was measured using an incapacitance tester, where rabbits were placed in the holder specially designed to maintain the animal comfortably positioned on two separated sensor plates. In the absence of hind paw injury animals apply equal weight on both hind paws, indicating a postural equilibrium. After unilateral hind paw tissue injury, a change in the weight distribution on the sensor can be detected, with less weight applied by the injured paw. After acclimation in the chamber for 5-10 min, five consecutive readings that did not require a shift in animal position or any other manipulation were taken. Additionally, to measure changes in rabbit activity in response to joint injury, rabbits were placed in an enclosed chamber $(2.5 \times 2.5 \text{ ft.})$ and the number of rearing episodes was recorded over a 15 min period of time.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

RESULTS AND CONCLUSIONS

Experimental results from WB analysis demonstrate a clear differentiation between naïve and OA animals that is statistically significant, thus indicating that weight bearing is an accurate and reproducible method that can be used to measure pain-like behavior in rabbits. On the contrary, rearing assessments were not able to detect differences between ACLT and sham groups and concluded that rearing is not a sensitive enough tool for the analysis of behavioral signs of pain in rabbits. Pharmacologic characterization of the validity of weight bearing for the detection of pain with commonly used analgesics is in progress.

Ethical Statement

Male New Zealand White rabbits (~3000 g at the start of the study) housed individually in ventilated cages with free access to food and water, were used for this study. Rabbits were maintained in temperature and humidity controlled animal rooms on a 12 h light/dark cycle in an Association for Assessment and Accreditation of Laboratory Animal Care accredited facility. All experiments have been approved by the Institutional Animal Care and Use Committee at Genzyme Corporation and are in strict accordance with the ethical guidelines laid down by the International Association for Study of Pain.

Progress With MINDS, a Testmanager for Psychological Assessment, Research and Education. Applications in the Forensic Psychiatric Domain

A. N. Brand nbrand@mindsware.nl A. K. L. von BorriesB. Ik.v.borries@pompestichting.nle.bulten@p

B. H. Bulten e.bulten@pompestichting.nl

bureau MindsWare Van Dam van Isseltweg 7, 4191 KC Geldermalsen, The Netherlands +31-345-570842 Pompestichting, Postbus 31435 6503 CK Nijmegen, The Netherlands

ABSTRACT

Testmanager MINDS is a program package running on Windows platforms, for administration of automated neuropsychological tests and questionnaires for use in several assessment settings, in research and for use in education. Administration of a test and report (or further processing) of test results is organized by separate modules. Continuously, test- and other modules, tables of norms are added to the package. Five newly developed test modules, specifically applied for use in forensic psychiatric settings, are presented. Applications of interest in this field particularly consist of tests for processing emotional material e.g. recognition of emotional expressions, and tests appealing to frontal lobe functions such as impulse control, planning and concept shifting. Instruments for assessment in these domains have been integrated into a testbattery ("forMINDS"). Preliminary results, comparing the performance of intramural forensic psychiatric patients with that of healthy control subjects, are presented at the congress.

INTRODUCTION

MINDS encompasses a growing number of automated neuropsychological tests and questionnaires. Its development started in the early nineties [1,2] in order to offer a flexible tool for use in psychological assessment, research and education. The program can be applied in several types of psychological settings with various target populations, e.g. child care institutions, psychiatric settings and other mental health care institutions. For this use most test modules have been supplied with several tables of norms, tables that can be adjusted by the publishing bureau or by the user. In addition, Minds can be used in specific research projects, since every test module has been equipped with a facility to quickly aggregate data into an SPSS file. Finally, Minds is currently being used in several practically-oriented psychology courses, where it runs on a network for student computers.

Recently, there is growing interest in Minds from forensic psychiatric settings. Applications of interest for work with the forensic population particularly consist of tests for the processing of emotional material, tests appealing to frontal lobe functions e.g. impulse control, planning and concept shifting, and tests for moral reasoning. This especially appeals to function domains which are problematic in borderline and antisocial personality disorder [3]. For that reason, several applications were developed and added to the assortment of performance test modules in Minds. These test modules are presented below.

The organization of Minds is quite flexible and intuitive. Menus for performance tests, questionnaires and other system parts can be adapted in order to suit the users' demands. Test modules and other parts are represented by pictograms that can be made visible or invisible at will. In principle, the administration of a test module is apart from the presentation of its outcome, by way of a separate report module. In addition, most test modules in Minds have a number of task parameters (e.g. time parameters, number of trials, etc), which can easily be adjusted. Most questionnaires in Minds are administered by a common program (QUEST) and outcomes are likewise reported by a common report module (REPORT). In addition, Minds is standardly equipped with a "wizard" for implementing a new questionnaire.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

TEST MODULES FOR THE FORENSIC SETTING

The following test modules developed for diagnostic use with forensic populations (embedded in a testbattery called forMinds) are currently being used in several forensic clinics to evaluate their usefulness in treatment planning and monitoring of treatment effects.

AFFGO, an Affective Go-Nogo Task

In this task the participant is only required to react to stimuli with a predefined affective load and to restrain responses to other stimuli. There are 3 categories of stimuli: positive and negative emotion, and neutral. The task can be set up as pictorial or verbal. In the pictorial version stimuli consist of pictures from the IAPS database [4]. Pictures in this database have extensively been evaluated and validated with respect to affective load and arousal. The verbal version uses words from the CELEX database. A response is given by pressing the spacebar. Up to 6 trial blocks may be predefined, with changing categories of stimuli. For each block it is possible to vary the proportion of GO and NOGO trials (default: 2:1). A short practice session precedes each block, in order to learn the distinction between the two affective categories being used in the block. Outcome variables are the mean reaction time (RT) and error percentages.

CASINO Game, a Shift Reversal Learning Task

This is a reversal shift task, intended to define in what way a participant is able to learn changes in response-outcome contingencies. It is based on the work of Cools et al. [5], and is operationalized as a casino game. In this game only two playing cards are used, and in the instruction the participant is presented as the casino manager. His task is to predict, on the basis of previous outcomes, whether an imaginary casino visitor wins or loses at drawing one of the 2 cards. Which of the two cards is winning, changes from time to time, and this shift depends on the participant's predictions. That is, a shift takes place following a predefined number of correct predictions.

In our research, the task was administered as two series of 3 blocks of trials (with the first block as a practice block, followed by 2 identical test blocks). In each block the valence condition (unexpected gain or unexpected loss) changes each time after reaching a preset but variable learning criterion. The two series differ in initial valence (gain or loss). The practice block in each series ensures that participants do understand the task at hand, as one cannot continue with the experimental blocks without a correct reversal response in the practice phase.

IDED, an Internal-External Dimension Learning Task

This task is intended to measure response reversal and the ability to concept shifting, i.e. to switch between an internal and an external dimension. The task is based on work from Mitchell et al. [6]. On the basis of visual feedback (correct / wrong) participants learn to choose between 2 stimuli that may consist of 2 dimensions: shape and line pattern. The

task consists of 9 stages in a fixed order, with changing target dimensions, of which the last 4 are crucial: Intra Dimensional Shift and Reversal, and Extra Dimensional Shift and Reversal. At each stage the rule to be learned changes, and the next stage is only started after reaching the (predefined) learning criterion in the current stage. According to Mitchell [6] this task can discriminate between two functions of the frontal cortex: response reversal and interdimensional shift learning. The proportion of errors at each stage is the main dependent variable although latencies are also registered.

Stop-Signal Task

This task is a two-choice GO-NOGO task, where the Gostimulus requires a left or right hand response [7]. On some trials shortly after the onset of the Go-stimulus a Stop signal is presented, requiring to restrain from responding. Several stimuli for Go and Stop signal are available, and the Stop signal can be visually, acoustically or combined. An important parameter, provoking the occurrence of impulsive responses, is the interval between onset of Go signal and Stop signal. The number of trials and trial blocks can be preset, and for each block a separate proportion of Stop trials can be preset.

GERT, Gradual Emotion Recognition Task

In this task, photographs of faces [8] are presented, portraying an emotional expression in a certain gradation. Task is to recognize, as quickly as possible and with the least possible mistakes, the emotion presented. The material consists of expressions of 6 emotions (anger, fear, sadness, disgust, happy and surprise), as well as the neutral expressions produced by 10 experienced actors. The pictures are in black and white and facial surroundings are omitted. Ten different gradations from each emotion have been made by way of a "morphing" technique, with steps of 10%. A researcher may predefine a task selecting stimuli from any of the emotions, actors and gradations. The task can also be preset in so-called continuous mode, in which an emotion is presented as in a film starting from the neutral face up to 100% emotion. Outcome variables are RT and accuracy for each emotion, actor and gradation. In addition a confusion matrix is supplied.

- 1. Brand, N. and Houx, P.J. MINDS: Toward a computerized Test Battery for Health Psychological and Neuropsychological Assessment. *Behavioral Research Methods, Instrumentation and Computers, 24* (1992), 385-389.
- Brand, N., Panhuysen, G., Kunst, H., Boom, J., and Lodewijkx, H. "From Theory to Research": Computer Applications in a Novel Educational Program for Introductory Psychology. In: F.J. Maarse, A.E. Akkerman, A.N. Brand & L.J.M. Mulder (Eds.), *Clinical Assessment, Computerized Methods and*

Instrumentation. Lisse: Swets & Zeitlinger, (2003), 136-143.

- 3. Dolan, M. and Fullam, R. Behavioural and psychometric measures of impulsivity in a personality disordered population. *The Journal of Forensic Psychiatry and Psychology, 15* (2004), 426-450.
- 4. Ekman, P. & Friesen, W.V. Pictures of facial affect. Palo Alto, CA: Consulting Psychologists, (1976).
- Lang, P.J, Bradley, M.M. & Cuthberth, B.N. International Affective Picture System (IAPS): Technical Manual and Affective ratings. Gainesville, Fl. The Center for Research in Psychophysiology, University of Florida, 1999.
- Cools, R., Altamirano, L. and D'Esposito, M. Reversal learning in Parkinson's disease depends on medication status and outcome valence. *Neuropsychologia*, 44, (2006), 1663-1673.
- Mitchell, D.G.V., Colledge, E., Leonard, A. and Blair, R.J.R. Risky decisions and response reversal: is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? *Neuropsychologia*, 40 (2002) 2013-2022.
- Overtoom, C.C.E., Kenemans, J.L., Verbaten, M.N., Kemner, C., Van der Molen, M.W., Engeland, H., Buitelaar, J.K. & Koelega, H.S. Inhibition in children with Attention-Deficit/Hyperactivity Disorder: A psychophysiological study of the Stop Task. *Biological Psychiatry*, *51* (2002). 668-676.

Effects of CGP7930 on Spontaneous Behavior, Anxiety and Learning in Immature Rats

Katerina Ticha

Anna Mikulecká

Pavel Mares

ticha@epilepsy.biomed.cas.cz

Department of Developmental Epileptology, Academy of Sciences of the Czech Republic Videnska 1083, Prague 14220, Czech Republic

INTRODUCTION

Gamma-aminobutyric acid (GABA), the major inhibitory mediator in the central nervous system, binds to two types of receptors - ionotropic GABA-A and metabotropic GABA-B.

There are many data on GABA-A receptors [1] but behavioral effects of drugs influencing GABA-B receptors not so numerous. GABA-B system represents the main inhibitory system at early stage of brain development. Our studies on anticonvulsant action of these drugs in immature rats had to be extended to their behavioral effects. GABA-B receptor agonist baclofen is for a long time used in the treatment of spasticity[2]), therefore also effects on motor performance should be studied. We found marked anticonvulsant action of positive allosteric modulator of GABA-B receptors - CGP7930, therefore we started to study its other effects. CGP7930 - 2,6-di-tert-butyl-4-(3hydroxy-2,2-dimethyl-propyl)-phenol, positive allosteric modulator of GABA B receptors increases the potency and efficacy of GABA at both native and recombinant GABA B receptors. This drug exhibits antidepressant, anxiolytic, and analgesic effect [3], and probably will find its use also in the treatment of drug addiction. Anticonvulsant drug with these positive side effects may find its place at least as the supportive therapy of epileptic patients.

Aim

This work was focused on the anxiolytic-like effect and on the possible negative effects on motor function and on spontaneous behavior in developing rats.

MATERIALS AND METHODS

Male Wistar rat pups were used - 12, 18 and 25 days old. The animals were maintained under 12/12h light/dark

regime in a temperature $(22\pm1^{\circ}C)$ and humidity controlled rooms. All experiments were approved by the Animal Care and Use Committee of the Institute of Physiology, Academy of Sciences of the Czech Republic. Three doses of CGP7930 were tested - 5, 10, and 20 mg/kg. Control siblings were treated with dimethylsulfoxide in a volume of 4 ml/kg (corresponding to the highest dose of CGP7930). Spontaneous behavior and possible anxiolytic-like effect were tested in the open field, motor abilities were examined using a battery of age-specific tests [4,5] - negative geotaxis, bar holding, wire mesh ascending test and rotarod test. Anxiolytic effect and influence on learning were tested in the elevated plus maze test. All these tests were repeated three times - 15, 60 min and 24 h after the injection (1st, 2nd and 3rd session).

1.) Motor Skills Tests

a) Negative geotaxis: rats were head down placed on an inclined rough surface (30°) and time to turn body to have head up was measured.

b) *Wire mesh ascending:* Rats were placed at the lowest part of wire mesh (45cm high and 15cm wide, inclined at a 70 degrees angle). Time to ascend to upper platform was measured.

c) *Bar holding:* Animals were put on the wooden bar (25cm long, 1cm in diameter, suspended 25cm above a padded soft surface) and they held with their forepaws. Time spent on the bar was measured up to a maximum of 120 s.

d) Rotarod test: The apparatus consisted of a plywood horizontal rod (10 cm in diameter, 11cm long) covered with sticking plaster in order to increase its roughness. The rod was placed 30 cm above a platform. It was programmed to rotate at 5 rpms. Rats were placed on the rotating rod with their heads against the direction of rotation. The duration of their stay on the rod was measured for 120s at maximum.

2.) Open Field

Spontaneous behavior and locomotor activity were monitored in the open-field arena (48x48x30cm). Each rat was tested for five minutes 15, 60 min and 24h after intraperitoneal injection. The arena was carefully wiped after each exposure. Behavior was recorded on a camera and distance passed, frequency and duration of various types of spontaneous behavior were calculated with

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

computer programs EthoVision XT and Observer XT (Noldus Information Technology).

3.) Elevated Plus Maze

The maze includes two open arms (30x10cm) and two closed arms (30x10x30cm) connected by central space localized 50 cm above the floor. The rat was placed at the end of an open arm – head direction from the center of the maze. Test was again repeated three times – 15, 60 min and 24 h after the injection. Latency to the entrance of the closed arm and time spent in open arms were measured. Each animal was tested for ten minutes.

Statistics

Two way repeated measures ANOVA with factor dose and session was used for statistical evaluation of results. Subsequent pairwise comparisons were performed with Student-Newman-Keul's test. The level of significance was set at p<0.05.

RESULTS

Main results demonstrated that CGP7930 does not significantly affect locomotion measured as a distance moved in the open field test. Experimental animals were calm and did not show high frequency of behavior and exploration. Motor skills were not significantly different in comparison with controls, but there were some differences between the first two sessions and the third session next day in 18day-old-rat. In 25 day-old-rat there is time spent on bar holding test in the group with dose 20 mg/kg CGP7930 significantly decreased in the first session. An anxiolytic-like effect was outlined.

DISCUSSION

Decreased locomotion might be caused by sedative effect of the drug or lower motivation as well as myorelaxation generated after injection. The differences between first two and third session might be due to the age factor experimental animals are one day older and this could play serious role in the motor tests. Another possible explanation is that the decrease in motor skills is caused by myorelaxation, which is gone the next day.

- Rogawski M.A., Löscher W. The neurobiology of antiepileptic drugs. *Nat. Rev. Neurosci.*, 5 (2004), 553-564.
- 2. Mareš P. Anticonvulsant action of GABA-B receptor agonist SKF97541 differs from that of baclofen. *Physiol. Res.*, 57 (2008), 789-792.
- 3. Frankowska M., Filip M., Przegaliński E. Effects of GABA-B receptor ligands in animal tests of depression and anxiety. *Pharmacol. Rep.*, 59 (2007), 645-655.
- 4. Bolles R.C., Woods P.S., The ontogeny of behavior in the albino rat. *Anim. Behav.*, *12* (1964), 427-441.
- Altman J, Sudarshan K. Postnatal development of locomotion in the laboratory rat. *Anim. Behav.*, 23 (1975), 896-920.

Stereoscopic Motion Analysis of Giant Honeybees

Michael Maurer maurer@icg.tugraz.at Matthias Rüther ruether@icg.tugraz.at Horst Bischof bischof@icg.tugraz.at

Institute for Computer Graphics and Vision (ICG), Graz University of Technology, Austria

Gerald Kastberger gerald.kastberger@uni-graz.at Thomas Hötzl thohoe@aon.at

Frank Weihmann frank.weihmann@uni-graz.at

Department of Zoology, University of Graz, Austria

Madhusudan Man Singh

Research Centre for Science and Technology, Tribhuvan University, Kirtipur Kathmandu, Nepal madhusudan_dongol@yahoo.com

ABSTRACT

Giant honeybees (Apis dorsata) nests display a series of layers of colony members around a single comb and may have more than 1.5 m in horizontal span. Nesting in the open, they have evolved a variety of defence strategies. Against predatory wasps, they produce highly coordinated wave-like cascades termed 'shimmering', whereby a collective of hundreds of bees flip their abdomens upwards simultaneously and in a cascadic way within a split second (see Figure 1). It has been already proved that these Mexican-wave-like traits form signals of antipredatory impact for external addressees and are prone of repelling predatory wasps [1]. However, the question is whether they also contribute for colony-intrinsic information such as reporting the momentary defensive state of the colony. To investigate this colony-intrinsic aspect, we analysed the functional architecture of the bee curtain with high spatial and temporal accuracy using a non-invasive measuring system that enables a 3D reconstruction of position and posture of colony members at the nest surface during shimmering waves [6]. A portable stereo recording setup with two high-resolution cameras (2352 x 1728 px; 60 Hz) acquired image sequences of the shimmering waves in the field on an expedition in Nepal 2009, allowing simultaneous assessment of the individual positions of hundreds of surface bees regarding the three dimensions of space (x,y,z) at the given frame rate and with an accuracy of fractions of a millimeter.

Author Keywords

Giant honeybee, *Apis dorsata*, stereo tracking, stereo reconstruction, 3D position, template matching.

INTRODUCTION

Shimmering behaviour in Giant honeybees is a main component of collective defence and is evoked by visual stimuli of mainly predatory impact [1]. It occurs in the bee curtain which is the changeable and reactive multi-layer cover of the nest (see Figure 1A), and aligns hundreds of colony members to flip their abdomens in a wave-like, highly coordinated reaction within a fraction of a second. Shimmering provides here dynamic visual cues which may confuse and repel predators [1]. The knowledge of how the curtain members are coupled together is fundamental for understanding of proximate and ultimate causes. In this paper we evoked shimmering by dummy-wasp stimulation providing visual cues with standardised velocities and directions. We measured the positional changes of the members of the bee curtain and the motion of the wave in the 3D space during shimmering using stereo tracking as the suitable method. For that, an image acquisition system has been constructed which allowed to monitor a complete nest from one side.

MATERIALS AND METHODS

The shimmering behaviour of the experimental colony was recorded with two high-resolution, high frame rate cameras observing a large part of the experimental nest. The cameras had a resolution of 4 Mpx, and a single pixel corresponded with roughly 0.4 mm length in the real world. The cameras were placed 3 m in front of the colony (see Figure 1B). This distance provided an undistorted view of the whole nest and kept the colony undisturbed.

We stimulated the experimental colony with a dummy wasp which was moved by a miniature cable car device at

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org



Figure 1. A, a nest attached to a residential house in Chitwan, Nepal. The nest displayed shimmering (see the blackish stripes at the right side of the nest), in particular in its quiescent zone, where the bees hang seemingly motionless and uniformly oriented with the head up and the abdomen down. The sample shimmering was evoked by a dummy wasp on the right side, and spread from right to left. B, the stereo camera setup with the working distance (d), the stereo angle (α) , the dummy wasp (w) on the cable car device, the nest (n) and the two cameras (c_1) and c₂). C, magnification of the white frame in A, above the mouth zone in the quiescent zone of the nest. Red and blue colours give the relative z-positions of 100 surface bees assessed by stereo tracking ; red represents positions offwards the nest, blue represents positions towards the nest (see the sketch symbolizing the nest and the two arrows in 14 steps of red and blue shading; maximal amplitude is 5 mm). The sample wave provoked initial z-movements offwards the nest surface (red)

which were followed by movements towards the nest (blue).

constant velocities of the dummy along a horizontal line, 10 cm in front of the "sunny" side of the nest. This stimulation method generated shimmering waves under controlled conditions with regard to position, direction and velocity of the dummy wasp [2].

For identification of the colony members at the nest surface during the shimmering waves, individual bees had to be segmented in the single images, tracked over time by matching them over the stereo views by stereo triangulation. One of the challenges in segmentation was due to the fact that the bees cover the comb in multiple layers. Thus the surface bees may overlap each other with their body parts, so the wings mostly cover parts of neighbouring bees. Segmentation of individuals was achieved by template matching using Normalized Cross Correlation (NCC). In a next step we identified the corresponding bees in the twin images of both frame-locked cameras. For that, we formulated the matching process as a discrete optimization problem and used reduced graph-cuts [3] to overcome the particular problem of "repetition" that neighbouring bees look very similar and may confuse the stereo matcher.

The rapidity of the individual abdominal flipping process during shimmering, taking only 100 ms, led to motion and contour blurring which additionally hampered the tracking of identified individuals. NCC was also used here to match the thorax position for a successive frame, based on the tracking data from a defined number of preliminary frames. In contrast to the conventional approach of tracking between pairs of consecutive frames, our method achieved a higher robustness of tracking of individuals in subsequent stereo images because we utilized body symmetry principles enforcing the epipolar constraints by projecting corresponding tracks in the rectified twin images.

RESULTS AND OUTLOOK

The presented framework enables a semi-automatic segmentation of individuals in a roughly two-dimensional matrix of agents. We performed an automated analysis of the individual movements in the three spatial dimensions. Up to now, tracking methods of insects refer mostly to single individuals in the 2D space [4,5]. The presented method [6] allows investigating the dynamics of the positions of hundreds of tightly packed surface bees in the 3D space at resolutions of fractions of millimetres and in time steps of 17 ms (see Figure 1C). This gained greater knowledge of the functional architecture of the bee curtain in Giant honeybees.

- Kastberger, G., Schmelzer, E. and Kranner, I. Social waves in Giant honeybees repel hornets. *PLoS ONE*, *3*, 9 (2008):e3141.doi:10.1371/journal.pone.0003141.
- 2. Kastberger, G., Weihmann, F. and Hoetzl, T. Complex social waves of Giant honeybees provoked by a dummy wasp support the special-agent hypothesis. *Communicative & Integrative Biology*, 2010, 1-2.
- 3. Zureiki, A., Devy, M. and Chatila R. Stereo matching using reduced-graph cuts. In *ICIP07* (2007), 237–240.
- 4. Veeraraghavan, A., Chellappa, R. and Srinivasan, M. Shape-and- behavior encoded tracking of bee dances. *IEEE Trans. Pattern Anal. Mach. Intell.*, *30*, 3 (2008), 463-476.
- 5. Maitra, P., Schneider, S. and Shin, M. C. Robust bee tracking with adaptive appearance template and geometry-constrainted resampling. *IEEE Workshop on Applications of Computer Vision*, 2009.
- Maurer, M. Rigid Body Reconstruction for Motion Analysis of Giant Honeybees using Stereo Vision. Master's Thesis, Graz University of Technology, Institute for Computer Graphics and Vision, 2010.

Use of Automated Tracking System Across Anxiety and Depression Models in Rodents

Patrick de Haes

Michel Meynen

Kim Hoeks

Theo F. Meert

J. Adriaan Bouwknecht ABouwkne@its.jnj.com

Department of Neuroscience, Johnson and Johnson Pharmaceutical Research & Development, Turnhoutseweg 30, Beerse, Belgium

Author Keywords

Behavior, anxiety, depression, animal model, validation

INTRODUCTION

A variety of animal models for anxiety and depression are available in rats and mice. The anxiety models are typically based on avoidance of an unpleasant stimulus like a large open space in the open-field test (OF), bright light in lightdark (LD), and elevation in elevated-plus maze (EPM). Here, the position of the animal within the arena is determined continuously resulting in measures for motor activity and anxiety separately. In depression models like the forced-swim (FS) and tail-suspension (TS) test, the resistance to an unpleasant situation (inescapable swimming or hanging by a tail) is observed where the key measure is development of immobility. It is important to work towards reliable and experimentator-independent observation in animal experiments in order to enhance reproducibility of results in an already difficult research area.

Ethical Statement

All experiments described in this abstract have been approved by the local Ethical Committee at Johnson and Johnson, Turnhoutseweg 30, Beerse, Belgium.

RESULTS

The Noldus Ethovision system allows tracking of both position of animal (focus on center of gravity) and movement (focus on change of animal surface). In our hands, we evaluated the usability of this tracking software across various assays and rodent species using infra-red (IR) background lighting where possible. The advantage of IR light is that detection becomes insensitive to the fur color of the rodent and also to the room light intensity and/or reflection resulting in highly accurate detection.

As an example, in the LD set up two problems have been solved in different ways. First, the IR-non transparent wall separating dark from light actually cut the animals in two parts when they moved through the opening. This resulted in inaccurate detection of the subject based on relative size of the two body parts. Second, anxious mice like the BALB/c strain tend to spend a large amount of time right at this separation resulting in very high levels of zone transitions, while the actual distance travelled indicates that these mice are not very explorative. This clearly hampers the usability of automated tracking systems above human observation. Software-matic extrapolation of object detection prevented the first problem, where the undetectable missing section between the two body parts was filled in automatically generating the actual shape of the subject. The second problem was solved by adding a small expansion of the dark zone into the light zone. Here, BALB/c mice sitting in the opening with very little movement were now recorded as still being in the dark. This is similar to the definition of zone-transition as frequently is being used for manual observations: in order to count a transition all four paws have to cross an arbitrary border. This however, cannot be replicated in the automated tracking system based on center point of gravity. However, the dark-zone expansion is the closest and most pragmatic solution to such a human observation. For the OF test, we show the importance of mouse strain as well as shape of arena (square versus circular). Mice were tested in a square open field where they spend an important proportion of their time in the corners. In a circular open field arena, in contrast, rats showed circular movements without spending time in a clearly preferred part of the arena. Still, in both cases, there was active avoidance of the center zone. In the EPM set up we demonstrate the importance of room illumination on exploration levels on the open arms. The level of exploration of the open arms seen with the room lights turned off was significantly higher compared to when

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

the room lights were turned on. For the depression assays based on movement, we successfully calibrated automated detection towards results by two independent human observations. Here, across a dose-response curve for the positive control imipramine (doses 0-5-10-20 mg/kg s.c.) individual results generated by two independent observers as well as the automated tracking system showed highly significant correlations. Finally, we were able to pick up relevant pharmacological effects of drugs that are known to influence anxiety- and depression-like behaviors in the models used.

CONCLUSION

In conclusion, a variety of experimental variables should be properly controlled for reliable baseline levels and pharmacologically induced responses in order to allow proper automated tracking. The Noldus Ethovision system can be used efficiently for tracking of position in the anxiety models as well as for movement in the depression models employed.

Induction and Measurement of Basic Emotions in Subjects and the Influence of Multisensory Stimuli

Rineke Valstar rineke.valstar@unilever.com

Marco Hoeksma marco.hoeksma@unilever.com

Unilever R&D,

Olivier van Noortlaan 120, 3133 AT, Vlaardingen, The Netherlands

ABSTRACT

In this study we are assessing whether we can induce positive and negative emotions in subjects as reflected in measures of the autonomic nervous system by administering pleasant and unpleasant odours. Subsequently we will assess whether a combination of visual, gustatory and olfactory stimulation results in stronger autonomic (emotional) responses as compared to either stimulus alone. We also want to assess whether there is a correlation between self reported emotional states and measures of autonomic nervous system reactivity. Four measurements of the autonomic nervous system are used: 1) the startle reflex, an eye blink reflex that is induced by an acoustic probe 2) the post auricular reflex, also induced by an acoustic probe 3) the skin conductance response and 4) the heart rate. They are chosen because of their differential susceptibility to habituation and positive and negative emotional states [1]. Next to the autonomic nervous system measurements an emotional state will be measured by self reports.

Author Keywords

Emotion, startle reflex, post auricular reflex, odour.

ETHICAL CONSIDERATIONS

The study has written approval by the Tilburg Medical Ethical Committee. The study has been conducted end 2009 according to the Declaration of Helsinki and The Medical Research Involving Human Subjects Act (WMO).

STUDY DESIGN

Thirty six healthy people between 18 and 55 years of age participate in the study. Odours are delivered through a custom built olfactometer. All psychophysiological measures are acquired through Psylab bio amplifiers (Contact Precision Instruments (CPI), London UK).

The study consists of two experimental sessions conducted on two separate days. In the first session, subjects are presented with only odours: pleasant (peach, vanillin) and unpleasant (rotten eggs) [2]. These odours are presented via an olfactometer in short bursts, interspersed with clean air (not odourised). Every 30-50 seconds, very abrupt and loud noise bursts are delivered through in-ear phones. Throughout the experiment, startle blinks and post auricular reflexes (Electromyogram, EMG), skin conductance and heart rate (electrocardiogram, ECG) are measured. Subject's emotional state is measured via self report.

In the second session, the same subjects are presented with only the pleasant odour peach, but this is put in the context of more elaborate product experience. In addition to experiencing the odour via the olfactometer, participants drink a peach flavoured soft drink and are presented with short movie clips that are thought to be congruent with product experience (a terrace in summer).

METHODS

Startle Reflex

The human startle response is a defensive reflex that is mainly characterised by rapid eye closure or blink, of which the magnitude can be measured from electrodes placed over the orbicularis oculi muscle beneath the lower eyelid [3]. In human research, short bursts (around 50 milliseconds) of white noise with instantaneous onset and a sound level of about 100 dB (A) are the most commonly used startle probes.

The startle response is measured by two Ag/AgCl electrodes attached below the lower eyelid over the orbicularis oculi and one ground electrode attached to the forehead. The skin is cleaned with alcohol (70%) and slightly abraded. Special electrode paste is used to enhance conduction and special adhesive rings are used to attach the electrodes.

All signals are fed into a special purpose bioamplifier (Psylab bio2, CPI) with a 30Hz high pass and 500Hz low pass filter and a gain of 50 μ V. The raw signal is sampled with 200Hz and recorded as well as the rectified and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

integrated signal that is derived from an online rectifier/integrator (CPI, London). A 50 Hz bandstop filter is applied to reduce mains noise.

A loud burst (100dB (A)) of white noise with instantaneous rise time and a duration of 50 milliseconds is delivered through in-ear headphones by means of a noise generator unit (CPI). We hypothesise that the startle amplitude will be increased by the unpleasant odour and reduced by the pleasant odour [4].

Post Auricular Reflex

The post auricular reflex (PAR), also induced by an acoustic probe, has been described as a useful tool for measuring appetitive responding in human [5]. The PAR is measured from the tendon of insertion behind the pinna (outer ear) of the right ear by means of two Ag/AgCl electrodes. In order to attach the electrodes correctly, the pinna is pulled forward. The skin is cleaned with alcohol and slightly abraded. Electrodes are attached by means of adhesive rings and special conductive paste is applied. All signals are fed through a special purpose bioamplifier Psylab bio2, CPI) with a bandpass filter of 10 - 1000 Hz and a gain of 50 μ V. The signal is sampled with 250 Hz. A 50 Hz bandstop filter is applied to reduce mains noise. We hypothesise that the (appetitive) PAR amplitude will be increased by the pleasant odour and reduced by the unpleasant odour, so opposite to the defensive startle. Moreover, in a second session the pleasant odour peach will be paired with a film clip creating a specific positive emotional context and a gustatory stimulus. We expect that the startle modulation and the PAR amplitude will be largest when the odour is paired with the gustatory stimulus and a film clip with a content that is congruent with the emotion elicited by the odour alone

Skin Conductance Response

Skin conductance measurement is based on changes in the resistance of the skin caused when glands in the skin produce ionic perspiration. An increase of conductivity is caused when the skin is sweaty. The electrodermal conductivity is measured by passing a small current through a pair of Ag/AgCl electrodes placed on the thenar and hypothenar eminences of the palm of the non-dominant hand. The areas where the electrodes are attached are cleaned with alcohol and slightly abraded to enhance skin contact. Special SCR gel is applied to the electrode to increase conductivity. The signal is fed through a Psylab electrodermal amplifier (CPI, London) with an on-line 10Hz filter. The gain setting is 0.4 μ Siemens. The data are stored as a continuous signal at 200Hz. We hypothesise that there will be a significant difference in SCR amplitude

between odour and no-odour conditions, but not between odour conditions, as skin conductance covaries more with arousal than valence of the stimuli.

Heart Rate

Heart rate is measured by means of three disposable Ag/AgCl adhesive gel-disc electrodes, two of which are attached to the collarbone and one (ground) electrode is attached elsewhere on the chest. The skin is cleaned with alcohol and slightly abraded before electrodes are attached. The signal is fed into a Psylab bioamplifier (CPI, London, UK) with a high pass filter of 0.5 Hz and a low pass filter of 40Hz. For heart rate, we do expect to find a difference between odour conditions, with unpleasant odours/stimuli eliciting the greatest initial declaration of heart rate and pleasant odours/stimuli the greatest peak acceleration.

Verbal Emotion: Affect Grid

The affect grid is a scale designed as a quick means of assessing affect along the dimensions of pleasuredispleasure and arousal-sleepiness [6].

RESULTS

We are analyzing the data and expect to show the results on the Measuring Behaviour Conference 2010.

- 1. Alaoui-ismaili O., Robin O., Rada H., Dittmar A., Vernet-Maury E. Basic emotions evoked by odorants: comparison between autonomic responsese and selfevaluation. *Physiol. Behav.*, 62, 4 (1997), 713-720.
- 2. Seubert J., Rea A.F., Loughead J., Habel U. Mood induction with olfactory stimuli reveals differential affective responses in males and females, *Chem. Senses*, *34* (2009), 77-84.
- 3. Bradley M.M. Emotion and motivation. In: Cacioppo JT, Tassinary LG, Berntson GG (eds) Handbook of psychophysiology, 2nd edition. Cambridge: Cambridge University Press, 2004.
- 4. Ehrlichman H., Kuhl S.B., Zhu J., Warrenburg S. Startle reflex modulation by pleasant and unpleasant odors in a between subjects design. *Psychophysiology*, *34* (1997), 726-729.
- 5. Sandt A.R., Sloan D.M., Johnson K.J. Measuring appetitive responding with the post auricular reflex. *Psychophysiology*, 46 (2009), 491-497.
- 6. Russell J.A., Weiss A., Mendelsohn G.A. Affect Grid: A single-item scale of pleasure and arousal. *J. of Personality and Social Psychology*, *57*, 3 (1989), 493-502.

Using Multivariate Analysis for Comparing Locomotor Patterns of Sprague-Dawley Rats from Different European Breeders

Theresa Andreasson theresa.andreasson@neurosearch.se Jenny Börjesson jenny.borjesson@neurosearch.se Elisabeth Ljung elisabethl.jung@neurosearch.se

Katarina Rydén Markinhuhta

Marianna Tharn

katarina.ryden@neurosearch.se

Marianne Thorngren marianne.thorngren@neurosearch.se

NeuroSearch Sweden AB Arvid Wallgrens Backe 20, 413 46 Göteborg, Sweden

ABSTRACT

Our aim was to characterize the locomotor activity pattern of Sprague-Dawley (SD) rats from European suppliers in both normal (n=44-91 per supplier) and hypoglutamatergic state (n=17-28 per supplier). Spontaneous locomotor pattern of male SD rats weighing 220-250g from Taconic, BK Scanbur, Harlan and Charles River Laboratories were tested in motility meter boxes. Both vehicle (NaCl) and MK-801 (0.7mg/kg) treated animals were tested for one hour. A locomotor pattern was generated by calculating the time series of x, y (horizontal activity) and z (vertical activity) coordinates into eleven main variables sampled at 25 Hz. Each main variable was studied at seven frequencies from 25 Hz to 0.25 Hz and pooled into 15 min bins, generating a locomotor pattern matrix of each animal consisting of 308 variables.

Multivariate Analysis (MVA) by means of Principle Component Analysis (PCA)[1, 2] was used for locomotor characterisation. PCA is a method that preserves as much (linear) information as possible in the data while projecting the observations onto a lower dimensionality (usually 2-3 PC's), reducing noise and making it possible to both investigate clusters among the studied objects as well as correlations among the variables. The first PC will always contain the largest amount of variation (i.e. information) among the PC's and all subsequently derived PC's will describe less of the total variation in data. Higher PC's with small eigenvalues (i.e. little variance) that are unstable (insignificant according to cross validation [3, 4]) are often regarded as unstructured noise. All derived PC's are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

orthogonal to each other (i.e. only containing variance not taken into account of by previously derived PC's), which also implies that they are linearly independent and uncorrelated.

As is the case with all statistic methods, MVA works better on certain types of data, for example normally distributed data is preferred as well as a decent number of observations especially if we want to look at group differences and have many variables. However, PCA is relatively distribution independent and can be performed on both short and wide as well as long and thin data matrices and can in most cases handle missing data in an acceptable way. PCA is also unbiased in the sense that the model doesn't take group belongings into account in the calculation but reveal those directions in data that contains the largest variation. The PC's are often interpreted as latent structures that reveal the "true" nature of the data in form of variances, correlations, similarities and dissimilarities even though they in a strict theoretical way are not statistically independent unless all variables follow a Gaussian distribution (normal distribution).

PCA is most commonly interpreted with mainly two closely related plots, the score-plot and the loading-plot. The score plot shows the projection of objects onto the principal components. Objects located close to each other are similar with respect to our measured data. Objects located on the opposite side of origo have opposite characteristics and if they are located in a perpendicular direction to each other (origo is the reference) have independent characteristics. Thus, a score plot is used to find clusters or patterns in the distribution of the objects in our model. Loadings are derived as the eigenvectors of the correlation matrix sorted by the eigenvalues. Loadings reveal the inherent correlation structure present in our data. Variables located close to each other have a positive correlation and variables located on the opposite direction of origo (approximately fitting a straight line) have a negative correlation. If lines from two different variables that cross each other in origo have an

approximate perpendicular angle they are uncorrelated and linearly independent. The scores and loading plots are tightly connected and can be approximately interpreted by overlaying the graphs. The experiments was approved by the Ethics Committee for Animal Experiments in Gothenburg (Göteborgs djurförsöksetiska nämnd)

Author Keywords

Phenotype, locomotor activity, multivariate analysis, rat, hypoglutamatergic, MK-801, PCA.

Development and Evaluation of a Rat Water Maze System: Impact of Maze Size, Water Temperature and Other Parameters on Performance

Liesbeth Mertens LMERTEN3@its.jnj.com Darrel J. Pemberton DPEMBER1@its.jnj.com

Johnson & Johnson Pharmaceuticals Division of Janssen Pharmaceutica NV, Turnhoutseweg 30, 2340 Beerse, Belgium

The Morris water maze (MWM) was first described over thirty years ago and is still one of the most commonly used cognition paradigms to measure spatial learning and memory in rodents. Throughout the literature, there are numerous and conflicting methodological approaches described but surprisingly very little is published concerning the impact of such procedural variables on performance. The aim of the present studies was to conduct a comprehensive assessment of variables such as pool size, rat strain, saliency of spatial and visual cues, training procedure and water opacity/temperature on a fixed platform acquisition water maze paradigm. The ultimate aim of these studies was to identify the most appropriate parameters to facilitate robust and reproducible learning curves for subsequent pharmacological studies.

For these investigations two mazes of differing diameters (1.4 and 2.0 m) were installed in separate procedure rooms that were near identical with respect to room size and internal layout. Preliminary studies focused on ensuring that both mazes were free of any potential confounding spatial bias. These studies comprised of determining the time spent in each of the four quadrants during a predetermined training time. Data from these studies demonstrated that both water mazes and their procedure rooms were free of any spatial bias that may affect performance in subsequent tests. As there are numerous publications from groups using non-oppacified water during their tests, the next study aimed to determine whether rats are able to visually locate the platform by diving. Using the small maze only, a single study was conducted using clear and oppacified water. Interestingly, there were marked differences in performance with animals trained in clear water locating the submerged

platform significantly quicker than those trained in oppacified water. Unless this was previously examined, it is quite likely that false interpretations of spatial learning could have been made in these prior published studies.

Another methodological variation in published studies is the use of a pre-training day where a curtain is drawn around the tank shielding the extra-maze spatial cues and a solid visual cue is suspended over the submerged platform. The intention of this pre-training day is to reduce anxiety and associated thigmotaxic behaviors prior to spatial cue testing. In our studies, there were clear benefits of pretraining on subsequent spatial learning. Essentially, animals that were trained to locate the platform using a suspended visual cue on day 1, performed significantly better during the subsequent spatial cue training sessions. Whilst it is generally perceived that there is an optimal water temperature for water maze studies, in our studies there were no effects of water temperature (21, 23 and 25 ± 1 °C) on spatial learning performance. During the spatial learning comparison studies between the two mazes, it become evident that performance of animals trained in the smaller of the two mazes (1.4 m) was considerably more variable between individual session and across studies. As reproducible learning curves were only evident in the larger maze, the final series of studies dedicated to examining the impact of spatial cues and rat strains were performed in the larger maze only. Surprising, whilst it is generally accepted that additional spatial cues (posters, lights etc) aid formation of enhanced spatial maps, in our studies, there was no difference in performance when these additional cues were added to the room.

The final series of studies aimed to quantify any potential performance differences between two hooded rat strains (Long Evans and Lister Hooded) and the most commonly used albino rat strains (Sprague Dawley and Wistar). Surprisingly, whilst it is generally accepted that hooded strains have enhanced visual acuity compared to albino animals, there were no significant differences between all four rat strains as measured by overall latency and path length to locate the submerged platform. There was

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

however a significant difference in swim speed between the Wistar strain and the other three rat strains, with Wistars consistently swimming faster across all trials/sessions. Although there were no overall significant differences in performance measures, there was however, a marked difference between the hooded strains and albino rat strains in time spent in periphery suggesting that hooded strains are less anxious in this task.

In conclusion, there were clear effects of maze size, use of visual cue, oppacified vs. clear water and subtle differences between rat strains in water maze performance. However, there were no effects of water temperature or use of additional spatial cue on performance.

Video Tracking Analysis of Contextual Fear Conditioning for the Screening of Anti-Alzheimer Drugs After Acute Administration in Tg2576 Mice

Luca Lorenzini luca.lorenzini8@unibo.it Alessandro Giuliani a.giuliani@unibo.it Marco Gusciglio marcogusci@libero.it Luciana Giardino luciana.giardino@unibo.it

Università di Bologna, Facoltà di Medicina Veterinaria - DIMORFIPA Via Tolara di Sopra 50, 40064 Ozzano dell'Emilia (BO), Italy

> Bruno Imbimbo Chiesi Farmaceutici, Italy B.Imbimbo@chiesigroup.com

ABSTRACT

Drug discovery and development demands highly standardized behavioral testing that should be up-grated to high-throughput screening open to cross-laboratory standardization. This is a challenging perspective for behavioral neuroscience. In this study, we present a videotracking procedure to evaluate the performance of Tg2576 Alzheimer mice in the fear-conditioning test, in order to identify potential drugs to improve cognitive performance after acute administration. Fear memory was evaluated in plaque-free Tg2576 treated with vehicle, acute CHF5074 (10 or 30 or 100 mg/kg), a novel γ -secretase modulator, and acute DAPT (100 mg/kg), a γ -secretase inhibitor, 24 and 3 hours before behavioural testing. Both drugs belong to drugs aimed to target amyloidogenic hypothesis of Alzheimer disease. CHF5074, but not DAPT, shown to contextual memory deficit reverse after single administration.

Author Keywords

Contextual memory, γ -secretase modulators, β -amyloid.

Ethical Statement

Animal experiments were carried out according to the European Community Council Directives of 24 November 1986 (86/609/EEC) and approved by the Ethical Committee for Animal Experimentation of Bologna University with Italian Ministry of Health (authorization number BQ/mc/af35295-X/10). All experiments were designed to minimize the number of animals used as well as their discomfort.

INTRODUCTION

The effects of compounds interfering with γ -secretase, the enzymatic complex responsible of the formation of the β -amyloid (A β) peptide from amyloid precursor protein (APP), on plaque deposition in transgenic mouse models of Alzheimer's disease are known but scanty data are available on the effects of these drugs on memory performance in early age, when plaque deposition is still absent. We evaluated the effects of acute treatment with a novel γ -secretase modulator CHF5074, and acute DAPT, a γ -secretase inhibitor on contextual in plaque-free Tg2576 mice. Fear-conditioning test explores hippocampal and non-hippocampal memory, allowing testing of anti-Alzheimer drugs.

METHODS

The Tg2576 transgenic mouse carries a transgene coding for the 695-amino acid isoform of human APP derived from a large Swedish family with early-onset AD. The mouse expresses high concentrations of the mutant $A\beta$, develops significant amyloid plaques, and displays memory deficits. The colony is currently maintained by breeding hemizygous males with B6SJLF1 females. The non-transgenic colony control (001349-W) may be used as an experimental control. Five-month old heterozygous transgenic females and aged-matched transgene-negative littermates were used. Transgenic mice (n = 16-17 per treatment group) received two subcutaneous injections of vehicle, CHF5074 (10 or 30 or 100 mg/kg) or DAPT (100 mg/kg) 24 and 3 hours before behavioural testing (training session of the contextual fear conditioning test). Aged-matched wild-type animals (n =26-27 per treatment group) received two subcutaneous injections of vehicle, CHF5074 (100 mg/kg) or DAPT (100 mg/kg) 24 and 3 hours before behavioural testing. Mice were trained and tested on 2 consecutive days (1). Training consisted of placing an animal in a chamber, illuminating

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

stimulus and house lights, and allowing exploration for 2 min (Ugo Basile Fear Conditioning, Calco, Varese, Italy). Afterward, an auditory cue [2 Hz] conditioned stimulus was presented for 15 sec A footshock [1.5 mAmp] was administered for the final 2 sec of the conditioned stimulus. This procedure was repeated and mice were removed from the chamber 30 sec later. Twenty-four hours after training, mice were returned to the same chambers in which training occurred (context), and freezing behavior was recorded by a computerized camera. At the end of the 5-min context test, mice were returned to their home cage. Approximately 1 h later, freezing was recorded in novel environment and in response to the cue. Freezing was expressed as a percentage of time in each portion of the test in which the animal remained immobile (at least 95% of his body for at least 500 msec). All phases of the test were recorded and immobility was detected by using the video tracking and analysis software ANY-maze (Stoelting Co., Wood Dale, IL). Data were analyzed with the appropriate model of analysis of variance (ANOVA) depending from the type of variable. Behavioral data were analyzed with two-way analysis of variance with "genotype" (wild-type and Tg2576) and "treatment" (vehicle, DAPT, CHF5074) as fixed factors and mouse as random factor. For balanced design (object recognition task), the ANOVA model included also "genotype by treatment" as fixed factor. Post hoc comparisons were directed only versus the transgenic control group (Tg2576-Vehicle) to reduce the loss of power due to multiple testing and were carried out with the Holm-Sidak's test. If needed, data were properly transformed (logarithmic or square root) to improve normality and homoscedasticity. Two-tailed p values were calculated. Calculations were done with the statistical software SigmaStatTM (Version 3.5, SPSS, Chicago, IL). Results were generally presented as mean \pm standard error of mean (SEM).

RESULTS

Subcutaneous doses of 10, 30 and 100 mg/kg of CHF5074 were administered 24 and 3 hours before the training session. During Day 1, freezing in baseline conditions was similar in all groups. During Day 2, vehicle-treated Tg2576 mice had significantly lower freezing to the context than vehicle-treated wild-type mice (59.1 \pm 3.9% vs 77.5 \pm 2.6%, p < 0.001, Figure 1a). CHF5074 showed a bellshaped dose-response curve with a non-significant increase of the contextual freezing with the 10 mg/kg dose (66.1 \pm 3.6%, p = 0.128), a significant improvement with 30 mg/kg $(71.5 \pm 3.9\%)$, p = 0.008) and no effects with the highest dose of 100 mg/kg (58.1 \pm 4.6%) (Figure 1a). Compared to transgenic controls, DAPT 100 mg/kg had not effects on contextual freezing (55.6 \pm 5.8%, Figure 1a). In wild-type mice, neither CHF5074 (100 mg/kg) nor DAPT (100 mg/kg) had effects on contextual freezing compared to vehicle-treated animals (78.1 \pm 3.1% and 76.3 \pm 3.0%, respectively, Figure 1a). There were no significant differences between treatment groups on freezing to sound.



Figure 1. Acute CHF5074 treatment attenuates contextual memory deficit in young Tg2576 mice. Columns represent averages (\pm SEM) of the percent of time in which mice are immobile when put in the old context (upper panel) or to sound in a new environment (lower panel). Vehicle-treated Tg2576 mice showed a significant impairment of contextual memory compared to control wild-type mice treated with vehicle or CHF5074 (100 mg/kg, s.c.) or DAPT (100 mg/kg, s.c.). A medium dose of CHF5074 (30 mg/kg, s.c.) significantly attenuated contextual memory deficit in Tg2575 mice (upper panel). No significant group differences were observed on freezing behavior in the new environment after sound stimulus. n = 26-37 in wild-type mouse groups and n 16-17 in the Tg2576 mouse groups. **p < 0.01 and ***p < 0.001 vs. vehicle-treated Tg2576 mice.

DISCUSSION

CHF5074 is a new NSAID derivative, in which the anticyclooxygenase activity has been removed and anti-A β_{42} secretory properties potentiated (Peretto et al., 2005). Chronic treatment with CHF5074 in the diet (375 ppm for 4 months) markedly reduced brain A β burden (-52% in the cortex; -77% in the hippocampus) (Imbimbo et al., 2007a) and completely reversed contextual memory deficit (Imbimbo et al., 2010). In the present study we found that CHF5074 improved memory after an acute treatment of young, plaque-free Tg2576 mice. The beneficial effects of acute and subchronic treatment with CHF5074 in young, plaque-free transgenic mice appear to be independent from an effect on soluble A β species, thus suggesting alternative hypotheses which require further investigations.

- Comery TA, Martone RL, Aschmies S, Atchison KP, Diamantidis G, Gong X, Zhou H, Kreft AF, Pangalos MN, Sonnenberg-Reines J, Jacobsen JS, Marquis KL. Acute γ-secretase inhibition improves contextual fear conditioning in the Tg2576 mouse model of Alzheimer's disease. J. Neurosci., 25 (2005), 8898-8902.
- Imbimbo BP, Del Giudice E, Colavito D, D'Arrigo A, Dalle Carbonare M, Villetti G, Facchinetti F, Volta R, Pietrini V, Baroc MF, Serneels L, De Strooper B and Leon A 1-(3',4'-Dichloro-2-fluoro[1,1'-biphenyl]-4-yl)cyclopropanecarboxylic acid (CHF5074), a novel γsecretase modulator, reduces brain β-amyloid pathology

in a transgenic mouse model of Alzheimer's disease without causing peripheral toxicity. *J. Pharmacol. Exp. Ther.*, 323 (2007), 822-830.

- Imbimbo BP, Giardino L, Sivilia S, Giuliani A, Gusciglio M, Pietrini V, Del Giudice E, D'Arrigo A, Leon A, Villetti G and Calza L CHF5074, a novel γsecretase modulator, restores hippocampal neurogenesis potential and reverses contextual memory deficit in a transgenic mouse model of Alzheimer's disease. J. Alzheimers Dis., 20 (2010), 159-173.
- Peretto I, Radaelli S, Parini C, Zandi M, Raveglia LF, Dondio G, Fontanella L, Misiano P, Bigogno C, Rizzi A, Riccardi B, Biscaioli M, Marchetti S, Puccini P, Catinella S, Rondelli I, Cenacchi V, Bolzoni PT, Caruso P, Villetti G, Facchinetti F, Del Giudice E, Moretto N and Imbimbo BP. Synthesis and biological activity of flurbiprofen analogues as selective inhibitors of βamyloid₁₋₄₂ secretion. *J. Med. Chem.*, 48 (2005), 5705-5720.

Measurement of Locomotor Activity in Rodents: Design of a Compact, Multifunctional, User-friendly and High Throughput System

Hansfried V. M. M. Van Craenendonck hvcraene@its.jnj.com **Darrel J. Pemberton**

Nancy Aerts

Koen Hens

Anton A. H. P. Megens

Department of Neuroscience, Johnson & Johnson Pharmaceutical Research & Development Division of Janssen Pharmaceutica N.V. Turnhoutseweg 30, B-2340 Beerse, Belgium

ABSTRACT

Locomotor activity (LMA) is typically used as a primary screen for pharmacological characterization of novel compounds. This can be as either a side-effect profiling screen or in LMA-challenge tests thought to be predictive of therapeutic efficacy. As alterations in locomotory behavior can be induced by modulation of numerous neuronal networks and their associated neurotransmitters, LMA can be altered by a broad range of drugs such as benzodiazepines, neuroleptics, opiates, and psychostimulants. Current video tracking systems are essentially unlimited with respect to the number of arenas that can be monitored (e.g. Ethovision XT: up to 100 arenas). However, within the laboratory setting it is generally unfeasible to fully utilize this large capacity for rodent studies. As locomotor activity can be influenced by multiple environmental factors (test chamber properties, ambient light, temperature, noise etc), we aimed to develop and construct a compact, user-friendly, multiple chamber behavioral recording system utilizing Ethovision XT (Noldus Information Technology BV, Netherlands).

The setup was designed and constructed with key principles such as ease of use, multi-functionality, low cost and robustness. The system comprises of sixteen recording chambers that can house arenas of varying size and shape making it suitable for both rats and mice. Controllable lightemitting diodes above each chamber provide illumination (white light) at a controllable light intensity and in addition, dark-period recording is also possible as an infra-red background illumination system was incorporated into the base of each chamber. This capability also permits tracking of animals with differing fur color. The system was designed such that cleaning between recording sessions is easy and its multifunctional design makes it possible to utilize the setup for different studies such as open field, light/dark box, marble burying, conditioned place preference and simple recognition memory tests.

To pharmacologically validate this new equipment, a series of pilot studies were conducted characterizing the effects of known hypo and hyper-locomotory agents that have been previously assessed in a standard beam-break LMA apparatus. Haloperidol (0.01-0.16mg/kg; s.c; 30 mins PTT) and amphetamine (0.31-20mg/kg; s.c; 30 mins PTT) was administered to NMRI mice (Charles River) and locomotor activity recorded for thirty minutes. As expected, both compounds elicited hypo-LMA (haloperidol) and hyper-LMA (amphetamine) effects that were in-line with historical data generated in the beam-break apparatus. Surprisingly however, there were important differences in both the magnitude of pharmacological-induced effects but



Figure 1. Experimental setup.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

perhaps more importantly a change in the sensitivity between detection of true LMA and stereotyped behaviors. For example, at higher doses of amphetamine where animals are clearly displaying stereotypic behaviors (sniffing, head weaving etc) the beam-break apparatus shows this as minimal LMA activity whereas the Ethovision detects this as LMA and therefore shows higher values.

In conclusion, we have designed, constructed and pharmacologically validated a novel, low-cost,

multifunctional, video-based LMA recording system that has significantly increased experimental throughput. Although, the preliminary pharmacological study has shown similarities between datasets generated in this new apparatus and historical studies, it has become apparent that due to subtle differences between beam-break and videobased recording systems, data may not always be comparable. Further development/calibration of this system may resolve some of these issues, but potentially improvements in the detection of LMA vs. stereotyped behaviors may require an acquisition system upgrade.

The Links Between Behavioral Markers of Performance, Personality and Communication

Márta Juhász juhaszm@erg.bme.hu Juliánna Katalin Soós soosj@erg.bme.hu Balázs Péter Hámornik hamornik@erg.bme.hu

Budapest University of Technology and Economics Department of Ergonomics and Psychology Egry J. utca 1, Budapest 1111, Hungary

ABSTRACT

In this study, we related team members' personality traits to team process behaviour, especially communication and team performance. We video registered 17 operator teams (N=90) in a Simulator Centre of a Hungarian Nuclear Power Plant. The analyses focused on personality traits related to Big Five factors and scales, communication pattern used under team process and the team performance based on overall evaluation and behaviour markers of 'soft' and 'hard' skills. Our results revealed that observable behaviour markers of team performance have strong relationship with the certain personality traits, and team related communication utterances.

Author Keywords

Big Five factor model of personality, team performance, task relevant hard and team relevant soft skills, teamoriented communication.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Synchronous interaction

INTRODUCTION

Numerous organisations tend to require effective professional teamwork in a high risk environment because expert teams have deep professional knowledge and are supposed to have fewer personal problems and conflicts between the members. In professional teams the personnel is strongly motivated to perform successfully and to maintain high standards of safety, so it can be assumed that they are all aware of the basic standards of professional proficiency. The strong professional cultures have its own positive and detrimental aspects: strong motivation to do

well their tasks, strong pride in their profession, training that stresses the need for perfection, sense of personal invulnerability, maintaining high individual standards, continual performance evaluation, pushing the limits of performance - "press-on", invulnerability to fatigue and other frailties. We were motivated by all these abovementioned characteristics to further examine the personality of the operator team members of the Nuclear Power Plant sector in details. Little is known about how personality traits predict work behaviour. In general psychologists analysing the personality traits predict the future behaviour in the workplace and rarely take into account those moderators that can influence prediction validity. Moderators can be for example the autonomy of the work situation. Testing the moderator role of the autonomy on personality-performance relations revealed positive relations between both Extroversion and Agreeableness and the performance criteria when the autonomy was high. Negative relations are found between Agreeableness and the criterion of job performance when autonomy is low. Otherwise Barrick, Mount [3] investigated the moderating role of autonomy on the relationship between the Big Five personality dimensions and supervisor ratings of job performance. Results indicate that the Conscientiousness and Extraversion factors are greater for managers in jobs with high autonomy compared with those in jobs with low autonomy. These results indicate that personality-contextual performance correlations vary across situations with different expectations for performance.

Despite the renaissance of teamwork, relatively little is known about how the individual contributes to the team intra-group processes and outcomes. The dominant way of thinking about the team is the *input-process-output model* [3]. The *inputs* can be divided into three categories: 1) individual-level factor (e.g., team member attributes, personality, skills), 2) group-level factors (e.g., structure and size) and 3) environmental-level factors (e.g., task characteristics, level of the autonomy). *Intragroup process* refers to interactions that take place among the team members and include interaction patters such as conflict, efforts toward leadership and those communication patters that differentiate teams from each other. Team *output* refers

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

to team outcomes associated with productivity, performance, as well as capability of team members to continue the work cooperatively. Based on this theoretical framework, our study focuses on the team-members' personality (input) and their relationship with the teams' communication patters (team-process) and the whole team performance based on experts' overall evaluation and behavioural markers of hard and soft skills (output).

METHOD

The data collection was based on 17 operator teams (N=90) interactions analysis in the Simulator Centre of Paks Nuclear Power Plant (NPP). The NPP Simulator Centre is a very realistic, high fidelity tool that is widely used in training and examinations creating the required level of face-validity, to be relevant for real life situations. The Paks Nuclear Power Plant's operator teams consist of four professional fields requiring the interaction of six members: Unit shift supervisor (USS), Reactor operator (ROP), Turbine operator (TOP), Field operator (FOP), Unit Electrician (UE), and Shift leader (SL).

Personality Measurement

The interest in identifying personality predictors of job performance has led researchers to use the Five Factor Personality Model (Big Five) as an important conceptual tool. In our research the NEO-PI-R personality Questionnaire by Costa and McCrae [4] was used. Each team member (N=96) was asked to fulfil the NEO-PI-R personality questionnaire some days prior to the scenario. The questionnaire based on Five Factor Personality Model measures major domains of personality:

- Neuroticism (N): the tendency to experience nervousness, tension, anxiety, emotional instability, hostility and sadness.
- Extraversion (E): an energetic approach to the external world, including sociability, assertiveness and positive emotionality.
- Openness to experience (O): describes the breadth, depth, originality and complexity of an individual's mental and experiential life.
- Agreeableness (A): the quality of one's interpersonal interactions along a continuum from compassion and altruism to antagonism.
- Conscientiousness (C): persistence, organization, and motivation in goal-directed behaviours, and socially prescribed impulse control.

NEO-PI-R personality Questionnaire measures the five major domains of personality, as well as the six scales that define each domain.

The predictive power of the model within the employment context has often been demonstrated [1,2,6,7].

Team Communication Measurement

Together with the instructors we chose a simulator scenario that every team had to perform. The mean duration of the

scenario is about 35 minutes. A video recording was made of the operators' activity during the selected scenario. At the beginning of the simulation study, the instructor informed the teams about this, but they did not know exactly which of the programmed scenarios would be videotaped. The role of instructors was to coordinate technically the team operation throughout the scenario and to evaluate the performance at the end of the scenario.

All the recorded conversations of the operators were transcribed in chronological order, identifying the operators' verbal utterances by two independent expert raters. This study only focuses on the communication utterances that are likely to be related to non-technical, team-oriented communication in the team and the major influence on the team atmosphere stemming from the individual personality. The major team-oriented communication dimensions were the following:

- **R**elation (R): The principle features of team work in high risk environment are the relation-related utterances, maintenance of contact, relationship, and vigilance in sentences, like "*Hold the line please!*", naming the addressee, etc.
- **P**oliteness (P): The speaker gives a command, information, question or affirmation formulated politely, including formulates such as "*thank you*", "*Would you so kind…*", "*Do it, please*". This communication form determines the team atmosphere, and indicates the mutual respect among team members.
- Motivation (M): Encouragement, formulated as reinforcement, completed with motivation, stimulation. For example "*It's perfect, just go on!*"
- First person plural (We): The speaker uses first person plural, expressed frequently in the form of "we, our, us, *let's*".
- Affection (A): Words describing emotions, someone's emotional status, indicating astonishment, exasperation, frustration, excitement, relieve happiness or contentment. For example "*I regret it*", "*I'm quite happy*" or laughing.
- Thinking, cognitive (T): Words indicating cognitive process. For example *"I think…"*, *"Attention!"*, *"If… than…"*, *"Check it!"*. These utterances may suggest problem-solving mechanism and can increase especially in facing with technical troubles. It shows how the teams are willing to endeavour the problem-solving mechanism.

Performance Indicators

The performance measurements were based on two kinds of evaluations. After every scenario, the team as a unit performance was evaluated by instructors applying a 3-point Likert scale (1: poor; 2: average; 3: high). Furthermore, the instructors were asked to evaluate each individual's skills according to task-relevant 'hard' skills (professional knowledge, problem-solving, comply with standards) and team-relevant 'soft' skills (communication, cooperation, impulsive control) using a 4-point Likert scale (1: weak, 2: acceptable, 3: good, 4: excellent). These

categories reflect every day used terminology in the organization and these skills are at the behaviour level directly observable.

Ethical Statement

The research has been authorised by the Management of Hungarian Nuclear Power Plant. All the research participants were informed about the research including the purpose of each measurement.

RESULTS

We found significant relation between some specific teamoriented communication dimensions and personality. A positive Pearson correlations has been revealed between politeness and Extroversion (E) (r=0.34, p<0,00), Openness to experience (O) (r=0,26, p<0,05) and Agreeableness (A) (r=-0,31, p<0,00), also a negative relation has been found between relation related communication dimension and Agreeableness (A) (r=-0,40, p<0,00). These types of communication utterances were more frequent where the team members are more sociable, assertive, active (Extroversion) and open to different feelings or fantasies (Openness to experience) but less straightforward, altruist or compliant (Agreeableness). Communication utterances related to problem-solving (Thinking) have a negative correlation with the Agreeableness (A) (r=-0,31, p<0,00) personality factor, too. This finding underlines Juhasz's [5] preview study in which the Cognitive Performance Factor had a strong negative correlation with the Agreeableness (A) personality factor.

The analyses related to behaviour markers of 'soft' and 'hard' skills proved to have significant relation to personality trait. Specifically, we found significant positive Pearson correlations between professional knowledge and Neuroticism (N) (r=0,29, p<0,05), problem solving and Conscientiousness (C) (r=0,23, p<0,05), cooperation and Neuroticism (N) (r=0,28, p<0,05) personality factors.

Team-performance as an output of the team process is Extroversion directly influenced by (E) and Conscientiousness (C) personality factors. Regression results for testing Team performance and the Big Five personality factors and scales show that Assertiveness $(\beta=0.248, p<0.05)$ and one of the scale of Assertiveness, the Activity (β =0,252, p<0,05) Conscientiousness(C) (β =0,290, p<0,05) and two of the scales of Conscientiousness factor, Order (B=0,332, p<0,005) and Achievement striving $(\beta=0,298, p<0,05)$ personality scales play a significant role in Team performance.

CONCLUSION

Our study reveals that relationship-oriented communication utterances were significantly related to Extroversion (E) and Openness to experience (O) personality factors. Taking into account the characteristics of these factors this result would have been expected. Even though, we found negative relation between relationship-oriented communication and Agreeableness (A) personality factor. Related findings were found in the Barrick, Mount [2] study, in which the predictive validity of Agreeableness (A) was investigated introducing autonomy as a moderator variable. The validity of Agreeableness (A) was also higher in high-autonomy jobs compared with low-autonomy ones, but the correlation was negative. These findings suggest that the degree of the job autonomy influences the validity of personality dimensions. It means that in NPP operator teams, where members work in high autonomy, so-called strong situation, some aspects of their personality traits are not permit to be evolved. During the team-process the hidden observable behaviour markers have got remarkable relationship with the personality traits. In conclusion, the professional knowledge and coordination behaviour markers show the most frequent relation with Neuroticism (N) and Conscientiousness (C) personality factors. It seems that these personality traits influence more the observable team behaviours. The stable role of Conscientiousness (C) in the work setting has been reinforced by Piedmont's [7] results, where the Neuroticism (N) has a good influence on professional knowledge, communication and cooperation in teamwork.

- 1. Barrick, M. R., Mount, M. K.: The Big Five Personality Dimensions and Job Performance: A Meta Analysis. *Personnel Psychology*, 44, (1991), 1-26.
- 2. Barrick, M. R., Mount, M. K.: Autonomy as a Moderator of the Relationship Between the Big Five Personality Dimensions and job Performance. *Journal of Applied Psychology*, 78, 1, (1993), pp. 111-118.
- 3. Barrick, M.R., Stewart, G.L., Neubert, M.J., Mount, M.K.: Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, 83, 3 (1998), 377-391.
- Costa, P. T., McCrae, R. R.: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): *Professional Manual*. Odessa, FL: Psychological Assessment Resources, 1992.
- 5. Juhász Márta: Influence of criterion change on supervisory rating concerning the Five-Factor Model. In: *Periodica Polytechnica*, *13*, 1 (2005), 45-59.
- 6. Gellatly, I.R., Irving, P.G.: Personality, autonomy, and contextual performance of managers. *Human Performance*, *14*, 3 (2001), 231-245.
- Piedmont, R. L., Weinstein, H. P.: Predicting Supervisor Rating of job performance Using the NEO Personality Inventory. *Journal of Psychology*, *128* (1994), 255-265.

Food Attractiveness and Gazing Behaviour

Martin Haindl

martin-haindl@gmx.at

Klaus Dürrschmid

klaus.duerrschmid@boku.ac.at

Department of Food Sciences and Technology, University of Natural Resources and Applied Life Sciences Vienna, Austria

ABSTRACT

The aim of this work was to verify whether the u-shaped curve of attention concerning canceling and recall versus attractiveness, published by Brunner et al. [1], was valid for the observation of food stuff, too. It was assumed, that both visual attractive and objects with little visual attractiveness get more attention than objects, which are characterized by mediocre visual quality. Two series of five pictures of foodstuff out of three product classes (fresh fruit, vegetables and sausages) were presented to the test persons. These series contained one of five objects in different stages of decay, both ascending and descending compared to the four residual objects. Using a Tobii® T60 eye tracking device, the gazing behaviour of the participants was analyzed and statistically evaluated. The pre-assumed u-shaped curve of attention was verified in general, but it was figured out, that foodstuffs of bad appearance attract attention much more than foodstuffs of good visual quality.

Author Keywords

Eye-tracking, gazing bahaviour, visual attractiveness, curve of attention.

MATERIALS AND METHODS

Inspired by Brunner, Reimer & Opwis' [1] paper "Cancellation and Focus: The Impact of Feature Attractiveness on Recall" we set out to ascertain the applicability of this model in a food science context by utilizing an eye-tracking device; that is, whether images of foodstuffs in various states of decay result in a U-shaped curve (with maxima in either the most appetizing and most disgusting image element) for certain eye tracking parameters. The work by Brunner et al. is, in itself based on Houston, Sherman & Baker's [2] model of "Cancellation and Focus", which is in turn an extension of a previously established feature-matching model by Tversky [3]. While previous research on these models was mostly based on textual information, the subject matter in our case dictated an alternative approach based on visual attractiveness (since food choice is largely an instinctual process, and "appetizing" and "disgusting" concepts are greatly reduced in emotional impact when we attempt to put them into words). Furthermore, previous research by this department suggested gender differences in the time necessary to detect food spoilage, which we hoped to confirm statistically due to a larger number of participants and product categories in this study.

While Brunner et al. [1] sought to extrapolate the attractiveness of certain features from their persistence in the test subjects' individual memory, eye tracking technology allows a more direct means of verification for the mental models mentioned above. By utilizing a currentgeneration Tobii® T60 eye tracking device, we were able to gain detailed insights into the impact of visual attractiveness (and, conversely, visual repulsiveness) on human gaze patterns. More specifically, we created 3 parallel lines of experiments, each representing a single major subgroup of food products (vegetables, fresh fruit and sausages). Each of these was further divided into two subexperiments: One where a single food object was shown in various stages of spoilage over time while four others stayed unchanged, and another where we began by showing all 5 food objects in an advanced state of decomposition while one of them continuously gains attractiveness (i.e. the photos are shown in a time-reversed order). The advantages of the Tobii® eye tracking technology for this line of experiments are manifold: Unlike previous generations of eye trackers where precision usually came at the cost of greater intrusiveness and set-up time (for example by necessitating specialized contact lenses, headmounted cameras or a fixed head position), the Tobii® system allows the participant to maintain a relatively relaxed and natural posture in front of the combined sensor/screen unit, also allowing a single operator to screen a large number of participants (183 recordings of females, 186 recordings of males) in a comparatively short amount of time (less than three weeks); the eye tracker comes prequipped with software for scheduling, managing and classifying participants and recordings and performing a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

basic analysis of the experimental data, which we exported for a more in-depth analysis in Statgraphics and Senstools. Lastly, its easy use of operation and general intuitiveness of the graphical user interface allowed us to familiarize new operators with the basic functions of the device in less than an hour.

It should be noted that as in previous works on this topic, in this paper the term "attractiveness" was used in a sense of "visually pleasing/appealing". In an eye tracking context, "attractiveness" might also be interpreted to mean "attracting the eye's attention", under which definition both appetizing and disgusting items would be considered "attractive". For this reason, we tried to avoid this potentially ambiguous term and replace it with "spoiled" and "unspoiled" when applicable. We also suggest using the term "attention-grabbing", commonly found in literature concerning the theory of incentive salience, as a neutral term for the tendency of an object to captivate a person's attention, regardless of that person's emotional response to it.

| | tomato | bell pepper | peach | banana | spicy sausage | turkey sausage |
|----------------------------|-----------------------|-----------------------|-----------------------|----------------------|-----------------------|----------------------|
| | • | ጥ 1. 0.010 | • | ۲ ۱ ۵ ۵۵۵ | • | Υ 1.0.000 |
| First fixation duration | b=0.135 | b=0.010 | b=-0.002 | b=0.009 | b=0.002 | b=0.008 |
| | ** | | | | | |
| | r ² =8.79 | r ² =0.36 | r ² =0.03 | r ² =0.32 | r ² =0.02 | r ² =0.53 |
| Fixations before | b=-1.337 | b=-0.186 | b=-1.43 | b=-0.356 | b=-1.037 | b=0.540 |
| | ** | | ** | | ** | |
| | r²=15.26 | r ² =0.29 | r ² =17.14 | r ² =0.43 | r ² =5.93 | r ² =1.14 |
| Fixation count | b=0.562 | b=0.905 | b=2.018 | b=0.593 | b=1.843 | b=0.126 |
| | ** | ** | ** | ** | ** | |
| | r ² =6.90 | r²=12.26 | r ² =29.67 | r²=4.14 | r ² =21.63 | r²=0.41 |
| Fixation length | b=0.608 | b=0.460 | b=0.803 | b=0.276 | b=0.673 | b=0.062 |
| | ** | ** | ** | ** | ** | |
| | r ² =23.97 | r ² =16.19 | r ² =31.31 | r ² =5.08 | r ² =25.17 | r ² =0.83 |
| Observation count | b=0.29 | b=0.208 | b=0.263 | b=0.138 | b=0.288 | b=0.035 |
| | ** | ** | ** | * | ** | |
| | r ² =5.65 | r ² =3.41 | r ² =5.40 | r ² =1.70 | r²=6.26 | r ² =0.11 |
| Observation length | b=0.609 | b=0.492 | b=0.879 | b=0.300 | b=0.709 | b=0.081 |
| | ** | ** | ** | ** | ** | * |
| | r²=21.21 | r ² =17.17 | r ² =32.43 | r ² =5.68 | r ² =25.19 | r ² =1.26 |
| Time to first fixation | b=-0.527 | b=-0.059 | b=-0.519 | b=-0.100 | b=-0.391 | b=0.204 |
| | ** | | ** | | ** | * |
| | r ² =17.92 | r ² =0.20 | r ² =20.16 | r ² =0.32 | r ² =7.13 | r ² =1.65 |
| ** 00% confidence interval | | | | | | |

**... 99% confidence interval

*... 95% confidence interval

b...slope of the regression line

 $r^{2}...coefficient \ of \ determination$

↑...image presented progresses from "spoiled" to "unspoiled"

 $\mathbf{\Psi}$... image presented progresses from "unspoiled" to "spoiled"

 Table 1. Eye tracking parameters vs. objects. Fields marked in light grey show values that increase with greater attention, while fields in dark grey show values that decrease as the object becomes more attention-grabbing.


Figure 1. Symbolic representation of the curve shapes for the measured parameters. Case 1: Fixation count", "fixation length", "observation count" and "observation length"; Case 2: "fixations before" and "time to first fixation".

RESULTS

Our analysis of the experimental data confirmed the strong attention-grabbing properties of a single visually repulsive food item surrounded by four unspoiled ones, while those of a single unspoiled item surrounded by four spoiled ones seem to be significantly less pronounced. The results are summarized in Table 1. The visual attention curve is thus more accurately described as a "half ready U-curve" than a "U-curve" for the parameters "fixation count", "fixation length", "observation count" and "observation length" (see Figure 1, case 1). For "fixations before" and "time to first fixation", this curve is inverted (see Figure 1, case 1).

It should also be noted that different forms of spoilage generate different amounts of visual attention. The bell pepper consistently shows a significantly higher coefficient of determination than the banana and turkey sausage, the latter only having a coefficient determination of 1.65 percent. While the banana undergoes a massive change in colour (almost completely yellow to almost completely brown) over the course of the pictures, the bell pepper retains its general coloration, instead undergoing a visible change in texture. The reason for the banana not producing as much visual attention is likely that a certain amount of brown spots on the peel are acceptable for the inside still to be considered edible; on the other hand, visible shriveling of a bell pepper indicates a loss of crispness and thus significantly reduced palatability. The turkey sausage showed only mild discoloration even in later stages of spoilage; furthermore, another sausage in the picture set drew significantly more visual attention due to being visibly overgrown with white-colored mould.

In contrast to the findings from previous experiments of our department suggesting gender differences in visual spoilage detection, we could not demonstrate a statistically significant difference in the gaze patterns of male and female participants in this study.

- Brunner, T.A., Opwis, K. & Reimer, T. Cancellation and focus: The impact of feature attractiveness on recall in *Proceedings of KogWis05: The German Cognitive Science Conference 2005*, eds. Klaus Opwis & Iris-Katharina Penner, (2005), 27-32.
- Houston, D.A., Sherman, S.J. & Baker, S.M. The influence of unique features and direction of comparison on preferences, *Journal of experimental Social Psychology*, 25 (1989), 121-141.
- 3. Tversky, A. Features of similarity. *Psychological Review*, 84 (1977), 327-352.

Delay Intolerance Task: Different Impact of Timeout Duration on Impulsive and Non-Impulsive Rats

Francesca Zoratto

Giovanni Laviola

Walter Adriani

francesca.zoratto@iss.it +39-06-4990-3171

> Sect. Behavioural Neuroscience, Dep. Cell Biology & Neurosciences, Istituto Superiore di Sanità, viale Regina Elena 299, I-00161 Roma, Italy

ABSTRACT

The intolerance-to-delay (ID) task, classically used to measure behavioural impulsivity in rodent models, requires the setting of specific temporal constraints: daily session length, reward delay intervals, and duration of timeout, i.e. the period following food delivery during which responding is without scheduled consequences. Here we focus on the impact of the timeout (TO) interval, to ascertain whether it affects or not the perception of delays. As expected, individual differences in the preference for large-late vs small-soon rewards emerged, with the identification of two distinct rat subpopulations: one with a nearly horizontal curve ("non-impulsive") and another with a very steep slope ("impulsive"). Noteworthy, the reaction to increasing delay length was affected by the TO duration, but only as a function of individual temperament. Interestingly, the use of extended timeout periods further decreased impulsive choice in already delay-tolerant rats. In conclusion, we demonstrate that TO is a key constraint of the ID task and therefore should be handled carefully.

Author Keywords

Cognitive impulsivity, temporal discounting, linear waiting model, animal models.

Ethical Statement

Animal experimental protocols were approved by institutional authorities, on behalf of Ministry of Health, in close agreement with European Community Directives and Italian Law. All efforts were made to minimize animal suffering, to reduce the number of animals used, and to use alternatives to *in vivo* testing.

INTRODUCTION

Impulsivity is a key symptom of attention-deficit hyperactivity disorder (ADHD), a neuropsychiatric

syndrome affecting infants and adolescents, and is also common in obsessive-compulsive and addictive disorders. Behavioural impulsivity is usually studied in rodent models with the intolerance-to-delay (ID) task, involving the choice, by nose-poking, between either immediate small amounts of food, or larger amounts of food after a delay [3].

Besides the experimenter-imposed delay, another major constraint within the ID task (to be also set by the experimenter) is the timeout (TO) interval following food delivery (i.e. the period during which nose-poking is recorded but is without scheduled consequences). Therefore, experimental subjects are forced to respond after at least the TO is elapsed. Subjects will spontaneously show a slight interval of further waiting, termed response time (RT). Hence, reinforcers and next responses will always be spaced by a mean inter-trial interval, i.e. the timeout interval plus the mean spontaneous waiting of subjects (mITI = TO + RT).

The introduction of delays is classically expected to generate a subjective state of aversion and to produce the shifting of preference towards the immediate delivery of a smaller-size reinforcer, despite lower payoff in the long term. However, we proposed that the mere absolute value of the delay duration has no universal significance per se, rather its impact on the subject could be dependent on other temporal features within the task [2]. Here, we discuss possible influences exerted by the timeout interval duration on decision making within the ID task. A refinement of ID tasks can be highly relevant to a deeper validation of preclinical models for ADHD and, more in general, of animal models for inhibitory control impairment.

METHODS

Food-restricted (88.13% \pm 0.27% of their free-feeding body weight) Sprague-Dawley male rats were tested in operant chambers provided with two nose-poking holes (Coulbourn Instruments, Allentown, PA, USA). Nose-poking in one hole (SS) resulted in the immediate delivery of a small amount of food (one 45 mg pellet, BioServ, Frenchtown, NJ, USA), whereas nose-poking in the other hole (LL) delivered a larger amount of food (five 45 mg pellets) after

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

a delay, which was increased progressively each day (from 0 s to 7 s, 15 s, 30 s, 45 s, 60 s, 75 s and finally 90 s, 8 daily sessions, preceded by 3 training sessions at delay 0 s). Following food delivery, the magazine light was turned on to signal the length of the timeout (TO), during which nose-poking was recorded but was without scheduled consequences. During the testing phase, a delay was inserted between nose-poking and large-reward delivery. The chamber light was kept on to signal the entire length of this delay. The small reward delivery was unchanged. Rats were assigned to three different timeout intervals (15 s, 30 s or 45 s; n=8 per group) and, consequently, to three different session lengths (20 min, 40 min, 60 min). This was intended to provide animals the opportunity to complete the same number of trials within the session.

Impulsivity can be measured by the steepness of the preference-delay curve. On the basis of the median value of steepness, we differentiated two distinct subpopulations [1]: an "impulsive" one, which shifted quickly towards the SS hole (i.e. with a very steep slope), and a "non-impulsive" one, with little or no shift. Therefore the two subpopulations were analyzed separately. In addition to the classical parameter of choice behaviour (percent LL preference), we calculated the mean spontaneous waiting (termed response time, RT) occurring between the end of each timeout (TO) and the next nose-poke. The pace between reinforcer deliveries and next responses, given by the mean inter-trial interval (mITI = TO + RT), was also calculated. Hence, we have recently proposed that the impact of any given delay may be proportional to this pace and be expressed as delay-equivalent odds, i.e. the extent by which delays are multiples of the mITI [2].

Data were analyzed using repeated-measures parametric analysis of variance (ANOVA). The general model was 8level delay x 3-level timeout, with timeout (the three different interval durations) as between-subject factor and delay (one per daily session) as within-subject factor. Statistical analysis was performed using Statview II (Abacus Concepts, CA, USA). Data are expressed as mean \pm SEM. Significance level was set at p<0.05. Since this study was a methodological pilot, the sample size is quite small but a replication study is already planned.

RESULTS

Choice Behaviour

As expected, all animals showed a shift in preference from the large (LL) to the immediate (SS) reinforcer as the delay length increased. However, animals belonging to the group with the shortest TO (TO15 group) experienced a clear-cut intolerance much earlier (already at delays 7.5 s, 15 s and 30 s) than TO30 and TO45 animals. In fact, at these delay values, LL choices were significantly higher in rats belonging to the groups with higher TO (TO30 and TO45 groups) than in the corresponding controls (TO15 group). Therefore, at a first glance, the magnitude of intolerance generated by the introduction of the delays seemed to depend critically on the value which was chosen as TO.

The separate analysis of these two subpopulations revealed an unexpected profile. TO15 animals belonging to the "non impulsive" subpopulation showed an interesting U-shaped curve. At lower delays, rats started shifting, with an apparent recovery of the percent LL preference starting from delay 45 s onward. Thus, at delay 90 s, subjects reached the same values they already showed at delay 0 s. As expected, "non impulsive" rats of TO30 and TO45 never shifted to a clear-cut SS preference, being relatively "tolerant" despite highest delays (see Figure 1). On the contrary, delays had a quite strong impact on animals belonging to the "impulsive" subpopulation and this independently from the duration of the TO (see Figure 2). As a matter of fact, our hypothesis (i.e. the intrinsic value of the delay may be a function of the TO) was true only in the case of "non impulsive" animals.



Figure 1. Mean (± SEM) choice (%) of the large reinforcer (LL), shown by "non-impulsive" rats belonging to the three different TO interval groups (n=4 per group).



Figure 2. Mean (± SEM) choice (%) of the large reinforcer (LL), shown by "impulsive" rats belonging to the three different TO interval groups (n=4 per group).



Figure 3. Mean (± SEM) response time (RT), i.e. the spontaneous waiting between the end of a timeout period and the following nose-poke for a reinforcer (either LL or SS), shown by rats belonging to the three different TO interval groups in the ID task (n=8 per group).

Spontaneous Waiting

Data revealed that, in the TO15 group, RT increased sharply (from around 18 s to around 27 s) when the imposed delay changed from 30 s to 45 s. In these animals, a clear recovery of LL preference comes along with a marked increase of RT values. Rats belonging to the TO30 group showed a gradual increase of response time when moving from no delay to a 90 s delay. It appears that the increasing delays directly influenced the length of rats' spontaneous waiting (RT) before next decision. In the TO45 rats, a marked discontinuity (response time increasing from around 5 s to around 20 s) was evident between imposed delays of 7.5 s and 15 s. Moreover, RT decreased (from around 21 s to around 11 s) when imposed delay changed from 45 s to 60 s. Interestingly, compared with the other two groups, TO45 rats expressed markedly lower RT values during all sessions at very high delays (see Figure 3).

DISCUSSION

Present data demonstrate that TO value is a crucial temporal constraint but only within subjects with little or no impulsivity.

Within the "non-impulsive" subpopulation, the main result is that TO30 and TO45 rats show some intolerance at higher delays when compared to TO15 rats. To explain this profile we propose that only delays that are extended enough, compared to the TO value, will generate a considerable drive to support the shift towards SS. Indeed, the specific delay length of 30 s had a quite low impact (i.e. a low odds value) in subjects used to a very long TO (TO45 and TO30 animals). Conversely, we can hypothesise that it was perceived as much more frustrating (i.e. equivalent to higher odds values) in subjects paced by quite a shorter TO (TO15 animals).

The RT profile observed in TO15 rats is consistent with previous work by our group. Indeed, a TO interval around 15-20 s and a session length around 20-25 min have been used so far in our hands. Under these conditions, a clear-cut discontinuity has been repeatedly observed between imposed delays of 30 s and 45 s, i.e. when imposed delay was equivalent to the mITI. Specifically, at that time point, the 15 s of TO plus the 18 s of RT give a mITI value of 33 s. As such, rats apparently begin to react when the delay value exceed the mITI.

The RT profile observed in TO30 animals are in agreement with the "linear waiting model", a formal model of voluntary waiting in experimental animals [4]. Indeed, according to this model, the duration of pauses following food presentation is determined by the preceding inter-food interval. Thus, spontaneous pauses are directly proportional to increasing temporal distances between food deliveries (a behaviour termed "temporal tracking"). Reinforcing events are therefore progressively rarefied.

The use of extended timeout periods further decreased impulsive choice in already delay-tolerant rats and also affected the profile of RT. This possibly reflects the ability of these rats to cope with long paces between reinforcing events. In summary, delay-induced states of aversion may depend on previous adaptation to the rate of reinforcement, at least for non-impulsive individuals.

- 1. Adriani, W., Caprioli, A., Granstrem, O., Carli, M., Laviola, G. The spontaneously hypertensive-rat as an animal model of ADHD: evidence for impulsive and non-impulsive subpopulations. *Neuroscience and biobehavioral reviews*, 27 (2003), 639-51.
- Adriani W., Zoratto, F., Romano, E., Laviola, G. Cognitive impulsivity in animal models: role of response time and reinforcing rate in delay intolerance with two-choice operant tasks. *Neuropharmacology*, 58 (2010), 694-701.
- Evenden, J.L., Ryan, C.N. The pharmacology of impulsive behavior in rats, the effects of drugs on response choice with varying delays of reinforcement. *Psychopharmacology*, *128* (1996), 161-170.
- Innis, N.K., Staddon, J.E.R. Temporal tracking on cyclic-interval reinforcement schedules. *Journal of the experimental analysis of behavior*, 16 (1971), 411-423.

Quality Monitoring and Analysis of Open Field Behavior in Rats by Means of Multivariate Analysis

Johan Kullingsjö

Ylva Sunesson

Theresa Andreasson

Henrik Ponten

NeuroSearch Sweden AB, Arvid Wallgrens Backe. 20, 413 46 Göteborg, Sweden johan.kullingsjo@neurosearch.se

ABSTRACT

Day to day variation among investigated objects, environmental disturbances and artifacts arising from hardware and data management systems, constitutes challenges to the quality of locomotor activity (LMA) data. We aim to demonstrate how multivariate analysis (MVA) can be applied for both quality monitoring and analysis enhancement of experimental data. To our knowledge, quality monitoring of animal experiments are in general almost absent and we are not aware of any commercial providers of equipment that provides good quality monitoring tools. The rationales for quality monitoring are several, most importantly, without comparable results between experiments it's usually impossible to draw conclusions of value. Even with a huge effort to keep experimental parameters as constant as possible there will probably be a relatively large day to day variation. If a lot of technicians are involved doing experiments and if we need to perform these kind of experiments on a daily basis both differences in the result between weekdays and between persons that perform the experiments are obvious factors to investigate as well as seasonal effects and a multitude of other factors. Other areas of investigation is the reliability of the equipment and to monitor and take care of unrealistic artifactual values . We would also like to know that our behavioral variables have a broad description of behavior (i.e. not only variables that have a strong correlation to activity). Finally, none of our efforts would be worth it if we couldn't replicate our trials and receive a similar result between replicates. This demand us to have sharp and reliable tools to evaluate our experiments. This novel method for quality monitoring has been developed inhouse and has never before been published on behavioral data even though the methodology previously has been described [1] for other types of data (neurochemical indices).

The locomotor analysis system consists of eight motility meters with infrared beams (25Hz) in two levels for horizontal and vertical measures. Data quality is monitored in a semi-automated system. First an automatic software filtering is applied to check for data consistency in format and magnitude. Secondly 308 locomotor variables are automatically calculated from the 25Hz data generating LMA patterns for each individual. MVA monitoring of quality is performed by manually evaluating the control animals in each new experiment in relation to historic controls (n>5000) in a number of automatically generated multivariate models. Outlier animals are marked as "weak" or "severe" as well as the whole experiment (good/no good). The LMA patterns are further used for MVA of treatment effects on both a single- and multiple number experimental level [2,3].

The experiments were approved by the Local Animal Ethics Committee in Gothenburg

Author Keywords

phenotype, locomotor activity, multivariate analysis, PCA, PLS, rat, QC, standardization.

- 1. Kullingsjö, J. Star Quality ,Three simple quality steps that aims towards the stars. , *Kemometri i västsverige conference(Gothenburg,Sweden,2005)*.
- 2. Waters, S., Kullingsjö, J., Waters, N. From beamcrossing to breakthrough compounds - behavioral pattern analysis based on PCA and PLS. Invited lecture at the UAM conference (Malmoe, Sweden, 2003).
- 3. Kullingsjö, J. Rats in space a multivariate odysée. Bästkustkemometri, conference (Gothenburg, Sweden, 2007).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Different Interpretation of the Hot Plate Test in Rats

Thomas E. Pitcher

tom.pitcher@pfizer.com +44 1304 644638

R. Neuman

Janet R. Nicholson Janet.R.Nicholson@Pfizer.com

+44 1304 646848

Ian Machin

Pfizer Global Research & Development Sandwich Laboratories, Pfizer Limited (IPC 351), Sandwich, Kent CT13 9NJ, UK

ABSTRACT

Thermal tests have been widely used in rodents; one of the main models used is the hotplate test. Classically the hotplate test has been used as an indication of analgesia using 52.5°C as the thermal stimulus (South, S.M. 2009). With the development of new targets for pharmacotherapy some concern has orison over disruption of thermal perception (eg TRPV1). Thus, we would like to propose that the hotplate can provide a useful tool to address these concerns pre-clinically when it is used at higher temperatures and can be useful in the detection of impairment to temperature perception. In the present study, pharmacology was assessed using a TRPV1 antagonist (PF-3864086, PF-386). TRPV1 antagonists have been reported to have effects on thermal perception. These effects have been highlighted in clinical studies, namely arm emersion and hot water sipping test (ASCPT March 2009). PF-386 has a similar profile to the MK2295 published clinically and is utilised in this study.

Animals were acclimatised to the hotplate (Ugo Basile) prior to testing (at 48°C, a non-noxious temperature). Animals were then assessed over a range of temperatures $(50^{\circ}C, 52.5^{\circ}C + 55^{\circ}C)$ on consecutive days in the same batch of animals, i.e. $50^{\circ}c$ day 1, $52.5^{\circ}C$ day 2 and $55^{\circ}C$ on day 3. Animals were placed on the hotplate and latency to respond was recorded in seconds (flinching, licking, biting or jumping). Baseline responses were taken on each test day to establish normal responses at a given temperature prior to pharmacological evaluation. PF-386 was assessed at 100+300mg/kg p.o. tested 1 +3hrs post dosing.

Results show that 100mg/kg significantly increased latency

to respond at 52.5°C and 300mg/kg at 50°C, 52.5°C and 55°C.

These results indicate that there is impairment to thermal perception at these doses; the lack of effect with 100mg/kg at 55°C could be due to other TRP channels being activated at this higher temperature which are blocked only by the 300mg/kg dose. The data aligns with clinical data indicating that potential safety risks can be detected with this model. This model could be a useful translatable biomarker which could be extended to projects beyond TRPV1, to determine changes to normal thermal perception.

Author Keywords

Hot plate, rat, TRPV1, thermal perception.

Ethical Statement

Experiments were conducted on male Sprague Dawley rats (175-200g) from Charles River UK. Animals were housed in 4s, maintained on a 12/12-h light/dark cycle with food and water available *ad libitum*. All procedures were performed in accordance with the UK Animals (Scientific Procedures) Act 1986.

- Smith. 1. South. S.M.. Edwards. S.R., M.T. Antinociception serum concentration versus following acute administration of relationships intravenous morphine in male and female spraguedawley rats: differences between the tail flick and hotplate nociceptive tests. Clini. Exp. Physiol., 36, 1 (2009), 20-28.
- 2. Crutlow M, Pharmacologic Inhibition of TRPV1 Impairs Potentially Injurious Heat in Healthy Subjects. American Society for Clinical Pharmacology and Therapeutics (ASCPT) in Washington, DC March 18-21. 2009.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Preliminary Study of Behaviour and Sound Emitted by European Spiny Lobster *Palinurus Elephas* During the Interaction with Predators

Giuseppa Buscaino giuseppa.buscaino@cnr.it **Francesco Filiciotto** francesco.filiciotto@iamc.cnr.it Gaspare Buffa gaspare.buffa@iamc.cnr.it

Antonio Bellante antonio.bellante@iamc.cnr.it Vincenzo Di Stefano vincenzo.distefano@iamc.cnr.it

Istituto per l'Ambiente Marino Costiero, Consiglio Nazionale delle Ricerche, Via del Faro no. 3, 91021 Granitola, TP, Italy

ABSTRACT

In this work it was analyzed the antipredatory behaviour of the European spiny lobsters (Palinurus elephas) in presence of two natural predators: the European conger eel and the common octopus. Behaviour of lobsters and predators was audio-video monitored using water-proof cameras and a pre-amplified hydrophone. Analysis of the data allowed the study of the antipredatory behavioural events (including the sound emissions) of lobsters exposed individually or in groups to a single specimen of predator. No antipredatory behavioral events were observed during the control phase, neither acoustic signals audible by the operator were emitted. Predators used to attack mainly the groups of lobsters. During the test phase, indeed, the most frequent behavioral events were the Alert (a kind of behavior, never described) the Point and the Whip while the Tail Flip and the Lunge were mainly associated to the sound emissions. Although uncertainties still remain about the function of acoustic emissions, is possible to hypnotize that these sounds represent an integral part of defense mechanisms against the predators.

Author Keywords

Lobster, behaviour, acoustics emission, predator.

INTRODUCTION

The Mediterranean lobster (*Palinurus elephas*) shows a antipredatory behaviour in response to an imminent threat [5,1,2,8]. This behaviour could also comprise acoustic emissions, associated to other defence-related behavioural

events or not, even if the motivational mechanisms determining this antipredatory attitude are not entirely known [4,7,6,9]. The aim of this work was to analyze European spiny lobster antipredatory behavioural events (*Palinurus elephas*) in relation to acoustic emissions, in presence of natural predators. The octopus, in fact, is an active predator of spiny lobsters [2] and the conger eel is known as a voracious predator of fish, cephalopods and crustaceans [3]. Consequently, this study employed the European spiny lobster (*Palinurus elephas*), the European conger eel (*Conger conger*) and the common octopus (*Octopus vulgaris*).

METHODS

The study was carried out using 50 European spiny lobsters, 5 European conger eels and 5 common octopuses. 20 experimental trials in total were carried out in a circular tank (3.0 m diameter), 10 of which with single lobsters (five with an octopus and five with a conger eel) and 10 (five by each predator) with groups of four lobsters. After an hour of acclimation, lobsters were monitored by an audio-video system for one hour without predators (control phase); then, a solitary predator was introduced into the tank and the monitoring was carried out for one hour (test phase). Visual monitoring was carried out using two water-proof cameras, the first one located on the bottom of the tank and the second one upon the tank. Signals coming from the video cameras and from a pre-amplified (VP1000, Reson, Slangerup, Denmark) hydrophone (8104, Bruel & Kjer, Nærum, Denmark) were synchronized and digitalized by DV-RT4 Real Time (D-Vision) card, managed by the DSE D-Vision (Torin, Italy) software. Audio-video files were analyzed in continuous and each behavioral event observed was noted. Eventual acoustic emission, associated to the other behavioral events, was also recorded. Behavioral events were identified and recorded on the basis of the general scheme proposed by Lavalli & Herrnkind [8] modified by the authors (Table 1). Moreover, a kind of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Mean number of antipredator behavioural events (n/h-1 ± standard deviation) of lobsters per test typology with predators.

behavioral event, never described before and named *Alert*, was recorded during the experiment. The entire work presented here complied with current regulations regarding animal experimentation in Italy.

RESULTS

During the control phase, no antipredatory behavioral events reported in the Table 1 were observed for all the lobsters, neither acoustic signals audible by the operator were emitted by the lobsters.

During the test phase, in relation to predator attacks, many behavioral events were observed and the most frequent of which were the Alert, the Point, and the Whip, respectively (Figure 1). Predators used to attack mainly the groups of lobsters and the conger eels carried out more predation attempts than the octopuses. In total, 2433 behavioural events were recorded of which 239 (10%) were Sound Emissions. (Table 2). Moreover, a positive relation was found between the total number of predator attacks and the total number of the acoustic emissions of lobsters, both individually and in group (Figure 2).



Figure 2. Linear correlation between the total number of predator attacks (both octopuses and conger eels without distinction) and the total number of acoustic emissions of the lobsters, both in the test in single and in group.

| Lobster Event | | | | | |
|---|--|--|---|--|--|
| Alert | Point Parry | | Whip | | |
| The lobster points towards the direction of predator with the antennae, rising the legs and the telson. | One or both antennae of the lobster move towards in the direction of the predator. | The tip or side of an antenna of the lobster makes contact with some part of the predator. | The lobster strikes the predator with either one or both antennae. | | |
| Pirouette | Tail Flip | Lunge | Sounds | | |
| The lobster turns on the spot with the antennae in the direction of the predator. | The lobster makes a rapid flexion of the extended abdomen, one or more times. | The lobster's antennae are swept together rapidly in front of the predator. | The lobster emits a single or multiple audible acoustic signal. | | |
| Predator Action | | | | | |
| Attack | The predator moves directly towards the lobster to within one antenna's length. | | | | |

Table 1. General scheme proposed by Lavalli and Herrnkind (2009) modified by the authors: description of the behavioural events of the European spiny lobster and predators.

CONCLUSION

The present study pointed out an important relation between the lobster anti-predatory behavioral events and the sound emissions. This relation is more evident in the Tail flip and in the Lunge, that represent the behavioral events displayed by lobsters when in extreme danger of predation (the Tail flip is a quick removal from the predator, while the lunge is the quick move toward the predator). Although uncertainties still remain about the function of acoustic emissions and the identity of the targets of these emissions (as the majority of marine predators is not able to perceive lobsters' emission frequencies, Buscaino et al. unpublished data), it is possible to hypnotize that the sounds represent, in association with other behaviors, an integral part of the defense mechanisms against the predators. Moreover, we described a behavioural event for which there is no evidence in the literature, which was called Alert and observed that the effectiveness of the "controlling" behavioural events (using with an higher frequency Alert, Point and Whip) reduce the frequency of the "last attempt" events (Tail Flip, Lunge).

REFERENCES

 Barshaw, D.E., Lavalli, K.L. and Spanier, E. Offense versus defense: response of three morphological types of lobsters to predation. *Marine Ecology Progress Series*, 256 (2003), 171–182.

| | Tail Flip | Lunge | Parry | Whip | Point |
|---|-----------|--------|----------|----------|------------|
| Total number of events | 185 | 15 | 386 | 352 | 595 |
| Event with sound emission (% of total) | 140 (76) | 8 (53) | 52 (13) | 37 (11) | 2 (0.3) |
| Event without sound emission (% of total) | 45 (24) | 7 (47) | 334 (87) | 315 (89) | 593 (99,7) |

Table 2. Audible sounds emitted by the lobsters in association behavioral events recorded during the test phase.

- 2. Bouwma, P. E. and Herrnkind, W. F. Sound production in Caribbean spiny lobster *Panulirus argus* and its role in escape during predatory attack by *Octopus briareus*. *New Zeland Journal of Marine and Freshwater*, 43 (2009), 3-13.
- 3. Cau, A. and Manconi, P. Relationship of feeding, reproductive cycle and bathymetric distribution in *Conger conger. Marine Biology*, 81 (1984), 147-151.
- 4. George, R.W. and Main A.R. The evolution of spiny lobsters (*Palinuridae*): study of evolution in the marine environment. *Evolution*, 21 (1967), 803-820.
- Hazlett, B.A., Bach, C.E., McLay, C. and Thacker, R.W. A comparative study of the defense syndromes of some New Zealand marine crustacea. *Crustaceana*, 73 (2000), 899–912.
- 6. Holthuis, L.B. Marine lobsters of the world: an annotated and illustrated catalogue of species of interest

to fisheries known to date. FAO Fisheries Synopsis, (1991), 13. Food and Agriculture Organization of the United Nations. Rome, Italy.

- Kanciruk, P. Ecology of Juvenile and adult Palinuridae (Spiny Lobsters). In: *The Biology and Management of Lobsters*. Edited by Cobb, J.S. and Phillips, B.F. ACM Press (1980), 59-96.
- 8. Lavalli, K.L. and Herrnkind, W.F. Collective defense by spiny lobster (Panulirus argus) against triggerfish (*Balistes capriscus*): effects of number of attackers and defenders. *New Zealand Journal of Marine and Freshwater Research*, 43 (2009), 15-28.
- 9. Patek, S.N. and Oakley, T.H. Comparative tests of evolutionary trade-off in a palinurid lobster acoustic system. *Evolution*, 9 (2003), 2082–2100.

Combining an Operant Chamber Paradigm with [18F]Fluorodeoxyglucose MicroPET Imaging: A Study on Conflict Processing in the Rat

Christine Marx cmarx@nf.mpg.de **Heiko Backes**

Wolfgang Hauber Hauber@bio.uni-stuttgart.de

Max-Planck-Institute for Neurological Research, Gleueler Str. 50, 50931 Köln, Germany

Department of Animal Physiology Institute of Biology, University of Stuttgart 70550 Stuttgart, Germany

Carsten Calaminus¹

Günter Mies

Rudolf Graf

Heike Endepols

Björn Lex¹

Max-Planck-Institute for Neurological Research Gleueler Str. 50, 50931 Köln, Germany

ABSTRACT

Response conflicts occur when spatial features of a stimulus, although task-irrelevant, are incongruent with spatial features of the response, which leads to longer reaction times and higher error rates (Simon effect). To identify brain areas relevant for conflict processing as well as associated neuronal mechanisms in rats, we combined auditory Simon tasks in an operant chamber with positron emission tomography (PET). As tracer we used [¹⁸F]fluorodeoxyglucose (FDG), which accumulates in active cells and therefore allows to perform the PET scan (FDG-PET) after the rat has completed the cognitive task. Rats showed a robust Simon effect including sequential modulation, similar to the findings in human participants. Our subtractive as well as correlative analyses of PET data revealed that a prefrontal cortical area, the prelimbic cortex, the striatum, and the premotor area M2 were involved in conflict processing. The results of this study support the dual route model established for the processing of response conflicts.

Author Keywords

microPET, imaging, operant chamber, Simon task, conflict processing, conflict monitoring.

INTRODUCTION

Everyday's activities such as writing an e-mail during a telephone call or driving a car while talking to a fellow passenger lead to conflicts in the cognitive system. Such phenomena are characterized by conflicting results of parallel information processing, which leads to higher reaction times and increased error rates. It is thought that the dorsal anterior cingulate cortex is involved in monitoring such conflicts and prefrontal areas in resolving them [5]. However, the exact location of involved brain areas are not known so far. Recently, it was shown that rats are able to perform an auditory conflict task, which generates a Simon effect [1]. The Simon effect is a neuropsychological interference effect in which reaction times are longer and errors more frequent when spatial features of the stimulus (although task-irrelevant) are inconsistent with spatial attributes of the response [7]. A dual route model has been proposed to explain the Simon effect [2]: A fast automatic route processes task-irrelevant stimulus location and selects a response directed towards the stimulus, whereas a slower intentional route uses the relevant stimulus feature to select the appropriate response established during training. The resulting response conflict delays performance in the current trial and initiate conflict monitoring and resolution mechanisms, which reduce conflict effects in the next trial (sequential effects, [3,8]) To study the neural basis of conflict processing in animal models, invasive techniques like brain lesions or pharmacological manipulations have frequently been used. However, brain imaging has the advantage of displaying focal activity of the entire brain in intact animals. While fMRI is not suitable for tasks where the animals are required to move freely, it was recently shown that combining microPET imaging using FDG-PET in

combination with a behavioural task is suitable to study

¹Present address: Department of Radiology, Laboratory for Preclinical Imaging and Imaging Technology, University of Tübingen, Röntgenweg 13, 72076 Tübingen, Germany, Bjoern.Lex@med.uni-tuebingen.de.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Rule: 10 kHz - go left; 15 kHz - go right.

Figure 1. Positions of inputs and outputs of the operant chamber for the auditory Simon task.

stress in rats [4,9]. In the current experiment, we set up an auditory Simon task for rats in an operant chamber, which can be easily combined with microPET imaging to study the functional anatomy of conflict processing.

BEHAVIOURAL TASK

Measuring reaction times in rodents is difficult because whole body movements are required for responses and therefore large individual differences occur. To overcome this problem, we developed a nose poke-induced Simon paradigm (Figure 1) placed in an operant conditioning chamber (Med associates Inc. Georgia, VM, USA).

Eight Lister hooded rats (*Rattus norvegicus*) were trained to perform an auditory Simon task. Auditory stimuli were presented unilaterally, and response side (left/right) was indicated by stimulus pitch (10 kHz, 15 kHz; relevant stimulus feature). When stimulus and correct response were on the same side, this was defined as a non-conflicting condition (compatible). In conflicting trials (incompatible), stimulus and required response were on opposite sides. Reaction times and error rates where measured. Four animals performed four different tests each combined with microPET. In the control task, sound was presented bilaterally (neutral task T_N). T_C consisted of 100 % compatible trials, and T_I of 100 % incompatible trials. In T_R , 50 % compatible and 50 % incompatible trials followed each other randomly.

MICROPET

The animals received an intraperitoneal injection of 2 mCi FDG during a brief anesthesia. After five minutes, rats performed a Simon task in an operant chamber for 30 min. During this period FDG was taken up by glucose transporters, phosphorylated by hexokinase but not further metabolized and therefore accumulated in cells with high metabolic activity. MicroPET scans took place under isoflurane inhalation anaesthesia in a Focus 220 microPET scanner (CTI/Siemens Knoxville, TN) with a resolution at center of field of view of 1.4 mm. Emission data were taken over 30 min starting 60 min after FDG injection. Following Fourier rebinning, data were reconstructed using OSEM3D/MAP reconstruction [6] resulting in voxel sizes of 0.38 x 0.38 x 0.82 mm.To rule out gross structural brain anomalies and to provide individual templates for coregistration of the PET images, T2-weighted structural MR images were acquired. MRI and PET data were analyzed with the help of the imaging tool VINCI [10].

RESULTS

As predicted, we observed significantly shorter reaction times as well as lower error rates in compatible compared to incompatible trials. Furthermore, we could demonstrate sequential modulation, i.e., a Simon effect was only present if the preceding trial was compatible. The analysis of the PET data revealed different activation patterns for different task settings. We detected increased metabolic activity in the right prelimbic cortex during T_I and T_R compared to T_N , indicating that this region is involved in conflict processing. Additionally, a part of the left striatum displayed decreased activation during T_I and T_R , which may be related to the suppression of unwanted movements. During T_I , metabolic activity of both left and right anterior premotor cortex was positively correlated to error rate, indicating that animals with high aM2 activity were prone to errors.

CONCLUSION

FDG-PET using high resolution microPET scanners is a promising technique to study the metabolic activity of brain areas related to conflict processing in small animals. We have identified a prefrontal area, the prelimbic cortex, to be involved in conflict processing, which is in line with human findings. Furthermore, our results are in accordance with the dual route processing model. Because rats show robust Simon effects including sequential modulation, they are suitable animal models to investigate conflict monitoring as well as conflict resolution processes and its modulation by brain function disturbances such as stroke.

All animal procedures were in accordance with the German Laws for Animal Protection and were approved by the local animal care committee and local governmental authorities.

- 1. Courtiere, A., Hardouin, J., Burle, B., Vidal, F., & Hasbroucq, T. Simon effect in the rat: a new model for studying the neural bases of the dual-route architecture. *Behavioral Brain Research*, *179*, 1 (2007), 69-75.
- 2. Dejong, R., Liang, C.C., & Lauber, E. Conditional and unconditional automaticity – A dual process model of effects of spartial stimulus-response correspondence.

Journal of Experimental Psychology-Human Perception and Performance, 20, 4 (1994), 731-750.

- Gratton, G., Coles, M.G.H., & Donchin, E. Optimizing the use of information – strategic control of activation of response. *Journal of Experimental Psychology-General*, *121*, 4 (1992), 480-506.
- Jang, D.P., Lee, S.H., Lee, S.Y., Park, C.W., Cho, Z.H., Kim, YB. Neural responses of rats in the forced swimming test: [F-18]FDG micro PET study. *Behavioral Brain Research*, 203, 1 (2009), 43-47.
- 5. Kerns, J. G. Anterior cingulate and prefrontal cortex activity in an FMRI study of trial-to-trial adjustments on the Simon task. *Neuroimage*, *33*, 1 (2006), 399-405.
- 6. Qi, J., Leahy, R.M., Cherry, S.R., Chatziioannou, A., Farquhar, TH. High-resolution 3D Bayesian image reconstruction using the micro PET small-animal scanner. *Phys Med Biology*, *43*, 4 (1998),1001-1013.
- 7. Simon, J.R., Hinrichs, J.V., Craft, J.L. Auditory S-R compatibility: reaction time as a function of ear-hand

correspondence and ear-response location correspondence. J. Exp. Psychology, 86 (1970), 197-202.

- 8. Stürmer, B., Leuthold, H., Soetens, E., Schroter, H., & Sommer, W. Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology-Human Perception and Performance, 28*, 6 (2002), 1345-1363.
- Sung, K.K., Jang, D.P., Lee, S., Kim, M., Lee, S.Y., Kim, Y.B., Park, C.W., Cho, Z.H. Neural responses in rat brain during acute immobilization stress: a [F-18] FDG micro PET imaging study. *Neuroimage*, 44, 3 (2009), 1074-1080.
- Vollmar, S., Hampl, J., Kracht, L., Herholz, K., Integration of Functional Data (PET) into Brain Surgery Planning und Neuronavigation, *Advances in Medical Engineering, Springer Proceedings in Physics*, 114 (2007), 98–103.

The FischFITMonitor – A New System for Monitoring Multiple Physiological and Behavioural Parameters in Fish

Georg Staaks oki@igb-berlin.de **Daniela Baganz** baganz@igb-berlin.de

Oliver Jauernig jauernig@igb-berlin.de

Leibniz-Institute of Freshwater Ecology and Inland Fisheries Mueggelseedamm 301, D-12587 Berlin, Germany

Carsten Brockmann Fraunhofer Institute for Reliability and Microintegration, Germany Carsten.Brockmann@izm.fraunhofer.de **Ullrich Balzer**

Institute of Agricultural and Urban Ecological Projects at the Humboldt-Universität Berlin, Germany ullrich.balzer@agrar.hu-berlin.de

ABSTRACT

The presented work reflects the scientific research on and development of the FishFITMonitor (FFM), an innovative telemetric system for simultaneous monitoring of multiple physiological and behavioural parameters of free swimming fish. The miniaturization of the developed sensor system allows the integrated measurement of temperature, DC resistance of the direct cell environment, cellular electrical potential, electromyography (EMG), three dimensional swimming acceleration, heart rate and breathing rate.

The FishFITMonitor is a new approach for the combined evaluation of physiological and behavioural parameters. The assessment of the fish status is carried out by the mean of chronobiological regulation diagnostics. Therefore a set of time series analysis functions is applied to prepare the sensor signals as input for an artificial neural network (ANN) which then analyses distribution, stability, changes and synchronisation of the chronobiological regulation states of all measured parameters. At the first developmental stage the system will be usable in restricted size aquaria, basins or smaller ponds.

Author Keywords

telemetric sensor system, multiple parameter monitoring, physiological and behavioural parameter, pattern analysis, neuronal network

INTRODUCTION

MacIntyre [6] suggested that animal health is a central tenet of animal welfare. A reliable evaluation and monitoring of health and welfare in laboratory, domestic or farm animals is still one of the most challenging issues in animal behaviour research and even for the development of appropriate measuring techniques and evaluation methods. Especially for fish the natural borders between us as human observers and the individuals of interest limit the access to behavioural and physiological data on which health and welfare estimations can be based. Therefore welfare in fish has not been studied to the same extent as in terrestrial animals. Deviations of measures from normality enable us to say that fish are stressed and have impaired welfare and health, but it is still difficult to evaluate the extent of impairment or forecast if the animals might return to normality or show mortality [4].

Following these considerations and based on the experiences with conventional telemetric systems (5; 7) and the methods of chronobiological regulation diagnostics in humans and mammals [1; 2] we decided to develop a miniaturized integrated sensor and evaluation system for the simultaneous online monitoring of several physiological and behavioural parameters in fish.

SYSTEM DESIGN

Mechanical Design

The design goals of the FFM-sensor node had to ensure a size small enough for implantation into the peritoneum of fish to pick up the vital parameters without impairing the observed individual. Therefore we constructed the FFM sensor node of water resistant bio-compatible material build in a rapid prototyping process. We gave it the shape of a flattened ellipsoid with embedded flat electrodes avoiding sharp edges. With eight micro screws and a rubber seal we obtain an absolute water resistant packaging. The screws

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Cross sectional scheme of the sensor node.



Figure 2. Photo of the opened sensor node.

are covered on top and bottom with a plastic faceplate to ensure hermetic encapsulation. The overall dimensions of the sensor node are recently 50x20x12 mm and its weight is about 10 g therefore applicable for fish of about 1 kg weight [3]. Figure 1 and 2 show the main design of the sensor node.

Electric Design

We are recently using a 1 Ah battery as a power source. With a sample and communication rate of one second the lifetime is about 40 days. Within the next year we are planning a further miniaturisation as well as an already tested wireless power recharging technology.

To ensure a high data rate of 250 kbps required by a sample rate of once per second with eight analogue channels we decided to use digital radio frequency (RF) communication.

Because of the limited energy resources we optimised the system for low power consumption. Therefore we use an energy saving microcontroller MSP430F147 and a RF transceiver CC1101 from Texas Instruments. For a dataset of eight channels containing 16 bytes of data, the transceiver is only busy for one millisecond. Because of a small bandwidth of the vital parameter signals we applied two instrumentation amplifiers and seven operational amplifiers with a low gain bandwidth but even very low power consumption. Despite of an high signal attenuation through the fish and the water we can cover a volume of one cubic meter of the swimming tank with an average signal level of -70.52dBm.

Sensors

We integrated five gold electrodes of 1 cm diameter into the flattened upper and lower sides of the sensor node to measure the electrophysiological signals. For the DC resistance of the direct cell-environment we use two electrodes with a distance of 2.5cm. The measurement current is limited to 5µA thus avoiding a stimulation of the nervous system of the fish. The three electrodes at the other side measure the cellular electrical potential and the EMG whereas the middle electrode is the reference connected to the circuit ground. All sensors are sampled by an ADC with a resolution of 12Bit. For temperature measurements from 0°C to 40°C two NTC thermistors are connected directly to the housing for minimum thermal resistance. With a 3D acceleration sensor we determine the orientation in the swimming tank and the momentary acceleration of the fish. The resolution is 2mG or 0.019N. Two recently integrated pressure sensors at the front and back end of the sensor node will measure the longitudinal pressure waves caused by heart beats and breathing of the animal.

DATA PROCESSING AND ANALYSIS

For the first experimental and developmental tests our sensor system had to cover a measurement period of a minimum of three weeks and to record one dataset every second. For the RF communication we decided to use a frequency division (FDMA) method and a point to point connection for the first implementation. This technique is very energy saving, avoiding collisions and overhearing problems. We could prove a reliable data transfer from all sensor channels to the post processing in the RF gateway and the PC software. The conversion of the digital signals to the physical units is then done by the PC software.

To avoid the problems of regular calibration of all sensors we focus on the chronobiological analysis and evaluation of datasets. Doing this the only exactly calibrated and reliable measurement we need to have is the sample timestamp of every single measurement. The chronobiological regulation diagnostics [Balzer] uses the frequency distribution of regulation states over a certain time span for the assessment of the state of the whole organism.

Therefore first a set of time series analysis functions are applied to a selected time span (normally two times 20 s) of the sensor signals during a pre-processing step. We remove trends, determine the main period lengths of the parameters oscillation by autocorrelation and power spectral analysis while stepwise moving the observed time frame through the whole dataset. Finally we determine the frequency distribution of the calculated main period lengths for each 20 s interval. These frequencies are the main input for an artificial neuronal network (ANN) which then determines regulation states according to the Periodic System of Regulation States by Balzer [2] The output of the ANN is recently composed of 42 output nodes which stand for regulation states. The node with the highest output enumerator represents most likely that of the fish during the observed time period. The working cycle of the ANN corresponds to the sensor signal sample rate, so every regulation second the state of а single physiologic/behavioural parameter is calculated.



Figure 3. Frequency distribution of regulatory states for one fish at three different time periods. The data based on 3D acceleration (3dx, 3dy, 3dz), temperature (T) and skin resistance (SR).

To visualise the results achieved by this procedure we show exemplarily the frequency distribution of regulation states of one fish for the x-y-z acceleration parameters, the temperature and the tissue resistance in Figure 3. The three observation periods represent:

- a) half an hour after implantation, still under the influence of anaesthetics and deactivated
- b) three to four hours after implantation, fish is already recovering, in an activated state
- c) three hours after light off at the evening, fish is more deactivated (relaxed) and slightly more adaptive

These are very first and preliminary results. For fish species we will still have to test and 'teach' the ANN by a human interpreter who 'informs' the ANN about the 'meaning' of the observed behaviour or status of the fish on the basis of distinct experiments with defined stress situations. Thus the ANN will learn figuratively from the fish.

CONCLUSION

With the FischFITMonitor we could present a miniaturized implantable wireless sensor node with eight sensor channels for real-time monitoring of vital parameters in fish. We could show the functionality and reliability of measurements and the possibility of an evaluation of recorded data by means of the chronobiological regulation diagnostics leading to conceivable results in the sense of health and welfare status of fish. Next steps must include a further miniaturisation of the sensor (by 30%), the wireless power supply recharge and a time division based communication (TDMA) for multiple sensor nodes.

By specific experiments we will train the ANN to recognise stress situations and health impairments in different fish species.

Ethical Statement

The experiments comply with the German Guidelines for Animal Care and were approved by the Landesamt für Gesundheit und Soziales, Sächsische Str. 28-30, 10707 Berlin, Germany, Reg-No. G 0378/08.

- Balzer, H.U. Chronobiology as a foundation for and an approach to a new understanding of the influence of music. in: Haas / Brandes (ed.) *Music that works*, p. 1-76, (2009) Springer Wien ISBN 978-3-211-75120-6.
- Balzer, H.U., Struwe, R., Kuhne, F. Evaluation of regulation states in comparison with behavioural analysis determining states of stress sensitivity in dogs. In Proceedings of the 3rd European Conference on Behavioural Biology, Physiological Mechanisms in Behavioural Ecology. (2006).
- 3. Baras, E., Lagardere, J-P. Fish telemetry in aquaculture: review and perspectives. *Aquaculture International*, *3* (1995), 77-102.
- 4. Davis, M.W. Fish stress and mortality can be predicted using reflex impairment. *Fish and Fisheries*, *11* (2010), 1–11.
- Fredrich, F., Ohmann, S., Curio, B. and Kirschbaum, F. Spawning migrations of the chub in the River Spree, Germany. *Journal of Fish Biology*, 63 (2003), 710–723.
- MacIntyre C., Ellis T., North B.P., and Turnbull J.F.. The influences of water quality on the welfare of farmed trout: a Review. In: *Fish Welfare* E. Branson. Blackwell Scientific Publications, London, (2008), 150-178.
- 7. Winter, H.V., Fredrich, F. Migratory behaviour of ide: a comparison between the lowland rivers Elbe, Germany, and Vecht, The Netherlands. *Journal of Fish Biology*, 63 (2003), 871–880.

Spatial ECG of the Rat During Spatial Navigation

Thomas Gener generthomas@gmail.com Lorena Perez-Mendez

Diego Lopez-Pigozzi

Institut d'investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

Christoph Guger

Maria Victoria Sanchez-Vives

Institut d'investigacions Biomèdiques August Pi i

Guger Technologies, Graz, Austria

Sunyer (IDIBAPS), Barcelona, Spain & ICREA, Barcelona, Spain

ABSTRACT

In the present study our aim was to spatially characterize the electrocardiogram (ECG) during the free navigation of rats in an open space via 2D maps. The development of the necessary tools (hardware, analytical) to carry out this mapping should allow us in the future to obtain a physiological correlate of different behaviours. With that purpose, we recorded the ECG from Lister Hooded rats with three superficial electrodes while the rat was being tracked during free navigation in an 80x80 open field. The heart rate and heart rate variability were represented on relation to the spatial position in the field. These physiological measures provide information about the internal state of the animal during the performance of a particular task. In this presentation, we present analytical techniques to treat the ECG signal and to obtain heart rate variability and to represent it in 2D. A similar approach can be applied to the spatial representation of other physiological signals.

Author Keywords

ECG, stress, free movement, home base, behaviour measurement, behaviour analysis, 2D maps, navigation.

INTRODUCTION

Willem Einthoven developed, in the early 1900's, tools to observe and measure electrocardiogram (ECG) [1]. He received Nobel Prize in 1924 for his work. Since this time, ECG is a commonly recorded physiological measure. The calculation of heart rate variability (HRV), a measure of the variability of the interval between heart beats, provides information about the stress level, which is relevant information during behavioural tasks [2, 3]. On the other hand, cardiovascular rhythms have been described to determine the time of motor acts initiation [4, 5]. Therefore, it should be possible to find temporal correlations between heart rate changes and behaviour. Our aim here was to detect the possible correlation between physiological values and behaviour by characterizing the parameters of the ECG spatial map during the free navigation of rats in an open space.

DATA RECORDING

The electrocardiogram data (ECG) was recorded from the surface of the thorax of Lister Hooded rats (n=5) with 3 electrodes attached to the skin and held by a vest. Rats were cared for and treated in accordance with the Spanish regulatory laws (BOE 256; 25-10-1990) which comply with the EU guidelines on protection of vertebrates used for experimentation (Strasbourg 3/18/1986). To increase the contact between the skin and electrodes, we use a conductive paste, a high-chloride abrasive electrolyte gel (Abralyt HiCl 1000 gr., EASYCAP GmbH). The animal was placed in a square maze (80x80cm) and allowed to freely navigate, being motivated by randomly delivered pellets. A camera synchronized with the ECG signal recorded the movement of animal in the maze. Data of the tracking (X; Y position) and data of ECG (raw ECG data) were acquired. In our study, we used two different acquisition systems to record ECG data, one from Axona Ltd, England (complete solution recording and video tracking) and the other one from g.tec, Austria (g.USBamp -recording-, g.ANTS - tracking-) in order to compare the performance of both systems available in our laboratory and to improve the compatibility of the analysis on different formats of datasets.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Recording of heart activity. QRS complex and the RR interval are visible.

TEMPORAL ANALYSIS

From the ECG recorded along time we extract several parameters which in turn are temporal signals. First of all, the RR interval (interval between consecutive QRSs complexes) was obtained.

For calculating the RR intervals, QRS complexes have to be detected. For that, we used an algorithm developed by *GTEC* in Matlab. This software includes a preliminary automatic detection of the QRS, which was supervised later by an experimenter. From RR intervals we obtained heart rate and different definitions of heart rate variability (HRV) over 5 seconds windows [3]:

• Heart Rate
$$\rightarrow$$
 $HR = \frac{1}{RR}$

• Root Mean Square of the Mean Squared Difference of successive RR (RMSSD)

$$RMSSD = \sqrt{\frac{1}{\#samp} \sum_{i=1}^{\#samp} (RR_{i-1} - RR_i)^2}$$

• The proportion of the number of pairs of successive RRs that differ by more than 0.05 ms (pNN0.05)

$$pNN0.05 = \sum_{i=1}^{\#samp} \begin{cases} 0 \ if \ |RR_{i+1} - RR_i| < 0.05ms \\ 1 \ if \ |RR_{i+1} - RR_i| \ge 0.05ms \end{cases}$$

(#samp \rightarrow number of samples fitting in 5sec. windows)

Those variables of HRV reflect the stress of the animal doing a behaviour task [2].

2D MAPS

In order to study the relation of heart activity with position we transformed data from the time domain to the spatial domain (2D). As a first step, we drew raw data recorded by the tracking system to examine the trajectory of the animal (Figure 2).



Figure 2. Rat's trajectory (80x80cm).

Then, we divided the space into square bins (1x1cm) represented by positions in a matrix object. This matrix could board any kind of data to be represented in the space. Next, the time spent by the animal in every spatial bin was represented (Figure 3). This was constructed with tracking data, by summing up all the periods elapsed in the same bin (dwell time map).



Figure 3. Dwell time map in ms (80x80cm).

Afterwards, the physiological signals that we had previously obtained in the temporal analysis (see above): HR, RMSSD, and pNN0.05 were represented (Figure 4). For this purpose, we aligned in time the position signal with the physiological signal, taking into account the difference in the number of samples, and upsampling or downsampling if necessary. From this alignment, we can know which the position of the rat for every value of the desired signal is, and we can assign this value to the corresponding position of the matrix object. Besides, it is common that the animal goes through the same place (bin) more than once.



Figure 4. Heart rate map in bpm (right, 80x80cm) RMSSD map (left, 80x80cm).

Further manipulation of this 2D data lead to correlation maps between them.

CONCLUSION

We describe here a technique to acquire and analyze ECG to obtain the spatial mapping of the heart activity, and thus a measure of the internal state of the animal during navigation. The novelty of this technique is the 2D representation of heart rate and heart rate variability. Several protocols were carried out in our laboratory in order to study ECG and animal's spatial behaviour. The results show changes in different ECG parameters depending on the navigational pattern and position. The next step should be to carry out the analysis on real time in order to be able to modify stimulation depending on the animal's internal state. Real time would also allow feedback protocols based on ECG analysis. We plan to continue using the aforementioned techniques during the performance of different spatial behavioural protocols. Eventually, a third dimension (time) could be added to have a better

understanding of the physiological and behavioural correlations.

- 1. Einthoven, W., Ueber die Form des menschlichen Electrocardiogramms. Pflügers Archiv European Journal of Physiology, 60, 3 (1895), 101-123.
- 2. Kitney, R.I. and O. Rompelman, *The study of heart-rate variability*. 1980, Oxford: Clarendon Press. xiii, 246 p.
- 3. Aubert, A.E., et al., The analysis of heart rate variability in unrestrained rats. Validation of method and results. *Comput. Methods Programs Biomed.*, 60, 3 (1999), 197-213.
- 4. Florian, G., A. Stancak, and G. Pfurtscheller, *Cardiac* response induced by voluntary self-paced finger movement. Int. J. Psychophysiol., 28, 3 (1998), 273-83.
- 5. Pfurtscheller, G., et al., Does conscious intention to perform a motor act depend on slow cardiovascular rhythms? *Neuroscience Letters*, 468, 1(2010), 46-50.

Characterization of Foraging Behavior Of Yellowjacket Workers and Worker-Larva Trophallaxis Towards Feeding Baits Laced with Fipronil

Tomislav Curkovic tcurkovi@uchile.cl **Carolina Ballesteros**

Jaime Araya

Crop Protection Dept., College of Agriculture, University of Chile, PO Box 1004, Santiago, Chile

ABSTRACT

Foraging and trophallaxis behaviors towards a honeybee (HB) and a protein (P) based feeding baits were studied in Vespula germanica workers and larvae. Behavioral responses were described by observational methods and characterized by conventional statistics. Stereotypia indexes (SI) were estimated and contrasted within behaviors and between HB treatments. For foraging, the insecticide fipronil at two concentrations (LC20 and LC80 estimated for larvae) was contrasted against pure HB (control) but for trophallaxis, fipronil was compared at one concentration with the control. HB contaminated with a low concentration of fipronil did not significantly affect foraging and trophallaxis, allowing for larval exposure to this bait intended for workers. On the other hand, the P-based bait including fipronil at both concentrations was not foraged. The HB-bait loaded with the LC₈₀ was foraged but trophallaxis did not occur. These results demonstrate the value of quantitative behavioral studies to develop pest management techniques.

Author Keywords

Ethogram, stereotypia index, chi-square test, Fisher test, Vespula germanica.

INTRODUCTION

The yellowjacket wasp, *Vespula germanica* (Hymenoptera: Vespidae), is a social insect species that has invaded several regions around the world, which is considered a serious pest in Chile. It forages in the vicinity of the nest to collect and provide food to larvae. During foraging the workers can bite and sting people and animals. They can also affect agriculture production (fruit bites), apiculture (destruction of hives), etc. Available techniques to manage this species

include the use of non-specific traps to capture adults and nest destruction by either physical or chemical methods, but both are considered relatively inefficient. We have been working on developing a specific and efficient bait to attract V. germanica workers, and to allow them to forage, the collection of food at a bait station, and transport it back to the colony, which is followed by stages of trophallaxis, the exchange of food between members inside the colony, which distributes the bait to many individuals (larvae and adults), making it possible to deliver chemicals incorporated (e.g. insecticides) into the nest. To develop this technique, it was necessary to determine a baseline on feeding behaviors (both foraging and trophallaxis), and contrast behavioral responses on workers exposed to baits based on honeybee (HB) or a protein (P) matrix, laced with concentrations of fipronil, and eventually to distribute low amounts of the insecticide into the colony to decrease its vigor in a pest management strategy.

MATERIAL AND METHODS

Fipronil, a neurotoxic insecticide, was added into both liquid pure honeybee or a protein matrix (in the process of being patented), at concentrations equivalent to the LC_{20} and LC_{80} , which were estimated in previous studies with larvae, to test whether the use of a relatively low concentration of this toxicant affects workers and larvae differentially.

Nest comb pieces containing developed larvae, and workers marked with color tempera on the thorax dorsum were used in observation arenas (Flanders batteries), where both foraging and trophallaxis behaviors were videotaped. The tapes were revised to identify homogeneous, discrete, and observable steps in the sequences, and their respective frequencies were calculated. They were used to build ethograms for all treatments, which consisted on both concentrations of fipronil and the control without insecticide. To test if dependence occurred or not between the steps within behavioral sequences, the chi-square or Fisher tests were used. Also, a Stereotypia Index (SI) was calculated and used to test for eventual differences between

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org

treatments in a particular behavior (during foraging or trophallaxis).

RESULTS

For both behaviors, foraging and trophallaxis, dependence between the steps included within sequences was found for HB with no insecticide, and also for all HB-fipronil-loaded with the LC_{20} , but not in the case of trophallaxis with the LC_{80} , since workers did forage but did not conduct trophallaxis, and died soon after exposure. With the P-based bait, dependence between the steps occurred only for foraging and trophallaxis both without fipronil. These behaviors were not observed when the P-bait was contaminated with the insecticide at either concentration.

When workers approach and recognize the protein bait without fipronil, they cut small pieces of food with their mandibles, which they hold between their front legs to carry them to the nest to conduct trophallaxis. On the other hand, pure liquid honey was swallowed and transported to the colony inside the crop of the worker. During trophallaxis with the P-bait, workers on the comb first manipulate the matrix with their front legs and mandibles, then approach and offer a flattened bread-like "tortilla" to larvae which take a piece with their mandibles. This process is repeated by workers until the whole food piece is consumed. Trophallaxis for honey occurs after regurgitation of fluid from the crop. The SI values for foraging indicated a high degree of stereotypia for the pure HB-bait and HB with fipronil at the LC_{20} , but not for the LC_{80} treatment. With honey, trophallaxis at the control and the LC_{20} were also highly stereotyped, but this conduct was not observed at the LC_{80} . When trophallaxis did occur, it was observed between workers and larvae and between workers themselves. With the P-bait with no insecticide, the SI was also high for both foraging and trophallaxis. These behaviors were not observed when the P-bait was loaded with fipronil at any concentration.

By comparing SI between two behavioral responses (e.g. those occurring to the control vs. those occurring to the LC_{80}) using the chi-square test, we were able to discriminate between HB baits at two concentrations of fipronil to manage yellowjackets. In fact, we found that a low level of contamination with the insecticide (i.e. LC_{20}) did not alter significantly both foraging and trophallaxis. Although honeybee is not an appropriate matrix for preparation of feeding baits for this wasp (it also attracts bees and other insect pollinators), these data suggest that some sublethal concentrations of fipronil can be used to prepare feasible slow action feeding baits. Different specific feeding attractants to this yellowjacket species require to be studied.

Software Employment for Automatic Emotion Recognition in Online Psychotherapy

Lorena Alejandra Flores-Plata unam.flores@gmail.com Georgina Cárdenas López

mgcl@correo.unam.mx

National Autonomous University of Mexico, Faculty of Psychology, Av. Universidad 3004, Col. Copilco-Universidad CP 04510, Del Coyoacán, México, DF. 55+56222292

ABSTRACT

Online therapy has been used in an increasing manner and is considered a powerful tool when it is used in addition to, or is adequately integrated with, the presence of a therapist. Nevertheless, it can replace face-to-face therapy in the following cases: a) when there is no access to psychological services; b) when people prefer to keep their anonymity in the beginning of the therapeutic process; and c) when people have problems in identifying the therapeutic benefits that they can obtain and therefore have the first contact with professionals through the Internet.

The present study in progress, proposes the implementation of monitoring systems for facial variables and the use of emotion recognition techniques during the psychotherapy on line session in order to evaluate expression emotions in repeated measures, emotions and feelings that will provide objectivity to patient evaluation, treatment strategies, and will serve as guidelines of the therapeutic gains obtained during and after treatment. Correlation between self-reports and automatic measures will be discuss.

For this purpose, will be compared two groups (N=24) of moderate non suicidal depressive adult patients who are not under psychiatric treatment. Each group will have 16 sessions treatment under a cognitive-behavior therapy approach, conformed by homework, daily records on behaviors, feelings and emotions. The first group (12 participants) will be assigned to psychotherapy via Internet daily manual sessions with recordings with videoconference, as well as self-reports questionnaires; the second group (12 participants) will be assessed through manual and electronic recordings by Analysis of Facial

Expression Emotions in order to measure the facial variables with *FaceReader*.

Author Keywords

Facial expression emotion, E-therapy, cognitive-behavior treatment, depression.

INTRODUCTION

Online Mental Health Services, which originate from the concept of Telehealth, arises due to the necessity to offer and receive treatment despite distances or conditions which would make the encounter between a professional and a patient impossible. Currently, online therapeutic practices, in addition to covering patient's needs living far from the psychologist's office or clinic and seeking for treatment, have become motivating spaces for receiving treatment, whether it be due to the comfort it represents or the familiarity some may have with technology, which plays a fundamental role in these treatment alternatives.

Whenever we speak of therapeutic practice we refer to different types of services. [1] classifies them in the following categories: *e-therapy, e-counseling, mental health information websites, self-guided treatment program websites, online support groups* and *online mental health screening and assessments.* We shall say that the feature all of these share is the use of Internet as a means to facilitate psychotherapeutic services. These authors define Internet therapy as the interaction between a consumer and a therapist through Internet in association with a Web based structured treatment program, supported by hardware such as: calls through Internet, videoconferences, use of cell phones to carry useful information for the patient's problem and even fulfill therapy itself.

Studies performed on effectiveness and dissemination [1-7] point at online therapeutic services as a tool that may benefit persons who otherwise do not have access to these services (due to physical or economic circumstances) or have an interest for a service under this modality given the familiarity they may possess with this electronic medium.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Marks et al. [3] describe the development, application and effectiveness of three systems, of which two were acknowledged and recommended by England's National Institute for Clinical and Health Excellence, for depression disorders *Beating the Blues*; for anxiety disorders *Fear Fighter*, and a third one that has proved effectiveness in obsessive-compulsive disorders *OCFighter*. Finfgeld-Connett [8] carries on research concerning online treatments for alcoholism problems, where user therapists are nurses, advantages of Internet for intervening and favourable results of these applications are discussed. Likewise, the feasibility of using information retrieved from face to face interviews for the patient's benefit is posed.

Precisely in the necessity to reinforce assessment and planning, the question to be dealt with arising is the role played by nonverbal facial and bodily elements in online therapeutic relations, given that in online psychotherapeutic experiences patient and therapist may: a) not see each other's faces at any moment; b) encounter in person at some occasions, or; c) perform a treatment in person with Internet based support. In these three cases (exceptionally the third) the therapist has no steady information on the patient's facial expressions and the modifications that could occur in these during treatment. In therapeutic practice, emotions and their expression play a very important role, un a therapeutic session there exists a wide spectrum of information disposed not only in speech, but also in nonverbal language, where facial expression is crucial, and therefore since several years ago facial responses to emotions have been studied, either for adaptive biological matters or for reasons more of a socializing tint.

The introduction of FaceReader software as an automatic face recognition system is added (see Fig. 1) to an online treatment protocol for non suicidal depression with the aim of disposing of an additional measure to patient's records and determining whether this additional measure favours a better assessment, planning and success in online psychological therapy. The variables measured by this software are: happiness, sadness, anger, surprise, fear, repulse, "neutral" state.

We expect to possess new elements in elucidating an assessment and planning a treatment for therapists offering online therapeutic services.

PROCEDURE

Participants

24 participants of moderate non suicidal depressive adult patients who are not under psychiatric treatment.

- 12 participants will be assigned to psychotherapy via Internet sessions with daily manual recordings with videoconference, as well as self-reports questionnaires.
- 12 participants will be assessed through manual and electronic recordings by Analysis of Facial Expression Emotions in order to measure the facial variables with *FaceReader*.



Figure 1. FaceReader.

ASSESSMENT INSTRUMENTS

- Depression Inventory of Beck
- Anxiety Inventory of Beck
- Questionnaire of Social Validity
- Inventory of Quality of Life

Phases of Treatment

Phase 1. First contact and evaluation. The first contact through the Psychological Services Center of the Faculty of Psychology at National Autonomous University of Mexico; Phase 2. Implementation of treatment. The work of the treatment was carried out through specific cognitive-behavioral techniques (16 sessions) with facial emotion recognition and without facial emotion recognition; Phase 3. Evaluation after treatment. Implementation of the instruments used in the beginning, which provide qualitative and quantitative indicators about the patient's improvement, Phase 4. Follow the maintenance and improvement of the patient. Patients were contacted by mail and sent the assessments that were applied in the pre-evaluation, after 3, 6 and 12 months.

CONCLUSION

We expect to see future growth in online psychological practice in Mexico. Additional empirical studies are needed in order to establish ethical regulations for practicing psychology under this modality of service, to assess results, cost-effectiveness, and to investigate the limits of distance intervention. These would lead to better practice of the emergent modalities in these changing times.

ACKNOWLEDGMENTS

The present work is part of research carried on at the Virtual Teaching and Cyberpsychology Laboratory led by Georgina Cárdenas- López, Ph D.

- 1. Abbot, J., Klein, B., Ciechomski, L. Best practices in Online Therapy. *Journal of Technology in Human Services*, 26, 2/4 (2008), 360-375.
- 2. Barak, A., Hen, L., Boniel-Nissim, M. & Shapira, N. A comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. *Journal of Technology in Human Services*, 26, 2/4 (2008), 109-160.
- Medina-Mora, M., Borges, G., Lara, C., Benjet, C., Blanco, J., Fleiz, C., Villatoro, J., Rojas, E., Zambrano, J., Casanova, L. Y Aguilar-Gaxiola, S. Relevancia de los trastornos mentales y uso de servicios: Resultados de la Encuesta Nacional de Epidemiología Psiquiatrica en México. Salud Mental, 26 (2003), 1-16.

- Marks, I., Cavanagh, K., & Gega, L. Computer-aided psychotherapy: revolution or bubble. *British Journal of Psychiatry*, 191 (2007), 471-473.
- 5. Bell, V. Online information, extreme communities and Internet therapy: Is the Internet good for our mental health? *Journal of Mental Health*, *16*, 4 (2007), 445-457.
- Pousada, .M, Valiente, L. & Boixadós, M. Intervención en la salud en la Red: estado de la cuestión y perspectias del futuro. *Monográfico Intervención de la Salud en la Red* en <u>http://uocpapers.uoc.edu</u> (2006) (Recuperado el 18 de marzo de 2009).
- Cárdenas, G., Serrano, B., Flores, L. & De la Rosa, A. Etheraoy: A training Program for development of clinical skills in distance psychotherapy. *Journal of Technology in Human Services*, 26, 2/4 (2008), 360-375.
- 8. Finfgeld-Connett, D. Web-Based Treatment for Problem Drinking. *Journal of Psychosocial nursing*, 44, 9 (2006), 20-27.

Use of Environmental Light Cycles to Distinguish Circadian Regulation from Direct Light Modulation of Learning and Mood

Cara Altimus caltimus@jhu.edu Tara LeGates tlegate1@jhu.edu

Department of Biology, Johns Hopkins University, Baltimore MD, USA

Alfredo Kirkwood

Department of Neuroscience, Johns Hopkins University, Baltimore MD, USA kirwood@jhu.edu Todd Weber Department of Neuroscience, Rider University, Lawrenceville NJ, USA tweber@rider.edu

Sunggu Yang

Department of Neuroscience, Johns Hopkins University, Baltimore MD, USA abiyang9@gmail.com

Samer Hattar Department of Biology, Johns Hopkins University, Baltimore MD, USA shattar@jhu.edu

ABSTRACT

Previous work has suggested that circadian rhythms regulate cognitive functions and mood in human and mice. However, previous experiments housed mice in regular daily light: dark cycles preventing the dissociation of light from circadian rhythms. In light of recent findings in human, that light directly modulates attention and mood, we wanted to create light cycles which would allow the dissociation of light from circadian rhythms. Here we describe two light cycles, the first allows the measurement of behaviors in the absence of "white" light, and the second prevents mice from photoentraining but does not cause sleep disruption or circadian arrhythmicity. These cycles have allowed us to determine the circadian and light effects on learning and mood.

Abstract Keywords

Circadian rhythms, light, learning and mood.

INTRODUCTION

Short day length and irregular light schedules cause mood disorders, cognitive dysfunction, and fatigue as observed in seasonal affective disorder (SAD), shift work, and transmeridian travel. A common feature of all these situations relates to changes in the duration or time of light input. The circuitry and function of how light regulates cognitive functions in SAD, shift work and transmeridian travel has been confounded by sleep and circadian disruptions. To begin to understand how light and circadian rhythmicity affect shift work differentially, it is necessary to have light environments that allow the isolation of each factor. Historically, this has been vary difficult to accomplish as daily light cycles are used to synchronize circadian rhythms. We have utilized two previously characterized light cycles, one which allow animals to photoentrain with only two hours of light each day, and another which does not allow circadian photoentrainment, but does not disrupt circadian rhythmicity.

METHODS

In our first light cycle, we administer two light pulses each day, which outline the typical light period (ie: first light pulse from 700-800 and second light pulse from 1800-1900). The remainder of the day is illuminated with dim red light, which does not affect the circadian system. We are then able to use behavior tests with red illumination to determine the circadian regulation of cognitive functions without any confounding issues of light on the circadian system.

The second light cycle consists of an ultradian light schedule of 7 hours (T7: 3.5 hours of light followed by 3.5 hours of darkness). We have shown previously that mice housed in this light cycle sleep equivalent amounts when compared to mice housed in a 24 hour light cycle. Using this light cycle we have found that mice are unable to predict the onset of lights and consequently adopt a 24.5 hour period. This slightly longer than 24 hour period allows the 3.5 hour light pulses to hit all phases of the circadian cycle within two days. In this setup, we are able to test the effect of chronic light pulses on behavior without the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



confounding factors of sleep disruption and circadian arhythmicity.

FINDINGS

Using these light cycles we have found that circadian phase does not influence the ability of mice to learn spatial or recognition tasks. However, mice housed in the T7 light cycle show decreased ability to learn a spatial task and are unable to discriminate objects in a simple recognition task indicating that chronic light pulses disrupt the ability of mice to form hippocampal dependent memories. Previously hippocampal dependent learning and memory has been linked to depression related increases in glucocorticoid levels. We therefore asked if the changes found in learning were due to increased depression like behavior in mice housed in a T7 cycle. We used forced swim tests and corticosterone measurements to assay depression like behaviors. Mice housed in the T7 light cycle found showed increased time spent floating and increased levels of corticosterone in indicating that chronic exposure to light pulses increases depression like behaviors which then leads to an decreased ability to learn. To test the hypothesis that depression like behaviors precede the decreased ability to learn, we treated the depression behavior with an antidepressant, desipramine, prior to performing the learning assay. We found that mice housed in the T7 cycle that were treated with antidepressants, showed a complete recovery in their ability to learn hippocampal dependent tasks.

CONCLUSIONS

Our ability to house mice in light cycles which dissociate circadian photoentrainment from the direct influence of light on behavior has allowed us to show that circadian rhythmicity does not affect hippocampal dependent learning and that light modulates learn and mood in mice, confounding previous experiments on the topic.

Ethical Statement

All work described here has been approved by the Johns Hopkins University IUCAC committee.

- Redlin, U. & Mrosovsky, N., Masking of locomotor activity in hamsters. J Comp Physiol A, 184, 4 (1999), 429-437.
- 2. Altimus, C.M. *et al.*, Rods-cones and melanopsin detect light and dark to modulate sleep independent of image formation. *Proc Natl Acad Sci USA*, *105*, 50 (2008), 19998-20003.
- 3. Beeler, J.A., Prendergast, B., & Zhuang, X., Low amplitude entrainment of mice and the impact of circadian phase on behavior tests. *Physiol Behav*, 87 5 (2006), 870-880.
- 4. Foster, R.G. & Wulff, K., The rhythm of rest and excess. *Nat Rev Neurosci*, 6, 5 (2005), 407-414.
- 5. Ma, W.P. *et al.*, Exposure to chronic constant light impairs spatial memory and influences long-term depression in rats. *Neurosci Res*, 59, 2 (2007), 224-230.
- Fonken, L.K. *et al.*, Influence of light at night on murine anxiety- and depressive-like responses. *Behav Brain Res* (2009).

Postnatal Exposure to Clonazepam Leads to Cognitive Deficits in Adult Rats

Martin Šubrt martin.subrt1@gmail.com Anna Mikulecká nmikul@biomed.cas.cz Hana Kubová kubova@epilepsy.biomed.cas.cz

Institute of Physiology, Academy of sciences of the Czech republic Vídeňská 1083, Praha 4, 14220 Czech Republic

INTRODUCTION

Benzodiazepines (BZs) belong to widely prescribed drugs in all age groups. Under the specific condicition they can be used during pregnancy and early postnatally in humans [1].

Clinical studies suggest possible risk of BZs exposure in early postnatal period for further development and behavior. Therefore, we designed the present study to examine possible cognitive alterations in adult rats after the short-term exposure to clonazepam (CZP) during development period (P7-P11).

METHODS

Animal and Drug

Wistar immature male rats were daily injected with CZP for five consecutive days (P7-P11). CZP, freshly suspended in saline with Tween 80, was administered intraperitoneally in three doses -0.1, 0.5, 1.0 mg/kg. Control group received a corresponding volume of saline. Body weight was checked daily during the CZP administration.

Habituation Test

Habituation is defined as a response decrement following continuous or repeated exposure to spatial novelty[2]. In our experiment, habituation was assessed in two ways: within-session and between-session habituation[3]. Between-session habituation to a novel environment is used as a model of non-associative learning.

We tested rats (P67-P70) for four consecutive days in an open field (OF). We placed rats individually in the center of OF and registered their behavior for 10 minutes. We evaluated following parameters: locomotor activity (distance moved in OF), exploratory activity (rearing) and displacement behavior (grooming). The evaluation was carried out using Ethovision software (Noldus Information Technology).

Morris Water Maze (MWM)

This test was used to investigate spatial memory. It is based on the premise that animal evolves an optimal strategy to explore the water maze and escape from the water onto submerged platform [4,5].

Animals were tested at the age of P81-P85. Each rat received 8 consecutive trials per day for 5 consecutive days. A trial began by placing the animal into 1 of 4 randomly selected sectors (north, west, south, or east; each starting point was used no more than 2 times per session) with the head facing towards the wall. The rat was allowed to swim until it found the escape platform or had swum unsuccessfully for 60 s. In the latter case, we guided a rat to the platform by hand and recorded its latency as 60 s.

RESULTS

Irrespective of the treatments, all animals showed withinsession habituation in the 4th session. Between-sessions habituation was found only in control animals. CZP exposed animals had deficit between-sessions habituation.

In the acquisition sessions (MWM), all animals reached the platform across sessions which suggest that they learned the task i.e. to escape onto the platform. However, animals exposed to CZP in dose of 1.0 mg/kg spent more time swimming. In retention test performed 10- and 40 days after acquisition sessions no differences were found between controls and CZP rats.

In conclusion, postnatal exposure to CZP led to cognitive deficit in a dose and task dependent manner.

ACKNOWLEDGMENTS

This study was supported by grant No. 305/09/0846 of the Grant Agency of the Czech Republic and project LC554 of the Ministry of education of the Czech Republic

REFERENCES

1. Riss J, Cloyd J, Gates J, Collins S: Benzodiazepines in epilepsy: pharmacology and pharmacokinetics. *Acta Neurol. Scand.*, *118* (2008), 69-86.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

- 2. Cerbone A, Sadile AG: Behavioral habituation to spatial novelty: interference and noninterference studies. *Neurosci. Biobehav. Rev.*, *18* (1994), 497-518.
- 3. Thiel CM, Muller CP, Huston JP, Schwarting RK: High versus low reactivity to a novel environment: behavioural, pharmacological and neurochemical assessments. *Neuroscience*, *93* (1999), 243-251.
- 4. Kubová H, Mareš P, Suchomelová L, Brožek G, Druga R, Pitkanen A: Status epilepticus in immature rats leads to behavioural and cognitive impairment and epileptogenesis. *Eur. J. Neurosci.*, *19* (2004), 3255-3265.
- 5. Morris R. Developments of a water-maze procedure for studying spatial learning in the rat. J. Neurosci. Methods, 11 (1984), 47-60.

Studying the Modulation of Brain Rhythms by Dynamic Cues

Diego Lopez-Pigozzi¹ dlopez1@clinic.ub.es Thomas Gener¹

Maria Victoria Sanchez-Vives ^{1,2}

Lorena Perez-Mendez¹

Julita De La Vega¹

¹Institut d'investigacions Biomèdiques August Pi i Sunyer, IDIBAPS, Barcelona, Spain ²ICREA, Istitució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

ABSTRACT

It is well known that the theta band (4-12 Hz) is a brain rhythm associated to the self-movement of the animal [1] as well as to several cognitive functions [2]. The aim of this study has been to determine the possible modulation of this rhythm in relation to dynamic cues. With that purpose we employed an e-puck robot as a dynamic cue in contrast to the few precedent studies where dynamic stimuli were presented in a screen [4]. To take advantage of this technique we created behavioural protocols to assure the capacity of the rats to track the movement of the robot. During the protocols the behavior of the animal was characterized by recording the electroencephalogram (EEG). The EEG signal was filtered and the relation between the theta band and the robot's movement was then analyzed. This combination of tools, as the behavioural tasks including mobile robots as the EEG recording, result in a powerful technique to reveal the brain mechanisms involved in the tracking of dynamic cues.

Author Keywords

Dynamic cues, robot, theta band, hippocampus, movement and electrophysiology.

INTRODUCTION

Despite the fact that the tracking of moving objects has been profusely studied in various animal species, little attention has been paid to it in rodents and the first experiments began in the last decade [6]. Most of the studies used images presented in a screen as the stimuli [4] and the aim was to understand how the visual system tracked moving cues. Only few recent works have used real objects in the behavioural protocols. When the "object" was another rat the experimenters found some issues because the tendency of the animals was to do social contacts [7]. These factors lead us to the substitution of the rat by a robot and to design an original configuration to separate the space where the object is moving from the place in which the subject performs the discrimination task by a see-through cylinder (Figure 1A). There are different possible zones of the brain where the information of the dynamic cues could be processed. We are particularly interested on the spatial





Figure 1. (A) Configuration of the behavioural protocol. The robot is covered by a white cylinder to increase its size and visibility. The rat should position its head above the operant platform to allow the beginning of the task (see black circle under the animal's head). (B) Learning curve showing the percentage of correct choices for three different subjects in colors and the averaged curve in black.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 2. Control model of the protocols. In the upper left corner the block of the tracking system sent the data to the customed algorithm (MATLAB S-Function) that control the robot in the middle right (Command). Also a trigger is sent to the Labview software in the lower right corner (To Instrument) that controls the behavioural devices, infrared sensors and water deliverers (from Guger Technologies, GTEC).

components of the object tracking on relation to the own animal's position. Large evidence has been accumulated that places spatial processing during navigation in the hippocampal and parahippocampal areas [5]. The visual information necessary for object tracking probably converges in the hippocampus. On one hand is important to know the position of these objects (the dorsal path) and on the other hand the identification of the object (the ventral path) [2]. For these reasons we choose the hippocampus as the place to look for the neuronal correlates of tracking moving objects.

BEHAVIOURAL TASK

We control the protocols by Labview and Matlab interfaces. The movement of the robot was remotely controlled by the task execution via Bluetooth. To know how the rat behaves we track its position. A diagram of the simulink model can be seen in the Figure 2 (GTEC, Graz, Austria). To start a trial the subject must remain in a platform placed in the frontal part of the cylinder, facing the robot. After one second in the platform the robot starts its movement randomly to the left or the right and the rat must go to obtain reward in the water dispenser of the same side (Figure 1A). We trained the animals for at least two weeks before they reached stable performances over 80% of correct choices, n=3 (Figure 1B). At this point the animals are chronically implanted as we describe in the next section.

ELECTROPHYSIOLOGICAL RECORDINGS

To record the EEG during the protocols we implanted the trained rats with microdrives (AXONA. Ltd, London, UK) carrying tetrodes under deep anesthesia (ketamine / xylazine). This commercial scaffold allows us to move the four mounted tetrodes of Platinum/Iridium (90%/10%, California Fine Wire) below the cranium searching the CA1 zone of the hippocampus. In this area theta is a dominant

rhythm and there are neurons (place cells) that fire when the animal is located in specific locations of the environment (place fields) [5]. We acquired the EEG signal during the behavioural protocols with an amplifier at 1200 Hz of frequency. The recordings were synchronized with the behavioural protocols in order to relate the behaviour with brain activity.

In Figure 3A the onset of the robot's movement is highlighted with a white line. We can see the averaged





power of theta for correct trials where a period of two seconds is considered, one before the movement of the robot and one after it. By testing the spectrogram in the presence / absence of the robot we are currently determining the specific modulation of the rhythms exerted by the robot movement which is distinguishable from the rat's movement and attention.

RESULTS AND CONCLUSIONS

We have achieved to combine protocols of sensory discrimination by using robots as dynamic cues with electrophysiological recordings. Our early results show that there could be a modulation of the theta band secondary to the robot's movement. In Figure 3 we illustrate an example where the theta band has a different power distribution during the experimental and control protocols. While in the control without robot the power increases at least 700 ms after the sound presentation in the robot's protocol there is an immediate increase after the onset of the movement sustained for 200 ms. This response is homogenous for different animals (n=3) but is not yet significant. Future experiments will try to quantify the origin of the theta band modulation by a dynamic cue.

- 1. Buzsaki, G. Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus*, *15*, 7 (2005), 827-840.
- 2. Knierim . Neural representations of location outside the hippocampus. *Lern. Mem.*, 13 (2006), 405-415.
- 3. Maurer, A.P., *et al.* Organization of hippocampal cell assemblies based on theta phase precession. Hippocampus, *16*, 9 (2006), 785-794.
- 4. Nekovarova, T. and D. Klement Rat's operant behavior can be controlled by the configuration of objects in an animated scene displayed on a computer screen. *Physiol Res.* (2005).
- 5. O'Keefe JA and Conway DH. Hippocampal place units in the freely moving rat: why they fire when they fire. *Exp. Brain Res*, *31* (1978), 573–590.
- Takanishi et al. Interaction between creature and robot. Proceedings of the 1998 IEEE/RSJ. Intl. Conference on Intelligent Robots and Systems, 1998.
- 7. Telensky et al. Enemy avoidance task: A novel behavioral paradigm for assessing spatial avoidance of a moving subject. *J Neurosci* (In press).

Combinatorial Measurement of Sleep and Wheel Running Activity to Examine the Interaction Between Light and Drug Administration in Mice

Tara LeGates

Department of Biology, Johns Hopkins University 3400 N. Charles St., 228 Mudd Hall, Baltimore, MD 21218 USA tlegate1@jhu.edu +1-410-516-7641

ABSTRACT

A billion dollar industry exists to manipulate sleep and alertness. These drugs are psychoactive compounds that work by interacting with neurotransmitter transporters or receptors in the brain to alter brain activity. Possibly, the most common and widely used example of one of these psychoactive compounds is caffeine. Caffeine has been shown to diminish drowsiness while increasing alertness. In high doses, it can cause increased activity and even tremors [1-4]. Another potent influencer of sleep and wakefulness, though commonly overlooked, is light. In humans, light input is known to increase alertness, while in nocturnal animals, such as mice, light input has the opposite effect [5-6]. Interestingly, an acute light pulse is sufficient to induce a dramatic increase in sleep in mice [6]. What is not known, however, is whether light and drugs interact to influence sleep.

METHODS AND RESULTS

Previous work has shown that administration of a 3-hour light pulse two hours after the lights offset is sufficient to induce a decrease wheel running activity, an effect termed masking[7]. Recent work from the lab has shown that a similar light pulse results in a decrease in the amount of time spent awake, which has been used to explain the decrease in wheel running activity [6]. To examine the potential interaction between light and drug administration, I utilized this masking light pulse. Specifically, I administered either caffeine or modafinil [8], both of which are psychostimulants, to mice in absence of as well as prior

Samer Hattar

Departments of Biology and Neuroscience (JHMI), Johns Hopkins University 3400 N.Charles St., 227 Mudd Hall, Baltimore, MD 21218 USA shattar@jhu.edu +1-410-516-4231

to the 3-hour masking light pulse and measured wheel running activity as well as sleep.

To measure wheel-running activity, adult male mice are individually housed in cages with a running wheel and access to food and water ad libitum. Mice are allowed to acclimate to the cage for one week prior to testing and are kept in a light cycle of 12-hour light: 12-hour dark. Treatments are always separated by one week to allow mice to readjust to baseline situation. Wheel running activity is recorded using Vitalview (MinitMitter; Respironics).

To measure sleep in mice, a two channel EEG and one channel EMG (Pinnacle Technology) head mount is affixed to the skull of mice while under ketamine/xylazine-induced anesthesia. Mice were allowed two weeks to recover in a 12:12LD cycle. Mice were then transferred to sleep recording cages, tethered with a preamplifier and allowed to acclimate to the cage for one day. The following night, treatments were administered and sleep measured. The sleep/wake stage of the mouse is determined by the experimenter based on the frequency and amplitude of EEG and EMG using Neuroscore (DSI). Sleep can essentially be separated into two stages, slow wave sleep (SWS) and rapid eye movement (REM). SWS is defined by low frequency (1-4 Hz), high amplitude EEG with little to no activity detected by EMG, whereas REM is defined by high frequency (8Hz) low amplitude EEG and no EMG. Wake is defined as high frequency (8-10Hz) low amplitude EEG and high frequency EMG.

Mice received an injection (i.p.) of caffeine 15-minutes prior to the 3-hour light pulse. Caffeine administration in the absence of a light pulse as well as vehicle administration were used as controls. The same procedure was used to examine modafinil.

In summary, I found that the administration of high levels of caffeine increased wakefulness in the absence of a light pulse and was sufficient to keep mice awake in the presence of a light pulse. However, this amount of caffeine decreased

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

wheel running activity in the absence of a light pulse and activity was decreased further during a light pulse. Administration of a decreased concentration of caffeine does not change wheel running activity or wakefulness in absence of a light pulse. Interestingly, while it is sufficient to keep mice awake during a light pulse, there is no change in their wheel running activity. These results were also found upon administration of the stimulant modafinil.

As the administration of a 3-hour light pulse has been shown to decrease both wheel running activity and sleep, these findings allow us to actually separate these two behaviors. This is in addition to the new insight that these two measures provide in the examination of the interaction between light and psychoactive compounds.

All animal experiments were done according to the institutional regulations of Johns Hopkins University (Baltimore, MD).

- 1. Rasmussen, JL; Gallino, M "Effects of caffeine on subjective reports of fatigue and arousal during mentally demanding activities". *European Journal of Clinical Pharmacology*, *37*, 1 (1997), 61-90.
- White BC, Simpson CC, Adams JE, Harkins D Jr. Monoamine synthesis and caffeine-induced locomotor activity. *Neuropharmacology*, 17, 7 (1978), 511-513.

- 3. Goldstein A, Kaizer S, Warren R. Psychotropic effects of caffeine in man. II. Alertness, psychomotor coordination, and mood. *J. Pharmacol. Exp. Ther.*, *150*, 1 (1965), 146-51.
- 4. Huang ZL, Qu WM, Eguchi N, Chen JF, Schwarzschild MA, Fredholm BB, Urade Y, Hayaishi O. Adenosine A2A, but not A1, receptors mediate the arousal effect of caffeine. *Nat. Neurosci.*, *8*, 7 (2005), 858-859.
- 5. Lockley SW, Evans EE, Scheer FA, Brainard GC, Czeisler CA, Aeschbach D. Short-wavelength sensitivity for the direct effects of light on alertness, vigilance, and the waking electroencephalogram in humans. *Sleep*, 1, 29, 2 (2006), 161-168.
- 6. Altimus, C.M., et al., Rods-cones and melanopsin detect light and dark to modulate sleep independent of image formation. *Proc. Natl. Acad. Sci. USA*, *105*, 50 (2008).
- 7. Redlin, U. and N. Mrosovsky, Masking of locomotor activity in hamsters. *J Comp Physiol* [A], 184, 4 (1999).
- 8. Caroline Kopp, Jean-Marie Petit, Pierre Magistretti, Alexander A. Borbe'ly, Irene Tobler Comparison of the effects of modafinil and sleep deprivation on sleep and cortical EEG spectra in mice. *Neuropharmacology*, 43 (2002), 110-118.

Measuring Stress and Cognitive Load Effects on the Perceived Quality of a Multimodal Dialogue System

Andreea Niculescu niculescuai@ewi.utwente.nl +31534894654

Betsy van Dijk e.m.a.g.vandijk@ewi.utwente.nl +31534893781 Yujia Cao y.cao@ewi.utwente.nl +31534892826

Anton Nijholt a.nijholt@ewi.utwente.nl +31534893686

HMI Group, University of Twente Postbus 217, 7500 AE, Enschede, The Netherlands

ABSTRACT

In this paper we present the results of a pilot study investigating the impact of stress and cognitive load on the perceived interaction quality of a multimodal dialogue system for crisis management. Four test subjects interacted with the system in four differently configured trials aiming to induce low/high levels of stress and cognitive load. Physiological sensors and subjective ratings were collected to measure the level of stress and cognitive load. After each trial the subjects filled in an evaluation questionnaire regarding the system interaction quality. In the end we conducted an in-depth interview with each subject. The trials were recorded with a webcam to facilitate the behaviour analysis. Results showed that both factors had an influence on the way subjects perceived the interaction quality, whereas the cognitive load seems to have a higher impact. Further quantitative experiments are needed in order to validate the results and quantify the weight of each factor.

Author Keywords

Multimodal conversational interactions, qualitative evaluations, behavior analysis.

ACM Classification Keywords

H.5.1 Multimedia information systems: Animations, Audio input/output, Evaluation/methodology, H.5.2 User Interfaces: Evaluation/methodology, Graphical user interfaces, Natural language, Voice I/O.

INTRODUCTION

The quality assessment of interactive systems is a complex construct of interdependent factors relying on system design and performance, and user perception. Among these factors are the cognitive load and stress experienced by users during the interaction [1].

In the literature these two factors are often summarized and measured together under the global concept of 'cognitive demand'. There are surely no doubts that these two factors are related but their relationship is not exclusive: stress can be caused not only by a highly loaded cognitive task but also by frequent input recognition mistakes or poor sound quality [2], whereas a highly loaded cognitive task would be perceived as stressful only in situations considered as exceeding available resources. [3]. Therefore, our long term research goals are to investigate whether stress and cognitive load can be successfully manipulated and measured separately. Also, we are interested in the impact these two factors might have on the perceived interaction quality of a multimodal dialogue system.



Figure 1. System screen shot.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

METHODS

Since crises represent situations in which people experience high levels of stress and cognitive load they offer perfect test environments for our experiment. Accordingly, we developed a small prototype of a multimodal dialogue system for crisis management. The system has attached an embodied conversational agent representing the crisis manager (see figure 1). The crisis manager provides information about a crisis event, such as event description, geographical maps, available rescue resources and estimated number of victims. Users can interact with the system using speech or mouse clicks and receive information in the form of text, speech, images or videos.

Experiment Design

The experiments were configured in four trials. Each trial combined several parameters in order to achieve low/high stress (S) and cognitive load (CL) (see table 1).

| Trial no. | Factor combination |
|-----------|--------------------|
| 1 | low CL / low S |
| 2 | low CL / high S |
| 3 | high CL / high S |
| 4 | high CL / low S |

Table 1. Combinations of cognitive load and stress levels per trial.

To manipulate the stress factor we used six parameters, such as background noise, speech speed, speech length, time limitation, simulated recognition mistakes and dramatic event description. For the cognitive load manipulation we varied task complexity and presentation format. A detailed description of the trials configuration can be found in [8]. A compulsory break of a minimum of five minutes was placed between the trials to lower the stress level.

Participants

Four male test subjects, aged between 24 and 30, participated in the experiment. All of them had a technical background, but vague or no idea about crisis management. Also, they were not particularly familiar with the use of multimodal dialogue systems; only one subject had spoken to a computer before. Due to the fact that our pilot study was performed with a small number of participants we decided to perform the trials in the same order for all four subjects (see table 1). However, in future we plan to counterbalance the trial order among participants to avoid learning effects or other biases that might arise by being exposed to a certain factor combination before the others.

Measurements

To collect measurements we used physiological sensors, questionnaires, a two-session in-depth qualitative interview and behavior analysis.

- **Physiological sensors**. We used the heart rate variability (HRV) [4, 5] as an indicator for the cognitive load and the galvanic skin response (GSR) [6, 7], as a measurement of the stress.
- Questionnaires. The NASA (TLX) questionnaire was used to control the results collected from the physiological sensors. TLX contains six workload-related factors: mental, physical and temporal demands, own performance, effort and frustration. Statements concerning subjects' concentration and degree of tiredness, ease of the system use, overall system quality and degree of understanding between subjects and the system were added to the TXL questionnaire. The factors were rated on a 20-point scale. The questionnaires were filled in after each trial.
- In-depth interview. Each test subject was interviewed based on a qualitative open questionnaire. The interviews were recorded on tape and manually transcribed. The interview aimed to explore relationships between the induced level of stress and cognitive load, and the overall interaction quality. The interview was performed in two sessions: before and after the experiment. In the first session subjects were interviewed about their expectations and background knowledge concerning the experiment topic (crisis management) and the use of multimodal conversational interfaces. In the second session subjects were asked about expectation fulfilling, problems encountered during the interaction, system comprehensibility and transparency, content informativeness, information presentation and interaction easiness. In the end the subjects could make additional comments if they had any.
- **Behavior analysis.** From the log files and videos analysis we extracted various parameters concerning response competition time, reaction time, number and type of errors, total number of words, verbal hesitations, breaks and mispronunciations. The videos also facilitated the qualitative analysis of several other behavioral cues, such as speaking style (polite, rude, key-words vs. sentences) verbal and non-verbal reactions to system errors or increased task difficulties, gestures and gaze.

RESULTS AND DISCUSSION

Our results showed that our manipulation was successful for the cognitive demand, and only partly for the stress.

Also, both stress and cognitive load were better indicated by subjective rating than by physiological measurements (for more details see [8]).

In the first trial the planned manipulation was altered by a "first impression" effect, the trial achieving a much higher level of stress than expected¹. Thus, the effect intended for trial 1 (low CL/low S) was instead achieved in trial 2.

¹ In the future we plan to add a "base-line" trial at the experiment's begin in order to avoid getting unplanned manipulations.

Accordingly, three out of four subjects ranked the interaction quality with the system for trial 2 as being the best. Trial 3 had the highest level of stress and cognitive load as perceived by subjects. However, only one subject rated the interaction quality in this trial as being the lowest; the other subjects chose instead trial 4 (high CL/low S) as having the lowest system interaction quality. This finding might lead to the conclusion that the cognitive load could have a bigger impact on the perceived interaction quality compared with the stress. Nevertheless, repeated measurements with a higher number of subjects are needed to confirm this assumption.

Further, the interview showed that subjects considered the system as presenting relevant and informative content in a clear and systematical way, a fact that decreased the cognitive load, according to one subject. They did not encounter communication problems, except for one subject and were pleasantly surprised to know how to handle the system right from the beginning. The interaction with the system was in general perceived as being easy and 3 subjects enjoyed it. However, the system was criticized as lacking basic functionalities (such as zooming in the crisis maps or help options) and being not transparent and not flexible enough: during the third trial most subjects were unsure how to answer the system's questions and were not allowed to return to previous conversation stages to ask for clarifications. Another negative point was the synthesized voice of the agent reminding the subjects unpleasantly that they were talking with a machine.

Analyzing the subjects' behavior we observed that they acted congruently to the golden rule "treat others as you want to be treated"; for example, one subject who expressed the wish to be treated politely used polite markers, such as "thank you" and "please" during the entire interaction (even in stress loaded trials); another subject wishing the system to present only facts and no other redundant information used only keywords or extremely short sentences. In general, subjects displayed a very different behavior in terms of performance and reactions to stress: some expressed their frustration using loud verbal expressions or showing a constant "joke" attitude; others became impatient and started clicking the mouse button to "increase" the interaction speed; some remained apparently calm, showing their response to stress only through frowning. Interestingly, most of the speech disfluencies, breaks and errors were made in the low stress condition and mostly by two of the subjects. These two subjects had in general a poor performance completing the trials and gave lower quality rankings, as compared with the other two subjects. The verbal response time values were very different among the subjects, but showed a common trend: they were lower in high stress conditions and higher in high cognitive load circumstances.

CONCLUSIONS

We performed a pilot study on the way users perceive the quality of the interaction with a multimodal dialog system for crisis management while being exposed to stress and cognitive load variations. Our manipulation was successful for the cognitive demand, and only partly for the stress. We encountered difficulties in achieving accurate stress manipulations, as stress appears to be a highly complex phenomenon to which humans respond very differently. For our experiment we used physiological sensors, questionnaires, qualitative interviews and behavioral analyses. Since objective measurements methods, (i.e. physiological sensors) could not provide meaningful results additional quantitative investigations are required to determine whether stress can be measured apart from cognitive load. Also, further analyses are needed to validate and weight the factors' impact on the interaction quality assessment.

ACKNOWLEDGMENTS

This research was funded by the EC Artemis project on Human-Centric Design of Embedded Systems (SMARCOS, Nr. 100249).

- 1. Moeller, S.: Quality of telephone-based spoken dialogue systems, Springer, New York, 2005.
- 2. Kryter, K.D.: The handbook of hearing and the effects of noise: Physiology, psychology, and public health. Academic Press New York, 1994.
- 3. Lazarus, R.S.: Theory based stress measurement. *Psychological Inquiry*, 1 (1990), 3-13.
- Scerbo, M.W., Freeman, F.G., P.J Mikulka., R. Parasuraman, and F. Di Nocero,: The efficacy of psycho-physiological measures for implementing adaptive technology. TP-2001-211018, (2001).
- Wilson, G.F. and F.T. Eggemeier: Psycho-physiological assessment of workload in multi-task environments. In: Damos, D.L. (ed.): Multiple-task performance. CRC Press, 1991.
- 6. Verwey, W.B. and H.A Veltman.: Detecting short periods of elevated workload. A comparison of nine workload assessment techniques. *Applied Experimental Psychology*, 2 (1996), 270-285.
- Boucsein, W., Haarmann, A., Schaefer, F.: Combining Skin Conductance and Heart Rate Variability for Adaptive Automation During Simulated IFR Flight. In: Harris, D. (ed.) HCII 2007 and EPCE 2007. LNCS, 4562 (2007), Springer, Heidelberg, 639–647..
- Niculescu, A.I., Y. Cao and A. Nijholt Manipulating Stress and Cognitive Load in Conversational Interactions with a Multimodal System for Crisis Management Support, in Development of Multimodal Interfaces: Active Listening and Synchrony, A. Esposito, N. Campbell, C. Vogel, A. Husain and A.Nijholt (eds), LNCS, vol. 5967 (2010), Springer, Heidelberg, 134-148.

A Systematic Review on Randomization and Permutation Tests in the Educational and Behavioral Sciences

Ming Huo ming.huo@ped.kuleuven.be

Wim Van den Noortgate wim.vandennoortgate@kuleuven-kortrijk.be Mieke Heyvaert mieke.heyvaert@ped.kuleuven.be

Patrick Onghena patrick.onghena@ped.kuleuven.be

Katholieke Universiteit Leuven Versaliusstraat 2, Leuven, Belgium

BACKGROUND

In many educational and behavioral studies, the assumptions of the classical parametric hypothesis tests (e.g., normality, homogeneity of variance, independence of errors) are often considered implausible [1, 2, 3, 4]. An alternative to the traditional statistical methods that does not rely on such strict assumptions is to use a randomization test (RT) or permutation test (PT). RTs and PTs constitute a set of distribution-free statistical tests that calculate the probability of getting a value as extreme or more extreme than an obtained value of a test statistic under a null hypothesis by recalculating the test statistic for all or many permutations of the data. They do not depend on a specific error distribution, and they use the original values of the data instead of the ranks. RTs and PTs were proposed in the early twentieth century, but were not widely used until much later. This is mostly because (a) they were too computationally intensive, (b) their applicability was limited to simple scenarios, (c) and they could be replaced by the available classical nonparametric tests based on ranks [5].

Over the past two decades, RTs and PTs have received much attention in the educational and behavioral sciences, with an accompanying multitude of applications. However, a general overview of the theoretical development and applications of RTs and PTs in the educational and behavioral sciences is still lacking, mainly because articles on RTs and PTs are spread out over the literature. Accordingly, a systematic review is called for.

Aim

There are three objectives of this paper: 1) to provide an overview of the theoretical development of RTs and PTs

and summarize several key areas of theoretical research; 2) to summarize several active areas of educational and behavioral applications of RTs and PTs; 3) to identify the experimental designs in which RTs and PTs have been applied.

METHODS

In order to realize the above-mentioned three objectives, the databases ERIC, PsycINFO and Web of Science were searched for articles on RTs and PTs, published in the educational and behavioral journals between 1989 and 2008. Searches were performed by using the keywords [randomization tests] and [permutation tests]. Articles written in languages other than English were excluded. Abstracts were read to identify the relevant articles. This review was carried out in two tracks: a theoretical track and an application track. The first track intends to summarize the theoretical evolution of RTs and PTs. The second track focuses on applications of RTs and PTs and intended to summarize the experimental designs as well as areas in which these methods have been applied.

RESULTS

124 articles were identified, which included 87 theoretical articles and 37 application articles. In the theoretical articles, seven major topics were identified: a) introduction to and instruction of RTs and PTs (e.g., [3], [6]); b) algorithms, programs and software for RTs and PTs (e.g., [7], [8]); c) RTs and PTs for group designs (e.g., [9], [10]); d) RTs for single-case designs (e.g., [11], [12]); e) multivariate RTs and PTs (e.g., [13], [14]); f) performance of RTs and PTs (e.g., [1], [15]); and g) advanced topics (e.g., [16], [17]). In the application articles, RTs and PTs were applied in the following three active areas: a) models of vocational interest structure (e.g., [18], [19]); b) eventrelated potential (ERP) and electroencephalogram (EEG) (e.g., [20], [21]); and c) animal behaviors (e.g., [22], [23]). Meanwhile, RTs and PTs were found to apply for the following experimental designs: a) one-group design and association analysis (e.g., [24]); b) paired-group designs

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.
(e.g., [25]); c) multivariate designs and multiple comparisons (e.g., [14]); d) distance matrices analysis (e.g., [22]); and e) single-case design (e.g., [26]).

DISCUSSION

From a theoretical perspective, methodological possibilities of RTs and PTs have been extended over the last 20 years. Among those articles, a majority of them focused on the implementation of RTs and PTs, RTs for single-case design and RTs and PTs for group designs, whereas multivariate RTs and PTs as well as advanced topics were less discussed.

From an application perspective, RTs and PTs have been employed in several active areas of behavioral research. However, compared with the range of topics discussed in the theoretical part, both the application areas and types of experimental design are limited. Some applications of RTs and PTs can be attributed to the theoretical development of RTs and PTs in the educational and behavioral sciences, such as the area of models of vocational interest structure. Among those application articles, RTs and PTs were used as alternatives to parametric hypothesis tests in order to avoid the stringent distributional assumptions.

CONCLUSIONS

The methodological possibilities of RTs and PTs have been extended substantially during the last years, RTs and PTs are not only applied in simple (e.g., two-group design) but also in complex contexts (e.g., multivariate designs). Moreover, recent developments of RTs and PTs have widened their application scope. RTs and PTs are employed in some more complex and exciting areas, such as ERP and EEG, since RTs and PTs are powerful tools to solve the multiple comparisons problem in these areas. While many theoretical articles were published in educational journals, more applications were published in behavioral journals.

- 1. Adams, D. C., & Anthony, C. D. Using randomization techniques to analyse behavioural data. *Animal Behaviour*, *51* (1996), 733–738.
- Edgington, E. S., & Onghena, P. *Randomization tests* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC (2007).
- Hunter, M. A., & May, R. B. Statistical testing and null distributions: What to do when samples are not random. *Canadian Journal of Experimental Psychology*, 57 (2003), 176–188.
- 4. Manly, B. F. J. *Randomization, bootstrap, and Monte Carlo methods in biology* (2nd ed.). New York: Chapman & Hall, 1997.
- Welch, W. J. Construction of permutation tests. *Journal* of the American Statistical Association, 85 (1990), 693– 698.

- 6. Bear, G. Computationally intensive methods warrant reconsideration of pedagogy in statistics. *Behavior Research Methods, Instruments & Computers, 27* (1995), 144-147.
- 7. Cai, L. Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation. *Behavior Research Methods*, *38* (2006), 51–59.
- 8. Chen, R. S., & Dunlap, W. P. SAS procedures for approximate randomization tests. *Behavior Research Methods, Instruments & Computers, 25* (1993), 406–409.
- Mielke, P. W., & Berry, K. J. Permutation tests for common locations among samples with unequal variances. *Journal of Educational and Behavioral Statistics*, 19 (1994), 217–236.
- Mielke, P. W., & Berry, K. J. Data-dependent analyses in psychological research. *Psychological Reports*, 91 (2002), 1225–1234.
- 11. Onghena, P., & Edgington, E. S. Randomization tests for restricted alternating treatments designs. *Behaviour Research and Therapy*, *32* (1994), 783–786.
- 12. Onghena, P., & Edgington, E. S. Customization of pain treatments: Single-case design and analysis. *Clinical Journal of Pain*, *21* (2005), 56–68.
- 13.Blair, R. C., Higgins, J. J., Karniski, W., & Kromrey, J. D. A study of multivariate permutation tests which may replace Hotelling's T² test in prescribed circumstances. *Multivariate Behavioral Research*, 29 (1994), 141–163.
- Blair, R. C., & Karniski, W. An alternative method for significance testing of wave-form difference potentials. *Psychophysiology*, 30 (1993), 518–524.
- 15. Peres-Neto, P. R., & Olden, J. D. Assessing the robustness of randomization tests: Examples from behavioural studies. *Animal Behaviour*, 61 (2001), 79–86.
- 16. Johnston, J. E., Berry, K. J., & Mielke, P. W. Permutation tests: Precision in estimating probability values. *Perceptual and Motor Skills*, 105 (2007), 915– 920.
- 17. Manly, B. F. J. Comments on a note on permutation tests of significance for multiple regression coefficients by long, et al. *Psychological Reports*, *101* (2007), 1041–1042.
- Hedrih, V. Structure of vocational interests in Serbia: Evaluation of the spherical model. *Journal of Vocational Behavior*, 73 (2008), 13–23.
- Yang, W. W., Stokes, G. S., & Hui, C. H. Cross-cultural validation of Holland's interest structure in Chinese population. *Journal of Vocational Behavior*, 67 (2005), 379–396.

- 20. Maris, E. Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, *41* (2004), 142–151.
- 21. Henderson, L. M., Yoder, P. J., Yale, M. E., & McDuffie, A. Getting the point: Electrophysiological correlates of protodeclarative pointing. International *Journal of Developmental Neuroscience*, 20 (2002), 449–458.
- 22. Mitani, J. C., & Brandt, K. L. Social-factors influence the acoustic variability in the long-distance calls of male chimpanzees. *Ethology*, *96* (1994), 233–252.
- 23. Sendova-Franks, A., & Franks, N. Spatial relationships within nests of the ant Leptothorax-unifasciatus (Latr)

and their implications for the division-of-labor. *Animal Behaviour*, *50* (1995), 121–136.

- 24. Boyd, J. P., Fitzgerald, W. J., & Beck, R. J. Computing core/periphery structures and permutation tests for social relations data. *Social Networks*, 28 (2006)., 165–178.
- 25. Murrihy, R., & Byrne, M. K. Training models for psychiatry in primary care: A new frontier. *Australasian Psychiatry*, *13* (2005), 296–301.
- 26. Cicchetti, D., Rosenheck, R., Showalter, D., Charney, D., & Cramer, J. Interrater reliability levels of multiple clinical examiners in the evaluation of a Schizophrenic patient: Quality of life, level of functioning, and neuropsychological symptomatology. *Clinical Neuropsychologist*, 13 (1999), 157–170.

Pattern Recognition Methods – A Novel Analysis for the Pupillographic Sleepiness Test

Jarek Krajewski krajewsk@uni-wuppertal.de

University Wuppertal Experimental Industrial Psychology Gaußstraße 20, 42097 Wuppertal Germany Thomas Schnupp t.schnupp@fh-sm.de

University of Applied Sciences Schmalkalden Neuroinformatics and Signal Processing Blechhammer, 98574 Schmalkalden, Germany

David Sommer d.sommer@fh-sm.de

Christian Heinze c.heinze@fh-sm.de

University of Applied Sciences Schmalkalden Neuroinformatics and Signal Processing Blechhammer, 98574 Schmalkalden, Germany

ABSTRACT

The aim of this paper is to improve the information gained by the most commonly applied fit-for-duty sleepiness test (Pupillographic Sleepiness test, PST) by using pattern recognition approaches. The pupil diameter based sleepiness detection is enriched by several new features and machine learning methods. Using all newly computed pupil diameter features we achieved on the two-class detection problem (moderate sleepiness vs. high sleepiness) an accuracy of 83.03% on participant-dependent data with a Random Forest classifier. This result suggested that the PST-standard feature set should be enriched by the here proposed enlarged feature set.

Author Keywords

Pupil diameter, sleepiness detection, machine learning.

ACM Classification Keywords

H.5.m Information interfaces and presentation: Miscellaneous

INTRODUCTION

Measuring sleepiness has been recognized as an important factor for the prevention of a broad range of traffic accidents. Hence, many efforts have been reported in the literature for developing sleepiness detection systems. One of the most promising fit-for-duty tests – the pupillographic sleepiness test (PST) focuses on instability of pupil size [1-3]. The background of this method is that in an alert participant, the pupil remains dilated in darkness with amplitude of change below 0.3 mm and a frequency of approximately 1 Hz. In sleepy subjects, the pupil shows spontaneous oscillations with predominantly low-frequency components and amplitudes reaching several millimeters. Furthermore, the pupil diameter (PD) decreases with time. These changes are measured by infrared video pupillometry during a recording session of usually 11 min. To quantify the sleepiness-induced changes the Pupillary Unrest Index (PUI), a feature reflecting spontaneous oscillations of pupil diameter, is most often applied (even if a few other features as e.g. interpolation rate, lnPUI, relative PUI are sometimes proposed).

Sebastian Schnieder

s.schnieder@uni-wuppertal.de

University Wuppertal

Experimental Industrial

Psychology

Gaußstraße 20, 42097 Wuppertal

Germany

Martin Golz

m.golz@fh-sm.de

Nevertheless, little empirical research has been done to examine the benefit of signal processing and pattern recognition based "brute-force" methods in addition to the original approach. Thus, the aim of this study is to apply multiple state-of-the-art pattern recognition methods [4-7] on the PST based detection of sleepiness. Attention is drawn particularly on the comparison of several commonly applied classifiers. The rest of this paper is organized as follows: Section 2, 3, and 4 describes the experimental design, feature extraction and classification. After the results of the sleepiness detection are provided in Section 5, the paper closes with a conclusion and a discussion of the future work in Section 6.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

METHOD

Twenty-seven students, recruited from the University of Wuppertal (Germany), took part in this study voluntarily. Initial screening through a questionnaire excluded those having sleep disorders or sleep difficulties (PSQI). The participants were instructed to maintain their normal sleep pattern and behaviour. Due to recording and communication problems, the data of the 6 participants could not be analyzed.We conducted a within-subject sleep deprivation design (8.00 p.m to 4.00 a.m). During the night of sleep deprivation a well-established, standardized selfreport fatigue measure, the Stanford Sleepiness Scale (SSS), was used by the subjects and two experimental assistants just before the recordings to determine the reference value of sleepiness. On this scale, a score of 1 point indicates "feeling active and vital, alert, wide awake" and a score of 7 points indicates "almost in reverie, sleep onset soon, losing struggle to remain awake". For training and classification purposes, the records were further divided into two classes according to the informative values of the PST: low and moderate sleepiness (LS) and high sleepiness (HS) samples with the boundary value SSS \geq 5.0 (5 samples per subject; total number of samples: 111 samples; 77 samples LS, 34 samples HS). During the night, the subjects were confined to the laboratory and supervised throughout the whole period. Between sessions, they remained in a room, watched DVD, and talked. Non caffeinated beverages and snacks were available ad libitum.

Feature Extraction. Due to measurement errors, e.g. during eye lid closures, the PD is preprocessed by deleting these segments and correcting them through linear interpolation. The following pupillometric variables (PST standard features; PST-SF) are computed: mean pupil diameter, the PUI, the interpolation rate (due to blinking) and the square root of the power within the frequency band 0–0.8 Hz in pupillary oscillations (fatigue waves). Furthermore, we calculated 'brute-force' signal processing based features of pupil diameter within the time, frequency and state space domain (PST-BF). This assignment follows the first processing step of computing frame level descriptors, independent of feature characteristics of the second, contour describing, functional based processing step.

Time domain features $(97\#+97\Delta\#)$. Within the time domain the following features can be extracted: regression descriptors (e.g. regression slope, intercept, maximum of regression error), class distribution measures (e.g. number of values within pupil diameterbin 0.0-0.1), peak amplitudes and distances (e.g. mean distance of peaks; maximum of peak amplitude), entropy, zero crossing distances and slope (e.g. maximum of distance between consecutive zero crossings; mean velocity of pupil diameterin zero crossings).

Frequency domain features (935#+935\Delta#). To capture fine temporal changes of spectral descriptors we performed a framing and windowing of the signal (frame size = 512, frame shift = 256, hanning window), and computed the

power spectral density per frame. The resulting frame-level descriptors (FLDs) were aggregated to FLD-contours. The next processing step captures temporal information of the FLD contours by computing functionals. Frequently used functionals are percentiles (quartiles, quartile ranges, and other percentiles), extremes (min/max value, min/max position, range), distributional functions (number of segments/ intervals/reversal points), spectral functionals (DCT coefficients), regression functions (intercept, error, regression coefficients), higher statistical moments (standard deviation, skewness, kurtosis, and zerocrossingrate), means (arithmetic mean and centroid). This procedure of combining FLDs and functionals results in 935 frequency domain features as e.g.: relative and absolute power spectral density (PSD) of raw and first derivate contours in 30 spectral bands (e.g. minimum of relative PSD of first derivate of 0.6-0.7 Hz spectral band contour), band energy ratios (PSD 1-5hz/ PSD 0-1 Hz), spectral flux (e.g. max of spectral flux = Euclidean distance of PSD between consecutive frames), and long term average spectrum descriptors (e.g. skewness of PSD distribution).

CLASSIFICATION

We conducted a validation experiment to examine whether automatically trained models can be used to recognize the SSS based classification of moderate vs high sleepiness. Our approach can be summarized in four steps: 1. Collect individual PST data and the associated sleepiness ratings for each participant; 2. Extract relevant features from the pupil diameter data (PST-SF vs. PST-BF); 3. Apply a correlation filter for feature selection (correlation > .30: PST-SF 4 features vs. PST-BF 44 features remaining); 4. Build PST-SF and PST-BF based classification models using dichotomized SSS in order to solve the two-class detection problem of LS vs. HS; 5. Test the learned models on unseen PST data. Classifiers typically used within pattern recognition based biosignal analysis include a broad variety of dynamic algorithm (Hidden Markov Models) and static classifiers. When choosing a classifier within this highly correlated and noisy feature space, several aspects might be of importance such as low memory, low computation time, quick converging, and no suffering from overfitting. With respect to these requirements, we applied the following static classifiers from the popular 4.5 RapidMiner [8] software using standard parameter settings: Support Vector Machines ('LibSVM', rbf kernel function; 'JMySVM', linear kernel function; 'FastLargeMargin' [9], linear kernel: 'W-SMO'. Sequential Minimal Optimization), Kernel Logistic Regression ('MyKLR'), Multilayer Perceptrons ('NeuralNetImproved', 2 hidden sigmoid layer, 5 nodes each), k-Nearest Neighbors ('NearestNeighbors'; k 3). Decision Trees = ('RandomForest', 800 trees), Naive Bayes ('NaiveBayes', 'KernelNaiveBayes'), and Fuzzy Lattice Reasoning ('W-FLR'). In a participant-dependent validation protocol, we applied a stratified 10-fold crossvalidation. The final

classification errors were calculated averaging over both classifications.

RESULTS

In order to determine the detection performance, different classifiers were applied on the 2064 features. The recognition rate (RR) of the different classifiers for the two class prediction problems is computed (see Table 1). Within the applied classification schemes the Random Forest classifier using the enlarged brute-force signal processing feature set reached the highest RR of 83.03%. The average RR benefit using the PST-BF instead of PST-SF feature set. For the PST-SF feature set, the 3-Nearest Neighbor classifier achieved the highest recognition rate of 77.42%. Within all applied classification schemes the PST-SF feature set achieved 72.96%, the PST-BF 79.86% accuracy (average improvement 6.9%).

| Classifier | Specification | PST-SF | PST-BF | Δ |
|------------|-------------------------------------|--------|--------|-------|
| SVM | LibSVM, kernel = rbf | 73.94 | 79.24 | 5.30 |
| SVM | FastLargeMargin, kernel = linear | 61.26 | 80.23 | 18.97 |
| SVM | JmySVM, kernel = linear | 75.68 | 79.32 | 3.64 |
| SVM | Sequential Minimal Optimization | 73.79 | 77.58 | 3.79 |
| KLR | MyKLR | 75.76 | 81.21 | 5.45 |
| NN | k = 3 | 77.42 | 81.97 | 4.55 |
| | NeuralNetImprove | | | |
| MLP | d, 2 hidden layer, 5 nodes | 73.87 | 75.76 | 1.89 |
| NB | NaïveBayes | 72.12 | 81.06 | 8.94 |
| KNB | KernelNaïveBayes | 68.33 | 81.86 | 13.53 |
| | Fuzzy Lattice | | | |
| FLC | Reasoning | 73.86 | 79.12 | 5.26 |
| RF | #Tree = 500 | 74.77 | 83.03 | 8.26 |
| | Average Classifier | 72.96 | 79.86 | 6.90 |

Table 1. Correct recognition rate (RR) (in %) of several classifiers on the test set using participant-dependent validation schemes. PST-SF: 4 standard PST features; PST-BF: 44 features (4 standard + 40 brute-force PST feature).

DISCUSSION

The main findings of the present study may be summarized as follows. First, using all pupil diameter features and all samples (without pre-selecting prototypical classes out of the whole database) we achieved on the two-class detection problem (moderate sleepiness vs. high sleepiness) an accuracy of 83.03% on participant-dependent data with a Random Forest classifier. This result suggested that the PST-standard feature set might be enriched in prospective studies by the here proposed enlarged feature set including known features as relative PUI, logarithm of PUI (lnPUI), logarithm of relative PUI and PSD of several frequency bands [10]. Our classification performance is in the same as has been obtained for comparable tasks, e.g. for sleepiness classification using speech, steering behavior, EEG, posturographic information, cf. [11-14].

Our results are limited by several facts. The present results are (which is a truism) preliminary and need to be replicated using enlarged databases and a greater diversity of participants. We relinquish to optimize the performance of SVM by a fine grained hyperparameter optimization. Therefore, the performance has to be recognized as lower border of the SVM capabilities. Moreover, it would seem advisable that future studies address the main topics of enriching the pupil diameter feature set with eye blinking and lid behaviour features. These camera based features carry further information and might therefore contribute to a higher detection rate of sleepiness. This performance gain might be probably higher than adding other fancy classifier for this sleepiness detection task (e.g. maximum-likelihood bayes classifiers, fuzzy membership indexing, hidden markov models, gaussian mixture density models).

- Wilhelm B, Lüdtke H, & Wilhelm H (1999). Spontaneous papillary oscillations. An objective measure for the level of tonic central nervous activation. In J Kuhlmann & M Böttcher (Eds.), Pupillography: Principles, Methods and Applications (pp. 27–36). München: W. Zuckerschwerdt Verlag.
- Regen, F (2009). Assoziation zwischen Pupillen-Unruhe-Index (PUI) und Korrelaten des zentralnervösen Aktivierungsniveaus im Wach-EEG. Dissertation, FU Berlin.
- Krajewski, J, Peters T & Wilhelm B (2009). Standardkennwerte des Pupillographischen Schläfrigkeitstest: Ergebnisse einer Schlafdeprivationsstudie. In G. Mayer, & T. Penzel (Eds.), Somnologie Schlafforschung und Schlafmedizin (p. 49-50). Würzburg: Steinkopff Verlag.
- 4. Sommer D (2009). Analyse des Mikroschlafs mit Methoden der Computergestützten Intelligenz. Aachen: Shaker.
- Zeng Z, Pantic M, Roisman G. I. & Huang T S (2009) A survey of affect recognition methods: Audio, visual, and spontaneous expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1):39–58.
- Krajewski J, Batliner A, & Wieland R (2009). Multiple classifier applied on predicting microsleep from speech. In ICPR (Eds.), 19th Conference on Pattern Recognition (4 pages, no pagination). Tampa/ Florida: IEEE Computer Society Press.
- Steidl S (2009) Automatic classification of emotionrelated user states in spontaneous children's speech. Logos Verlag, Berlin,. PhD thesis.

- Mierswa I, Wurst M, Klinkenberg R, Scholz M, & Euler T (2006) YALE: Rapid prototyping for complex data mining tasks, in Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06).
- Fan RE, Chang KW, Hsieh CJ, Wang XR, & Lin CJ (2008) Liblinear: A library for large linear classification, Journal of Machine Learning Research 9.
- 10. Lüdtke, H (2005). Mess- und Analyseverfahren in der Pupillografie. Aachen: Shaker.
- Krajewski J, Batliner A, & Golz M (2009). Acoustic sleepiness detection – Framework and validation of a speech adapted pattern recognition approach. Behavior Research Methods, 41, 795-804.
- 12. Krajewski J, Golz M, & Sommer D. (2009). Detecting sleepy drivers by pattern recognition based analysis of

steering wheel behaviour. In A. Lichtenstein, C. Stößel, & C. Clemens (Eds.), Der Mensch im Mittelpunkt technischer Systeme (*S. 288-291*). Fortschritt-Berichte VDI, Reihe 22, Mensch-Maschine-Systeme. Düsseldorf: VDI-Verlag.

- 13. Sommer D, Golz M, Schupp T, Krajewski J, Trutschel U, & Edwards D. (2009). A measure of strong driver fatigue. Proceeding International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 4, 9-15.
- 14. Schnupp T, Schenka A, Edwards D, Krajewski J, & Golz M. (2009). Is posturography a candidate for a vigilance test? World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings, 25 (9), 388-392.

Group Preparedness For Risk in the Environment of Social Instability

Natalya Miheyeva

Faculty social psychology, Daugavpils University, Latvia natmih@inbox.lv

ABSTRACT

This article analysed primary theoretical sources with the view of operationalisation and systematisation of the notions of risk as well as the notions that are allied and interdisciplinarilly connected to it. In the outcome of the theoretical study both classical and the most recent risk studies and correlating events have been analysed and systematised staying in the focus of the study against the backdrop of social instability.

Besides, different factors influencing group cohesion and organizability of behaviour in risk situation are studied.

Therefore, this article is seen as an attempt to analyze a risky situation and recognize whether the phenomenon of risk is the factor of organisational behaviour. And still there are relevant though unsolved theoretical and practical issues concerning the problem of individual psychological disposition to risk in connection with individual personal peculiarities and social conditions..

Author Keywords

Risk, preparedness for risk, social instability, group, decision making.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Trackmate (V. 1.0): A Versatile Program for Automated Tracking of Behaviour in Laboratory Rodents

Esther Remmelink VU University, Amsterdam, The Netherlands e remmelink@student vu nl

Paul Halasz Motion Mensura Ltd, Australia

ABSTRACT

An increasing array of commercial software packages allow video tracking and automated analysis of rodent behaviour in standard models such as the elevated plus maze, radial arm maze and conditioned place preference paradigms. Trackmate is a new program developed as a collaboration between Motion Mensura Ltd and the University of Sydney Psychopharmacology Laboratory and can be used to automate data gathering in a wide variety of behavioural paradigms involving rodents. Trackmate is written in the LabviewTM programming environment and can be run on any adequately powered PC. Users are required to install Labview RunTIme software and Apple Quicktime, both of which are free downloads. The only other requirement is an appropriate USB video grabbing device, many of which are available at very low cost. We typically run the software on MacMiniTM computers (2.8 MHz, 4GB RAM, 320 GB hard disk) with KWorld DVD Maker 2TM USB interfaces. Trackmate allows data to be gathered in real time from up to 4 camera inputs, via an appropriate quad splitter box, allowing 4 individual animals in different apparatus to be tracked simultaneously (see Figure 1). For each animal, a blob (corresponding to the body of the animal) is detected as well as the centre of mass and head direction of the rodent. Up to 4 different regions of interest (ROIs) can be defined for each animal in the apparatus in which they are located. Location and distance travelled by the blob and centre of mass of each animal within each of these ROIs can be measured and logged. "Blob Logic" functions allows sophisticated measures to be made at the intersection of ROIs: for example "head out" behaviour in the emergence test and object investigation in novel object recognition test.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Michael Bowen

Iain McGregor

University of Sydney, Australia

Robert Dielenberg



Figure 1. Screenshot of Trackmate while simultaneously tracking 4 rats in separate cat odour boxes.

Trackmate logs an exhaustive array of test results to Excel compatible spreadsheets and also allows hard disk recording of raw video footage to be made in real time for archival purposes. Travel paths of animals can also be recorded. *Trackmate* is easy to set up and has been successfully used in various Australian universities to acquire data in a variety of paradigms including: locomotor activity and drug-induced behavioural sensitization, home cage activity, elevated plus maze, open field test, conditioned place preference, novel object recognition, Morris water maze, forced swim test, and predator odour avoidance.

Trackmate Social (V 0.9): Beta Software for the Automated Measurement of Rodent Social Behaviour

Michael T. Bowen School of Psychology, University

of Sydney, NSW 2006 Australia mbowen@psych.usyd.edu.au Esther Remmelink VU University, Amsterdam, The Netherlands esrem@hotmail.com

Robert A. Dielenberg Motion Mensura Ltd, Cooks Hill, NSW 2300, Australia

Paul Halasz VU Motion Mensura Ltd, Cooks Hill, NSW 2300, Australia Iain S. McGregor School of Psychology, University of Sydney, NSW 2006 Australia

ABSTRACT

Many commercial software packages now exist to allow video tracking and automated analysis of behaviour in individual rats and mice across a variety of behavioural paradigms. However, automated analysis of social behavior in pairs, triplets or quads of rodents is technically much more challenging and has proved difficult to successfully achieve due to the inherent challenges presented by developing algorithms capable of successfully distinguishing between, and tracking multiple subjects simultaneously. At the same time, there is increasing interest psychopharmacology, psychology in and neurobiology to characterise the effects of gene manipulations, stressful interventions and drug treatments on sociability in rodents. Trackmate Social is a new program under development as a collaboration between Motion Mensura Ltd and the University of Sydney Psychopharmacology Laboratory and is being used to automate data gathering and analysis of social interaction in rats and mice. It has successfully overcome some of the challenges presented by this type of tracking through use of sophisticated algorithms that allow simultaneous tracking of multiple subjects. Trackmate Social is written in the LabviewTM programming environment and can be run on any adequately powered PC. Users are required to install Labview RunTImeTM software and Apple QuicktimeTM, both of which are free downloads. The only other requirement is an appropriate USB video grabbing device, many of which are available at very low cost. We typically run the software on MacMini[™] computers (2.8 MHz, 4GB RAM, 320 GB hard disk) with KWorld DVD Maker 2TM USB interfaces. Trackmate Social can be used to analyse social behaviour in groups of 2-4 rats or mice (see Figure 1). For each animal in the group, a blob (corresponding to the body of the

animal) is detected as well as the centre of mass and head direction of each individual rodent. The distance between blobs over time is automatically computed allowing proximity measures and time spent in dyads, triplets or quads to be derived. Various regions of interest can also be defined in the test environment, allowing time spent by social groups in specific regions to be measured. Trackmate Social logs an exhaustive array of test results to Excel compatible spreadsheets and also allows hard disk recording of raw video footage to be made in real time for archival purposes. Ongoing research and development is aimed at automating the detection of specific social behaviours such as adjacent lying, anogenital sniffing (see Figure 1), head to head exploration and chasing/following. Trackmate Social has recently been used in a variety of experiments, including (1) demonstrating increased huddling behaviour in groups of rats exposed to predator odor and other stressors, (2) showing decreased social interaction in adolescent rats given chronic paroxetine, and (3) showing increased anogenital investigation in unfamiliar compared to familiar rats in the social interaction test.



Figure 1. Screenshot of Trackmate Social logging anogenital investigation in an unfamiliar pair of rats.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

The Use of Wildlife Cameras to Monitor Pig Behavior During Commercial Long Distance Transport

Katja van Driel

Richard Heath

Janet Talling

katja.vandriel@fera.gsi.gov.uk

The Food and Environment Research Agency, Sand Hutton, York, YO41 1LZ, UK,

Heleen van de Weerd

Peter Kettlewell

ADAS Drayton, Alcester Road, Stratford-Upon-Avon, Warwickshire, CV37 9RQ SAC, Sir Stephen Watson Building, Bush Estate, Penicuik, Midlothian, EH26 0PH

Malcolm Mitchell

ABSTRACT

In this paper we describe how a wildlife stills camera can be used to monitor behavior during long distance transport. Wildlife stills cameras are normally used to capture images of wild animals *in situ*. A number of features inherent in their design allows for the adaptation of them for monitoring behavior during the transport of live pigs on commercial long distance journeys.

Author Keywords

Wildlife camera, commercial long distance transport, pigs, behavior.

INTRODUCTION

Within the European Union millions of slaughter and breeder pigs are transported annually. The duration of transport can vary widely from a few hours up to several days. European regulations EC1/2005 [1] differentiate between short journeys, 8 hours, and longer journeys, those over 8 hours. These require the use of a higher specification vehicle i.e. one with fans and drinkers onboard. However it is unknown whether there should be a maximum total length of any one journey. For example breeder pigs, are transported for 11 days, moving from Western Europe to Russia. A report of the Scientific Committee on Animal Health and Welfare [2] found a paucity of information on the welfare of pigs during long journeys and the need for further research. A project was set up to identify the risk factors associated with long journeys. An epidemiological rather than experimental approach was taken as this would allow collection of pertinent data from a large number of commercial journeys, rather than a large amount of detailed variables from a limited number of experimental journeys. A new technique to record behavior during commercial journeys lasting for up to 11 days had to be developed.

METHODS

Working closely with commercial hauliers restricted the options for data collection and lead to a novel solution. To capture all the animals' behaviour during a journey video recorders would need to be used. There are two major problems with video cameras, firstly, they would need a large independent power supply, and secondly they would need a massive memory capacity to capture images from an 11 day journey. Direct observations would require a team to follow each transport from start to finish and the behaviour of the animals could only be observed at stops rather than throughout the journey.

A solution was found in the use of a digital stills camera originally designed to take images of wildlife in the field. This device should allow recording of the behaviour of the animals during the journey, no requirement for research technicians to be present at the end of the journeys, and information on the behaviour patterns of the animals.

These types of camera have been used by wildlife researchers and biologist for some years under field conditions. The Reconyx professional digital infrared camera (PC85) was chosen, as it is a self-contained unit with a built-in, independent battery power supply, which can last at least 10 days at low temperatures. For digital stills cameras the power supply is usually the limiting factor; memory cards can hold 1000s of images (250-400 kb per image), sufficient for 11 day journeys. The Reconyx has a built-in automatic infrared flash, and so can take pictures in low light conditions without disturbing the subjects.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Figure 1. Placement of PC-85 with protective housing and brackets.

The camera can be set to take pictures either on detection of movement or at programmed time intervals. Each picture is stamped with the time, date, temperature and a moon-phase icon. Once programmed and started there is no requirement for further monitoring of the device.

Work was needed with the cameras and housings to fit them to the vehicles to be used for the research. Although the camera is supplied in a plastic housing a further layer of metal protection was needed to shield them from the pigs and potential damage. This also allowed for brackets to be fitted to mount the housing in the truck (see Figure 1).

To obtain the best angle and field of view for the cameras, a mock-up of a truck interior was created and test photographs shot at different angles and heights using a model pig were taken. The ideal position for fitting was found to be at a height of 90cm in the corner of the pen pointing towards the opposite corner of the pen. The view covers most of the pen apart from a small amount of space in the corners to the left and right of the field of view (see Figure 2). It was originally thought that a universal bracket could be made to fit all types of livestock container, however detailed examination of several vehicles has indicated that some customisation will be required.



Figure 2. Showing field of view of pen, with pigs.



Figure 3. Comparison of pig behavior recorded at different time intervals.

A validation trial was undertaken on a 24-hour transport of slaughter pigs. The camera was fitted by the experimenter, left on the truck for the journey and removed by the haulier, and then the images were downloaded by the haulier and transferred via the internet to the experimenter.

RESULTS

The pen of pigs from which the images were taken contained 12 pigs. The camera was programmed to take photos at 5-minute intervals, and no images were missed. From the images an ethogram focusing on posture changes, was devised. The behaviors scored were: Standing, Dog Sitting, Mounting (other pigs) and Lying Down. During analysis of the images in addition to the behaviors of all the pigs, a note was also made of the number of pigs visible. The proportion of images where only one pig was visible was low (4%). The proportion of pigs in view demonstrating each type of behaviour was then calculated (Figure 3). The results showed that sampling every 15-minutes gave similar results, as 5-minute sampling.

CONCLUSION

The digital stills camera will allow for the first time, data collection on the behavior of pigs during commercial long distance journeys.

- 1. EC Council Regulation (EC) No 1/2005 on the protection of animals during transport and related operations and amending Directives 64/432/EEC and 93/119/EC and regulation (EC) No 1255/97.
- The welfare of animals during transport (details for horses, pigs, sheep and cattle). Report of the Scientific Committee on Animal Health and Animal Welfare, European Commission, Health and Consumer Protections Directorate-General, Brussels 11th March 2002.

An Automated Tracking Algorithm for Quantitative Group Behavior Studies

Haishan Wu

School of Computer Science, Fudan University Zhangheng Road. Shanghai, P. R. China hswu@fudan.edu.cn +86-159 0092 2460

ABSTRACT

In this paper, we present an automated algorithm for tracking large groups of targets. To quantitatively investigate the behaviors in group context, it is vital to measure the motion trajectories of each target. This is task is challenging because large numbers of interacting targets lead to frequent occlusions. The proposed method first automatically detects the targets from video sequences, and then tracks the detections vial solving a global optimization problem. Several parameters of the proposed method can be determined automatically. Using this approach, we successfully measured the motion trajectories of large numbers of *Mycoplasma mobile*, which may facilitate the biologist to explore their gliding behavior in group context.

Author Keywords

Group behavior, multi-target tracking.

INTRODUCTION

Behavior of organism groups, such as human crowds, insect swarms and bacterial groups, is a fascinating natural phenomenon. Quantitative investigation of such behavior may lead new insights to the underlying interaction mechanism. Tracking a group of targets by human visual inspection is obviously very laborious and tedious, and thus developing an automated algorithm for measuring individual motion trajectories will be of great value.

With one video camera, target tracking involves two steps: target detection in each frame and detection association in consecutive frames. Most current tracking approaches [1] are used to track single or several targets. When handling large numbers of targets in a group, frequent occlusions in video sequences will result in challenges in both object detection and identity association. Recently, several approaches were proposed to track different organism groups [2][3][4]. The main drawbacks of these methods are



Figure 1. Threshold selection.

that their performances were highly dependent on the parameter selections, most of which should be determined empirically.

In this paper, we present an automated tracking algorithm for quantitative group behavior studies. It is able to detect the targets with occlusions automatically by utilizing the statistical information of detected blobs. It then associates the detections across frames by solving a linear assignment problem, which is designed to handle the occlusions and solved efficiently by using Hungarian algorithm.

METHODS

Taking the video sequences as input, we first removed the background in each frame, which was estimated as the time average of the whole sequences. We then segmented the targets by a global threshold, which was chosen automatically in the following way: we first found the smallest and largest pixel values in each frame, which were denoted by p_s , and p_l , respectively; then the threshold was then chose from [p_s , p_l], and the numbers of the detected blobs were counted. As shown in Figure 1, the numbers of detected blobs decreases when the threshold is too large, and too small threshold will lead to large numbers of blobs, because many of them are tiny pieces. The numbers of detected blobs are almost unchanged if the threshold is proper. Such threshold is selected to segment the targets from backgrounds.

We then associated the detections between two successive frames. We formulated this problem as a linear assignment problem, which was solved by minimizing the assignment costs. Let C(i,j) be the cost by assigning the *i*-th detection in

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

t frame to the *j*-*th* detection in t+1 frame. We used the Euclidean distances between them to define the cost matrix *C*, then the optimal assignment matrix A was obtained by solving the following optimization problem:

$$A = \arg \min \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} C(i, j) A(i, j)$$

subject to $\sum_{i=1}^{N_i} A(i, j) = 1$, $\sum_{j=1}^{N_j} A(i, j) = 1$

A is a binary matrix with entries 1 (association) and 0 (otherwise)[4]; N_t and N_{t+1} are the detection numbers in t and t+1 th frame. Generally, $N_t \neq N_{t+1}$ because of occlusions and false positive detections et., and therefore one detection in current frame might not be assigned to one detection in the next frame. We overcame this difficulty by the following way: for any detection without assignment, its position in next frame was predicted by the velocity in current frame, which were regarded as a new detection, and the corresponding track was labeled as prediction status. If a track remained prediction status for several frames (5 to 10 in our experiment) would be terminated. Experiment results demonstrated that this method could result in more complete and accurate tracks.

RESULTS

The gliding behavior of *Mycoplasma mobile* has attracted the attention of many biologists [5][6]. With a video of 700 frames, we here applied our method to measure the motion trajectories of a large group of *Mycoplasma mobile* comprising nearly one hundred individuals. The results were shown in Figure 2 and Figure.3.

CONCLUSION

We presented an automated method for measuring the motion trajectories of organism groups. This method is general-purpose and thus applicable to other tracking tasks, such as moving locust groups [7]. One future work is developing machine learning methods to mine the acquired data.



Figure 2. Tracking large number of *Mycoplasma mobile*. Round dots denote the detected positions in current frame.



Figure 3. Twenty-four acquired trajectories with lengths longer than 300 frames are shown. Different colors represent different individuals.

- 1. Yilmaz, A., Javed, O., Shah, M., Object tracking: A survey. ACM Computing Surveys, 38, 4 (2006).
- Betke, M., Hirsh, D.E., Bagchi, A., Hristov, N.I., Makris, N.C., Kunz, T.H., Tracking Large Variable Numbers of Objects in Clutter, In *IEEE Proc. ICCV* 2007.
- Branson, K., Robie, A., Bender, J., Perona, P., Dickinson, M. High-throughput ethomics in large groups of *Drosophila*. *Nature Methods*, 6, 6 (2009), 451-457.
- Jaqaman K, Loerke D, Mettlen M, Kuwata H, Grinstein S, Schmid SL, Danuser G. Robust single-particle tracking in live-cell time-lapse sequences. *Nature Methods*, 5, 8 (2008), 695-702.
- Uenoyama, A. and Miyata, M.. Gliding ghosts of Mycoplasma mobile. PNAS, 102, 36 (2005), 12754-12758
- 6. Jaffe, J.D. and Miyata, M. and Berg, H.C., Energetics of gliding motility in *Mycoplasma mobile*. *Journal of bacteriology*, *186*, 13 (2004).
- Buhl, J., Sumpter, D. J. T., Couzin, I. D., Hale, J. J., Despland, E., Miller, E. R., & Simpson, S. J.. From disorder to order in marching locusts. *Science*, *312*, 5778 (2006), 1402-1406.

The Compensatory Tracking Task: A Pattern Recognition Based Approach for Classifying Vigilance

Corinne Schenka c.schenka@googlemail.com Thomas Schnupp t.schnupp@fh-sm.de Christian Heinze c.heinze@fh-sm.de

University of Applied Sciences Schmalkalden Blechhammer, 98574 Schmalkalden, Germany

Jarek Krajewski krajewsk@uni-wuppertal.de Martin Golz m.golz@fh-sm.de

University of Applied Sciences Schmalkalden

Blechhammer, 98574 Schmalkalden, Germany

University Wuppertal, Experimental Industrial Psychology, Gaußstraße 20, 42097 Wuppertal, Germany

ABSTRACT

In this paper we apply methods of pattern recognition on data of the Compensatory Tracking Task (CTT) in order to classify between two different vigilance states. Ten subjects attended a partial sleep deprivation study. The study design included baseline measurements and measurements of two separate nights. Adaptive signal processing was applied in time and spectral domain. Different classifiers were applied in order to find a generalized discriminant function to discriminate between two pre-defined classes, "vigilant" versus "hypovigilant". The classification performance was evaluated in terms of test set error rates. Results show that best performance was obtained utilizing spectral domain features in combination with Support-Vector Machines. Regarding the test duration results indicate that a test length of six minutes may be sufficient.

Author Keywords

Compensatory tracking task, vigilance testing, signal processing, computational intelligence, support-vector machines.

INTRODUCTION

Although sleepiness is a main cause of accidents at work and in traffic, it is not easy to measure reliably immediate consequences on different performance abilities of a subject. On the one hand sleepiness leads to decrements in attention, cognition and motor control which oftentimes happen suddenly. On the other hand this is difficult to reproduce in test situations. Signal analysis is mainly concerned to continuously measurable decrements and fails in correctly quantifying sudden events. One performance test which continuously demands visuo-motor coordination and which measures performance continuously is the Compensatory Tracking Task (CTT) [1]. The question of this paper is, if the continuously measured cursor-target distances contain information on continuous performance decrements. As a first step, we applied adaptive signal processing and pattern recognition and asked for discriminability in a two-class problem.

MATERIALS

Description of the CTT

The test was executed on a standard personal computer with a trackball as the input device. In the centre of the screen a fixed annulus was presented as the target. The cursor was shaped circularly. The goal of the CTT is to locate the cursor such that the distance of the centre of the cursor to the centre of the target is zero. Only in this case a solid circle is displayed. During the duration of the test the target-cursor-distance is measured continuously at a rate of 12 Hz. The cursor is driven by three virtual forces [1]. A buffeting force calculated as a superposition of six sine functions with randomly initialized phase angles acts with limited dynamics and is still not predictable by the user. A second force acts radially like a Coulomb distraction force. The user interactions generate the third component. Even alert and trained subjects don't succeed in trying to keep the performance measure at zero.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

Subjects and Study Design

Students of the University of Applied Sciences Schmalkalden were recruited for this study. They had to fill-out three questionnaires (PSQI [2], D-MEQ [3] and SSQ [4]). Several criteria for inclusion or exclusion were checked based upon these questionnaires. Ten subjects were selected randomly. Their age ranged between 18 and 32 vears (mean 24.6 \pm 3.7). They were invited to two training sessions. During the two days before experiments all subjects had to wear a wrist actometer to assess main biorhythmical variables and to check adherence to the given sleep-wake schedule. Every subject participated in two experimental nights starting at 8:00PM and finishing at 4:00AM, each night was divided into 8 hourly sessions. Within one session 6 different kinds of tests had to be performed; their order was randomized. This contribution focuses on presentation and discussion of CTT results.

METHODS

Pre-Processing

The x- and y-components of the cursor-target distance (r_t) and of the current cursor position (x_t, y_t) were analyzed. Adaptive segmentation was performed. Their parameters were optimized empirically [5]. Samples of baseline experiments where time since sleep (TSS) was lower than five hours were labeled to class # 1 ("vigilant") whereas samples of the late night (TSS > 18 hours) were labeled to class # 2 ("hypovigilant").

Feature Extraction

Signal Analysis of recorded performance time series were performed in time and spectral domain. In the time domain 29 features were extracted of each segment of the 3 time series (x_t , y_t , r_t). They are used by other authors as well [1]. Power spectral densities are usually utilized as features in time series analysis. We estimated them by WOSA (Weighted Overlapped Segment Averaging) in order to get low variances at the cost of bias and of spectral resolution. At the same side of "costs" are the consequences of band averaging which further reduces estimation variance. The three parameters (lower / upper cut-off frequency and bandwidth) were optimized empirically [6] and resulted in 0.12 Hz, 3.16 Hz and 0.76 Hz, respectively.

Classification

Three different algorithms were compared: Learning Vector Quantization (LVQ), k-Nearest-Neighbor (kNN) and Support-Vector-Machine (SVM). LVQ is an Artificial Neural Network which training stage is relatively fast. Thus it is appropriate to empirical parameter optimizations. The main LVQ parameter to be optimized is the number of neurons [6]. kNN is a non-parametric method of statistical pattern recognition, well-known for decades. The algorithm is very simple, but is not able to perform adaptation and has relatively high computational costs. The parameter k is to be optimized empirically in order to regularize the piecewise linear separation function of kNN. SVM is a



Figure 1. Mean classification errors versus test duration. Features extracted in time domain and spectral domain had been used as input to discriminant analysis utilizing LVQ. Error bars indicate ± 1 standard deviation.

class of high-performance classification analysis. It is able to regularize between empirical classification error minimization and structural risk optimization. SVM solves the classification task through implicit transformation to high-dimensional space. It belongs to nonlinear discriminant analysis and is an important method in Computational Intelligence. SVM classification performances were validated using leave-one-out scheme. Validation of LVQ and kNN utilized 50-fold delete-d cross validations with test-training ratios of 80:20.

RESULTS & DISCUSSION

Frequency-domain feature sets led to lower classification errors compared to time-domain feature sets. Frequencydomain features showed a minimum in classification error rates at simulated test durations of 6 minutes which is in contrast to time-domain features. Here the minimum was observed at test durations of 10 minutes (Figure 1). Best results are observed using SVM, worst using kNN.

This study gives further evidence that the CTT is able to measure continuous decrements due to operator fatigue. In the future it has to be shown if these results keep stable when sample size is increased. Intra- and inter-subject variability has also to be estimated. Our results indicated that test durations can be reduced to 6 or 4 minutes in future. Further research should also answer the question, if there are single features of the signals which indicate performance decrements.

REFERENCES

1. Makeig, S. and Jolley, M. COMPTRACK: A compensatory tracking task for monitoring alertness. Technical Document 96-3C. San Diego: Naval Health Research Center, 1996.

- 2. Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R. and Kupfer, D.J. The Pittsburgh Sleep Quality Index *Psychiatry Res.*, 28 (1989), 193-213.
- 3. Horne, J.A. and Östberg, O. A questionnaire to determine morningness-eveningness in human circadian rhythms. *Int. J. Chronobiol.*, *4* (1976), 97-110.
- 4. Kennedy, R.S., Lane, N.E., Berbaum, K.S. and Lilienthal, M.G. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol. 3*, 3 (1993), 203-220.
- 5. Golz, M. and Sommer, D. Automatic knowledge extraction: fusion of human expert ratings and biosignal features for fatigue monitoring applications. In: Mandic DP, Golz M, Kuh A et al. (Eds.) *Signal Processing techniques for knowledge extraction and information fusion*. Springer, New York, (2008), 299-316.
- 6. Sommer, D. Charakterisierung des Mikroschlafs mit neuroinformatischen Methoden. PhD Thesis (in German). Shaker Press, Aachen, Germany, 2009.

Automated Home Cage Assessment of Sca17 Mouse Model

Esteban Portal

University of Tübingen, Department of Medical Genetics Calwerstraße 7, 72076 Tübingen, Germany estebanportal@gmail.com

ABSTRACT

The study of neurodegenerative diseases presents a great challenge in modeling and treatment, requiring that behavior be carefully recorded and analyzed. There is an increase in the use of rodent models used for this purpose, but behavioral tests pose the challenges of being time and resource intensive, as well as demanding specific training and careful planning to ensure relevant results. The use of automated home cages can improve these aspects of behavioral testing. For this reason, we have assessed a SCA17 mouse model with the LabMaster automated home cage system. We report the results for this model as well as the methodology used to improve analysis by automated home cage system.

Author Keywords

Automated home cage, automated behavior assessment, rodent models, SCA17, neurodegenerative disease.

Spinocerebellar ataxia 17 (SCA17) is progressive neurodegenerative disease of autosomal dominant transmission. SCA17 is induced by expansion of the CAG repeats (>42) in the TATA-binding protein (TBP), a transcriptional initiation factor. This leads to atrophy in the cerebellum and Purkinje cell loss, with less pronounced neurodegeneration in other parts of the brain (1,2).The clinical phenotype is heterogeneous, exhibiting ataxia, dementia, dystonia, parkinsonism and dystonia (1,2). Mouse models for SCA17 have been generated and studied using Rotarod, weight and survival. We chose one of these models, SCA17Q-16J, to assess in an automated homecare system with the aim of finding whether we could explore more phenotype parameters in the disease progression.

Large scale behavioral assays of animal models using classical behavioral analysis face limitations due to manpower and training required, time consuming techniques, and costs of breeding and maintaining large amounts of animals. Furthermore, experimenter effect and differing protocols for the number of tests used can give differences when comparing studies. When the same animals are tested several times in a battery of tests, problems of carry-over effects and stress induced by manipulations increase with the number of tests used. On the other hand, when different animals are assigned to different tests, a large number of animals is needed thus increasing the costs. A limitation with using several tests to measure activity is that only one aspect is recorded with each trial, with no concurrent or complex behavior analyzed. Animal models with severe phenotypes pose additional challenge due to the reduced life expectancy and time available for experimentation.

The automated home-cage system LabMaster (developed by TSE system, Bad Homburg, Germany) provides an excellent tool to measure a wide range of behavioral parameters (5). While such automated approach results in concurrent and complex behavioral data, the amount of data produced requires much time to explore and process. A second objective was to find a methodology to reduce the time necessary to explore and interpret the output generated by this system. To process the data, we used the R statistics program for automatic aggregation, grouping, correlation and multivariate analysis that permits a quick screen of behavioral characteristics. The results where also processed with GraphPad Prism for final presentation.

CONCLUSION

We found that the Labmaster automated home cage systems permits a more deep analysis of behavior phenotype for SCA17Q-16J mouse with less time and training requirements than the classical behavior tests published. Also, we were able to produce a methodology and tools that greatly improves the efficiency of analyzing and exploring automated home cage results.

- 1. Koide, R. A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Hum. Mol. Genet.* 8, (1999), 2047–2053.
- 2. Nakamura, K. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

in TATA-binding protein. Hum. Mol. Genet. 10, (2001), 1441-1448.

- 3. Friedman MJ. Polyglutamine domain modulates the TBP-TFIIB interaction: implications for its normal function and neurodegeneration. *Nat Neurosci.* 10, 12 (2007), 1519-1528.
- 4. Friedman MJ. Polyglutamine expansion reduces the association of TATA-binding protein with DNA and induces DNA binding-independent neurotoxicity. *J Biol Chem.* 283, 13 (2008), 8283-8290.
- 5. Urbach, Y.K. *et al.* Automated phenotyping using intrahome-cage technology. *Proceedings of Measuring Behavior 2008*, 330.

The BacHD Rat: A New Rat Model of Huntington's Disease Expressing the Full-Length Mutant Huntingtin

Laura Emily Clemens

Laura.Clemens@med.uni-tuebingen.de

Libo Yu Libo.Yu@med.uni-tuebingen.de

Jasmin Simone Ehrismann

Jasmin.Ehrismann@med.uni-tuebingen.de

Esteban Portal

Esteban.Portal@med.uni-tuebingen.de

Olaf Riess

olaf.riess@med.uni-tuebingen.de

Huu Phuc Nguyen

hoa.nguyen@med.uni-tuebingen.de

University of Tuebingen, Department of Medical Genetics Calwerstraße 7, 72076 Tübingen, Germany

ABSTRACT

The quality of behavioral studies on neurodegenerative diseases relies on the employment of adequate animal models. We recently developed a unique rat model of Huntington's disease, which carries the full-length form of the causative gene huntingtin, inserted via a bacterial artificial chromosome (Bac). Advantages of the BacHD rat model are the replication of the protein context of the human disease due to the expression of the protein in its native form, combined with species-related benefits of the rat as an animal model. To characterize the BacHD rat, we implement classical behavioral tests as well as an automated home-cage system. First behavioral studies in BacHD rats demonstrate an early phenotype that resembles the human pathology. Thus the BacHD rat represents an appropriate new model to carry out basic research as well as therapeutic studies for Huntington's Disease.

Author Keywords

Huntington's disease, full-length huntingtin, rat model, behavioral screening, PhenoMaster.

Ethical Statement

All of our animal experiments have been approved by an ethical committee and have been carried out in accordance with the German Animal Welfare Act (Deutsches Tierschutzgesetz).

INTRODUCTION

Huntington's Disease (HD) is a fatal hereditary neurodegenerative disorder caused by an expanded CAGrepeat in the first Exon of the gene coding for the protein huntingtin. The patients suffer from chorea, psychiatric and metabolic disturbances as well as cognitive decline, leading to their death 10-15 years symptoms onset. Up to the present, there is no specific therapeutic for the disease available and the molecular pathogenesis still needs to be deciphered. For this reason, it is fundamental to have an animal model which replicates properly the human condition and therefore allows for more relevant results in basic research as well as in therapeutic studies.

THE BACHD RAT

We have developed a rat model of Huntington's Disease by inserting the full-length mutant huntingtin into Sprague Dawley rats' genome using the bacterial artificial chromosome (Bac) as a carrier. This BacHD rat model expresses human full-length huntingtin under the control of the human huntingtin promoter and its regulatory elements. Additional 20 kb upstream and 50 kb downstream sequences reduce the position effect of the transgene. The expansion in the huntingtin gene is a mixed CAA-CAG repeat, which produces a particular stability of the repeat length. Two LoxP sites, flanking the mutant huntingtin exon 1, allow for Cre-mediated excision.

The advantages of the BacHD rat model compared to other models of the disease are twofold, based on the speciesspecific differences between the animal models and the genetic background. The larger size of rats relative to mice facilitates repetitive physiological measurements and surgical manipulations as well as structural and functional imaging. Furthermore, cognitive abilities and their behavioral correlates are much better comparable between humans and rats. Due to the presence of the full-length huntingtin in its native 3-D structure (and unlike the situation in the fragment models of HD), cleavage and degradation processes as well as interactions with other proteins, which are instrumental to the disease pathogenesis in humans, are replicated.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

MEASURING BEHAVIOR IN THE BACHD RAT

In order to characterize the BacHD rat, we have assessed behavior, metabolism and motor function with classical behavioral test approaches as well the automated homecage system PhenoMaster (TSE Systems, Bad Homburg, Germany). With the age of 1 month, the BacHD rats already display impaired motor function demonstrated by less grip strength and a shorter latency to fall from a rotating rod. With 3 months, BacHD rats have reduced locomotor activity and food and water intake, measured in the PhenoMaster. Six months old animals show decreased exploratory behavior as well as decreased anxiety concluded from less locomotor activity and a higher ratio of locomotor activity in the centre versus the periphery during the first 15 min of a PhenoMaster recording. 9 months old BacHD rats have lower oxygen consumption rates and a constantly low respiratory quotient of 0.7, calculated from carbondioxide production and oxygen consumption sampled in the PhenoMaster. Furthermore, the animals display a higher amount of activity during the light cycle and reduced activity during the dark phase pointing to an incipient disruption of the circadian rhythm.

CONCLUSION

The BacHD rat is in many ways advantageous to other animal models of HD. It displays an early behavioral, metabolic and motor phenotype that resembles many of the human pathologies. Thus, the BacHD rat represents an appropriate new model to carry out basic research as well as therapeutic studies for Huntington's Disease.

Crowd Safety Architecture - Measuring Safety Levels

Maarten Douwe Bredero

Office Architect Maarten Douwe Bredero and lecturer at Saxion University of Applied Sciences Kortestraat 27, 7419 CK Deventer, The Netherlands mail@architectmaartendouwebredero.com

Author Keywords

architecture, crowd control, crowd safety, safety, orientation, perception, space.

INTRODUCTION

Guiding football fans or supporters to their favourite pastime has become an increasingly delicate matter. The cost to the public purse of police supervision during football matches has become extremely high. Consider the recent case when two matches were played instead of one so as to exclude visiting supporters (Ajax versus Feyenoord).

A specific location (FC Twente's "Grolsch Veste" stadium) allows research to be carried out on how to optimize the process of guiding supporters from the train station to the stadium [1].

The hypothesis states that smart guidance compensates for the sense of danger. Smart guidance here focuses on either separating or combining flows of supporters in relation to other pedestrian and cyclist traffic.

Both the approaches dealt with in this research proposal are to be implemented in a real-life situation and in an architectural design.

RESEARCH QUESTIONS

The questions to be considered are:

- To what extent can a virtual model correlate with an actual physical site?
- Is visiting the site prior to engaging in the virtual environment mandatory for deriving valid conclusions?
- To what extent does mingling the general public with supporters diminish feelings of comfort and safety?
- Does optimizing the situation require complete separation of supporters, the general public, and cyclists?

- Does physical separation by means of transparent barriers in fact create a feeling of caging, which will be counterproductive to establishing a safe environment?[2]
- Can the situation be improved by constructing a "supporter-proof" cycle bridge (in the form of a folding cage), combined with cycle parking facilities?

TYPE OF RESEARCH

The research involved is empirical with an inductive approach. Based on specific observation of behaviour, an attempt will be made, on the one hand, to deduce general rules. On the other hand, it is assumed, as a deductive hypothesis, that smart separation of supporters from the general public will enhance the feeling of safety.

The research will consist of questionnaires administered at specific intervals while subjects explore the virtual environment. A survey will therefore be combined with a case study in order to test the hypothesis.

Depending on the number of measurements, the research will primarily be qualitative in nature. Only limited statistical calculations will therefore be necessary.

Recording of behaviour during the simulated process of passing through and exploring the real architectural space will provide an insight into the psychological perception of spatial information.

The objective is naturally to achieve a neutral result by applying an appropriate methodology and precise focus.

RESARCH METHODOLOGY

Precise measurement can be achieved by working with four distinct groups.

The subjects will consist of:

- 1) the general public attending the sports event;
- 2) cyclists passing through the area;
- 3) supporters;
- 4) police.

Test 1 REAL ENVIRONMENT: Within a given timeframe, subjects within a group explore the actual site without a cycle bridge and respond to specific questions put to them. Set-ups A and B.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.





TEST 1: visiting public and cyclists mix, fans mixed or separated, A or B



TEST 2: visiting public and through cyclists separated, fans mixed or separated, C or D









Figure 1. Test configurations and bridge design projected on site (length 432 m).

Test 2 VIRTUAL ENVIRONMENT: Within a given timeframe, subjects within a group explore the virtual model with a cycle bridge and respond to specific questions put to them. Set-ups C and D.

All subjects first work alone and then together during both tests. It may be necessary to combine different types of subject but this will depend on the results of the initial tests. Consider mixing fans with the police as depicted in Figure 1.

Moreover, relevant questions will be put regarding orientation - i.e. the mental condition of subjects before and after the tests - with the answers being compared with the actual behaviour of the subjects.[3]

IMPLEMENTATION OF RESEARCH

The actual site will be transformed into corridors by means of special fences as outlined below.

The present architectural Definite Design (see below) will be further converted into a semi-realistic environment using advanced visualization and gaming techniques.

Subjects, either as crowd members, cyclists or lawenforcement officials, will explore the environment interior – either individually or in groups – by means of computer interfaces. They will interact with computer screens while sitting on chairs at gaming consoles. This can be done at the advanced gaming facilities at the University of Twente (T-Xchange), which is directly adjacent to the study location. This non-direct method (registration of observations) will make it possible to measure the mental process of safety assessment [4].

GENERAL CONSTRAINTS

Accurately zoned site;

Realistic model (Architectural Definite Design converted into real-life visualization);

Close correlation between physical site and design model;

Representative measurement (large enough number and proper kind of subjects);

Applicable (usable protocol for professional architects as result).

- Seer, S., Bauer, D., Brandle, N., Ray, M., Estimating Pedestrian Movement Characteristics for Crowd Control at Public Transportation Facilities. International IEEE Conference on Intelligent Transportation Systems, 2008.
- 2. Perkins, L.B., Crowd safety and Survival. Lulu, Morrisville NC, USA, 2004.
- 3. Jager, W., Popping, R., Sande, van de, H., Clustering and Fighting in Two-party crowds: simulating the Approach-avoidance Conflict. *Journal of Artificial Societies and Social Simulation*, 4, 3 (2001).
- 4. Jones, I., Mixing Qualitative and Quantitative Methods in Sport Fan Research. *The Qualitative Report*, *3*, 4 (1997).

Free Exploration of an Elevated Plus-Maze in Mice

Arnaud Arabo

Claire Potier

Vincent Roy vincent.roy@univ-rouen.fr

Affective and Cognitive Neuroscience & Psychology Laboratory (EA4306) University of Rouen, France

ABSTRACT

In this research, we compared the behavior of mice allowed to freely explore an elevated plus-maze (EPM) with the behavior of mice exposed to the classical testing situation (i.e. forcible centre exposure). A detailed temporal analysis (min-by-min scoring) raises interesting results for practical and theoretical views on the elevated plus-maze task.

Author Keywords

Elevated plus-maze, free exposure, anxiety, emotional behavior, exploration, mice.

INTRODUCTION

The EPM is the most widely used model to evaluate rodent "anxiety-like behaviors". It is based on the a priori postulate that rodents exposed to the apparatus will respond to a conflict between safe parts of the maze that are protected (closed arms), and aversive parts of the maze that are unprotected (open arms). Several discrepancies emerge from the interpretation of an animal's behavior in the EPM. In particular, whether avoidance of open arms is really an unconditioned process, or whether it is rather something acquired during exposure to the test, is still very much an open question [1]. Indeed, rodent's aversion for open space is supposed to be natural; but the fact that open-arm entries actually decrease during exposure to the EPM argues in favor of some kind of open-arm avoidance learning. This is especially relevant with respect to some potential explanations of the one-trial tolerance effect, consisting of a high avoidance of open arms and a decreased effectiveness for anxiolytic drugs during re-exposure to the EPM.

In a recent study, a procedure of free exposure to the EPM was used to demonstrate that open arm avoidance is really a natural tendency in rats [3]. In order to generalize and extend these results, we conducted a follow-up experiment in mice using a detailed temporal analysis.

MATERIALS AND METHODS

Animals

Male CD1, BALB/c and C56BL/6 mice (70 days-old) were tested in the EPM, either in the free exploration or in the forcible exposure situation.

Procedure

The mice were isolated for 24h in a compartment (20 x 20 x 20cm) with bedding, food and water access. On the testing day, animals from the "free exposure group" were given free access to the EPM thanks to a small removable door connected to the end of a closed arm (see Figure 1). Mice were thus allowed to freely explore the EPM with the possibility to return to the familiar compartment. Their behavior was recorded for a total of 5 minutes in the EPM. Control animals were handled from their familiar compartment and forcibly exposed to the EPM for 5 minutes. Control animals were placed at the centre of the maze, facing an open arm.

Maze and Video-Tracking

The elevated plus-maze (Intellibio, France) was made of ivory Perspex with a central zone area (5 x 5 cm) and facing closed and open arms (25 x 5 cm). It was elevated at a height of 50 cm, the closed arms were surrounded by 25 cm



Figure 1. A mouse freely explores the EPM from a familiar compartment connected to a closed arm.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

walls and the open arms had 0.25 cm edges. The testing room was illuminated with a dim white light that provided 100 lux for the open arms and 50 lux for the enclosed arms.

The ANY-maze[®] video-tracking system (Stoelting, version 4.60) was used to record the distances covered, the arm entries and the time spent in the open arms min-by-min. Entering an arm was counted when 98% of the animal's tracking area overlaid the arm whereas the animal was considered to be in the centre of the maze when 50% of the tracking area was out of the arm. Closed-arm returns were recorded when an animal exited a closed arm and re-entered the same arm. The ethological keyboard was used to record rears, head-scans and protected stretched attend postures (pSAPs).

Ethical Statement

The research was conducted in accordance with the guidelines for the care and use of laboratory animals established by the National Institute of Health of the United States of America [2].

Statistical Analysis

Data were analyzed with ANY-maze[®]. Two-way ANOVAs with condition and segment of test as factors were performed, followed by Tukey's tests for post-hoc comparisons when allowed.

RESULTS AND DISCUSSION

The results that will be presented in details at the meeting confirm that open-arms avoidance is more pronounced in the free exploration condition than in the forcible exposure situation. This difference is observed for the three mice strains and does not depend on locomotion (total distance covered) or exploration (head-scans) in the maze. The minby-min analysis points out that this difference is evidenced during the first two minutes of the test (see Figure 2 which presents for instance the open-arm entries in CD1 mice).

First, these results confirm that open arm avoidance is also a natural tendency in mice and that it does not require some kind of learning during the first trial in the EPM. In addition, the present study also reinforces the hypothesis that initial open-arm entries during a forcible confrontation with the apparatus should possibly be interpreted as the attempts to escape the situation, rather than as the indications of a low anxiety level. This hypothesis is also consistent with the fact that pSAPs and closed-arm returns were more frequent in the free exploration condition. Indeed, the approach / avoidance conflict is more clearly



Figure 2. Open-arm entries in CD1 mice forcibly exposed or freely exposed to the elevated-plus maze.

defined in this situation since the familiar compartment stands really in contrast with the unknown elevated plusmaze and its anxiogenic open-arms.

CONCLUSION

Our work confirms that during initial exposure to the EPM, the rodents' behavior is not driven only by the approach / avoidance conflict towards the open arms, but also by some attempts to avoid the whole situation. These attempts can lead during the first minutes of the test to open-arm entries. Thus our work emphasizes the relevance of a detailed temporal analysis to avoid misinterpretations of the results in the EPM procedure.

- Carobrez, A. P., & Bertoglio, L. J. Ethological and temporal analyses of anxiety-like behavior: the elevated plus-maze model 20 years on. *Neurosci. Biobehav. Rev.*, 29, 8 (2005), 1193-1205.
- Institute of laboratory animal resources Commission on life sciences - National research council. *Guide for the care and use of laboratory animals (in french)*. National Academy Press, Washington, DC, USA, 1996.
- 3. Roy, V., Chapillon, P., Jeljeli, M., Caston, J., & Belzung, C. Free versus forced exposure to an elevated plus-maze: evidence for new behavioral interpretations during test and retest. *Psychopharmacology (Berl), 203,* 1 (2009), 131-141.

Work Units in Tutorials Software for Learning Mathematics

Laurent Souchard

Ministère de l'alimentation, de l'agriculture et de la pêche, France LDAR, University of Paris Diderot, France CREAS, University of Sherbrooke, Quebec, Canada laurent.souchard@dbmail.com

ABSTRACT

In this paper, we describe one of the tools we use to construct the comparison of four tutorials software dedicated to the learning of mathematics in early secondary. One of the biggest differences between the traditional school and the use of computer for learning is the number of interaction that the student obtains with software. After this report, it is necessary to measure these interactions. The tool we have built and we present here, is the Work Unit, which is the interval of time in the course of which the learner receives no answer of the software. It helps us reach even more precisely in the description of CTS of our study as an institution of teaching and learning.

Author Keywords

Moments of working, Work Unit, Work Unit Factor, learning time, tutorial software.

INTRODUCTION

The offer of tutorial software [1] for learning mathematics and many other disciplines is becoming increasingly important.

The impact of a technological environment on the organization of education and classroom has been known for many years [2]. The research we conduct is to describe the characteristics of an environment using Closed Tutorial Software for learning mathematics at the beginning of secondary education in France [3]. These characteristics fall within the description of CTS as an institution of teaching and learning.

Before a didactic analysis we have constructed in our research, we chose to describe one of the specificities of the CTS: the temporary management of the moments of the study [4] which proposes every software. For it we set up, from the software *The Observer 5* [5], the following protocol:

- moments of study and research (AER);
- moments of institutionalization;
- moments of the technical work (M5);
- moments of evaluation (M6);
- moments of manipulation software (Manip).

The moments most often encountered in the four software by pupils and by the expert, are the moments of the technical work (M5, continuous moment) and evaluation moments (M6, instantaneous moment).

Among the different areas of comparison of four software, math education, algebra or calculus, educational organization of the software, student assessment, we decided to take into account the pace of each software through the repetition of evaluation moments.

Encoding scheme that we chose, we can see significant differences between software and the chapters of each software. It's possible to see these differences in the beginning of the *Visualize Data* [4] of the expertise of the four CTS.

So, we need to create a new tool to describe these differences, the Work Units [3].



Figure 1. Pupils working with Closed Tutorial Software.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

| Legend | | | | |
|----------|---------------|--|--|--|
| Behavior | Color/Pattern | | | |
| AER | | | | |
| M5 | | | | |
| мө | | | | |
| Manip | | | | |

Figure 2. Colors for the moments.

WORK UNIT

A work unit is the interval of time between two consecutive time assessment of student work or expert. For example, if t_0 , t_1 and t_2 are the first three moments of assessment in a series of exercises, W1 and W2 are the first two Work Units:

$$WU1 = [t_0, t_1]; WU2 = [t_1, t_2].$$

COMPARISON OF WORK UNITS IN FOUR SOFTWARE

The first remark concerns all Work Units (WU) of the expert. We have listed about 1800 that the expert must make all four Tutorial Software. Of the results, the average duration of WU is approximately 24.7 seconds. The average of tutorial software not very different and is 23.75 seconds (3%). This shows that the deviations of the averages by the average total Tutorial Software are not unimportant. We noted in our analysis that the serious game in the software *Smao* is singular from the rest of the software.

By the calculations of average years by WU and we find the



Figure 3. Visualize Data of the expertise of the four CTS in The Observer 5.



Figure 4. The first Works Unit.

average values of 20.5 and 20.8 seconds. We keep in mind that the average value of Work Units of the expert for all four Tutorial Software is around 20 seconds off the serious game in *Smao* part, that is to say for 1700 UT expert. The serious games in *Smao* are not thus comparable to the other types of exercises in four CTS and would require a more specific analysis.

The Figure 5 shows the average duration of the WU expert and provides us a ranking of the four Tutorial Software. *Tdmaths* and *Les Maths, c'est facile* are fast enough with an average of 17.3 and 19 seconds by WU on all software. *LiliMath* is slower with an average of 22.7 and *Smao* is much slower if all the software is taken into account with an average of 36 seconds by WU. If the serious games are excluded from the calculations, *Smao* is quite comparable in speed to *LiliMath* with an average of 24.4 seconds. The first calculation of average says that:

- *Tdmaths* and *Les Maths*, *c'est facile* are fast;
- *LiliMath* is moderately slow;
- *Smao* is rather slow.

The second set of remarks on the division of Work Units of the expert into three categories: very short WU under 10



Figure 5. The duration of Work Units by software in seconds.



Figure 6. Median duration of WU of the expert by exercises.

seconds, those who are average and whose duration is between 10 and 40 seconds, and finally, WU's rather long that lasts more than 40 seconds.

Going further in the statistical analysis, we can notice in the Figure 6, some differences with three additional software and *LiliMath*.

WORK UNIT FACTOR

Work Units allow us a first description of the software on their use by learners. A second possible use of this concept is one that compares the student work with that of the expert. For this, we define the Work Unit Factor as the ratio: duration of the Work Unit of student on the duration of the Work Unit of the expert.

Work Unit Factor =
$$WUF = \frac{Duration WU_{student}}{Duration WU_{expert}}$$

Figure 7. Work Unit Factor.

For example, here are the WUF four students using the software *Smao*. Two of the students seem to work faster than the expert. Going to check the actual work of these two students, we find that they have actually met many difficulties to solve their exercises. This timeliness can translate a shoddy job.

| Camille | Jules | François | William |
|---------|-------|----------|---------|
| 2,10 | 1,47 | 0,75 | 0,74 |

Table 1. Work Unit Factor in Smao.

The time management is the heart of the successful education of all students: a good student is one who often knows how to manage his time during the homework, quizzes and other competitions or examinations in a limited time, as we remember Christine Pierrisnard-Robert: Note that the best performing pupils academically, who have specific knowledge and meta command of their ''role as students, tend to perform the tasks required quickly, which joined the assumptions of teachers who generally believe that'' good student ''works quickly and well. [6]

Nothing escapes the clock and the first sentence which hears a professor on behalf of his pupils in failure is: «I did not have time to finish ". But the tools available to enable students to learn to manage their time are not numerous. Time management is actually one of key to success, particularly in mathematics.

The research has made Christine Pierrisnard Robert confirmed.

It is particularly difficult and rare to explain what we know and what we make of "time" of the school activity. [7]

The concept of Work Unit, or WU, in Tutorial Software provides us a tool to measure and a different approach to this "time" of the school activity. The time when, to be exact, the different types of times are not the same as in conventional institution: teaching time, decided and managed by the teacher and the learning time, experienced by the student are not exactly transferable to the new institutional environment where the student works alone with a Tutorial Software. The teaching time is integrated and programmed into the Tutorial Software by the creators of the software but the student is also an actor in the management of this didactic time. He may decide to recover control, for instance, by repeating as many times as desired exercise. This suggests that the learning time, instructional time, time of activity, all these times and their acceptance by students and teacher need multiple institutions to express themselves more effectively.

Work Unit Factor is ultimately for us a tool that allows us to focus on parts of the student work through this indicator on working time. Based on a repository of expertise achieved by the Tutorial Software, we can define an optimal duration of the work unit by year and Tutorial Software. Once the factor of a WU student deviates too far from the average value, an analysis is necessary. This does not mean of course an average value close to the standard value is sufficient proof. From these averages, we could imagine a system that assigns each student a series of WUF type exercise gradually as the student advances in the Tutorial Software.

CONCLUSION

The concept of Work Unit has allowed us to enter a new dimension of analysis of student work. The method of behavior analysis and computer tools allowed us to use this approach. The time of activity can be approached in the second and the student has to have access there to manage the evolution of his learning. As well as a sportsman who trains for a competition, a pupil must be able to measure his progress and have access to tools which allow such an approach.

- 1. Bitter G., Pierson M., Using technology in the classroom, Allyn and Bacon, 2007.
- Haymore J., Sandholtz C., Owyer D. C., La classe branchée, collection de l'Ingénierie éducative, CNDP, 1998; titre original : Teaching with Technology: Creating Student-Centered Classrooms, Teachers College Press, 1997.

- Souchard, L., Les Logiciels Tuteurs Fermés: institutions d'apprentissage et d'enseignement des mathématiques? le cas du début du secondaire. Université de Paris Diderot, France, 2009. http://tel.archives-ouvertes.fr/tel-00421836/fr/
- 4. Chevallard, Yves, *Organiser l'étude :1. Étude et fonctions*, Actes de la 11e École d'Été de Didactique des Mathématiques, Grenoble, La Pensée Sauvage, 2002.
- 5. The Observer from Noldus Information Technology: http://www.noldus.com/office/fr/observer-xt-2 /.
- 6. Pierrisnard-Robert, C., Conceptualisation des aspects temporels en situation d'apprentissage scolaire. *Temporalistes*, 40 (1999), 9.
- 7. Ibid, p. 8.

Workshops

Proceedings of Measuring Behavior 2010 (Eindhoven, The Netherlands, August 24-27, 2010) Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

Measuring (Ambient) Persuasive Technologies

Wolfgang Reitberger,

wolfgang.reitberger@sbg.ac.at

Alexander Meschtscherjakov alexander.meschtscherjakov@sbg.ac.at

Manfred Tscheligi

manfred.tscheligi@sbg.ac.at

Christian Doppler Laboratory for "Contextual Interfaces" HCI & Usability Unit, ICT&S Center, University of Salzburg, Salzburg, Austria

Boris de Ruyter

Philips Research Europe HTC34, AE 5656 Eindhoven, The Netherlands boris.de.ruyter@philips.com Jaap Ham

Department of Human-Technology Interaction, Eindhoven University of Technology PO Box 513, 5600 MB Eindhoven, The Netherlands j.r.c.ham@tue.nl

ABSTRACT

In this paper we describe a half day workshop about the measuring of (Ambient) Persuasive Technologies at Measuring Behavior Conference 2010. The workshop aims to bring together some of the leading experts in the field in order to foster an interdisciplinary dialog regarding challenges and possibilities of this novel research area.

Author Keywords

Ambient Intelligence, Ambient Persuasion, persuasion, persuasive technologies, measurement, evaluation.

ACM Classification Keywords

H.5.m Information interfaces and presentation: Miscellaneous, H.1.2 User/Machine Systems: Human factors

INTRODUCTION

Recent research has started investigating Persuasive Technology, which can be defined as technologies that are "intentionally designed to change a person's attitude or behavior or both" [2]; see also [1]. For example, persuasive picture frames can be used to improve sitting posture [5], dashboard instruments can give car drivers feedback about their fuel consumption, and robots that take on the role of social actors and praise or criticize users depending on their performance (e.g., [3]).

The current workshop will assess the measurement of

Persuasive Technologies. That is, crucial in the effective development and employment of these technologies is the measurement of persuasive effect in general, long term effects in particular, underlying mechanisms and processes and appropriate methods to measure such effects. Measuring Persuasive Technology inherently demands a multidisciplinary approach: Amongst these are perspectives related to human-computer interaction, psychology, sociology, and ethics/ morality.

Next to stimulating conscious behavior changes, (ambient) persuasive technology [8] can also influence behavior in unconscious ways [4]. The current workshop will not only spend attention to variables related to conscious behavior change (e.g., explicit attitude change), but also to unconscious determinants and effects (e.g., implicit attitude change).

The workshop builds on previous efforts by the authors, such as a workshop at CHI 2008 [6], which aimed to establish and further develop the concept of Ambient Persuasion and more recently a workshop at AmI-09 [7], which focused on applications in this novel area and the potential ethical dilemmas.

CHALLENGES

- *Measuring the persuasive effect of persuasive technology* on different levels, e.g. attitude change, awareness, behavioral change. This challenge raises the question which kind of factors and outcomes are suitable to determine the success of a persuasive intervention and how these factors can be evaluated.
- *Methods for measuring long term persuasive effects* (e.g. the Experience Sampling Methodology –ESM).
- Assessing the persuasive potential of early prototypes is motivated by the need to identify promising persuasive

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

designs already at an early stage in the user centered design process. Existing approaches such as the paratype method combine low fidelity prototyping and experience sampling to help achieving this goal, but further research and methodological innovations are still required to improve persuasive prototyping.

- *Modalities of Persuasive Technology*, e.g. Ambient and Ubiquitous Technologies, Mobile Devices, Robots for Persuasion, Embodied Conversational Agents, etc.
- Specific Contexts for Persuasive Technology such as:
 - Context car (e.g. fuel efficient driving behavior)
 - o Context home (e.g. energy consumption)
 - o Context factory (e.g. motivation and productivity)
 - Context shopping (e.g. advertising or specific consumer behaviors)
 - o Context health (e.g. changes in lifestyle behavior)
- *Moral dilemmas of using persuasive technology* (e.g. when social norms and value conflicts arise) and ethical concerns regarding persuasive technology.
- *Persuasion strategies* (e.g. obtrusive vs. unobtrusive, subliminal persuasion)

SPEAKERS

The workshop is going to be half day. A specific group of internationally renowned researchers in the field will be invited to present their work through a selective call for contributions. Each talk will take 20 minutes including discussion and we aim for a total of 6 talks, with sufficient time for discussion among the partcipants.

The topics and their respective invited researchers are:

Assessing the Social Psychological Consequences in Behavior and Cognition of Persuasive Technology" Cees Midden

Measuring Feelings - What Can Persuasion Learn from User Experience Virpi Roto

Prerequisites for Successful Measuring of Ambient Persuasive Technology Harri Oinas-Kukkonen

Trust, Discourse Ethics, and Persuasive Technology

Andreas Spahn and Philip Nickel

Approaches for Measuring Behavior Change in Energy Consumption Interventions

Magnus Bång

Audience Measurement for Digital Signage Jörg Müller

- 1. Fogg, B. J. (2003). Persuasive Technology. Using Computers to Change What We Think and Do. San Franciso: Morgan Kaufman Publishers.
- IJsselsteijn, W., de Kort, Y., Midden, C. J. H., Eggen, B., and Van de Hoven, E. (2006). Persuasive Technology for Human Well-Being: Setting the Scene. Proceedings of Persuasive 2006, 1-5.
- Ham, J., Midden, C., & Tak, S. (2008). The persuasive effects of positive and negative social feedback from an embodied agent on energy conservation behavior. Proceedings of Persuasive 2008, Oulu, Finland, June 4-6, 2008. Oulu, Finland: University of Oulu.
- Ham, J., Midden, C., & Beute, F. (2009). Can Ambient Persuasive Technology Persuade Unconsciously? Using Subliminal Feedback to Influence Energy Consumption Ratings of Household Appliances. Proceedings of Persuasive 2009, Claremont, USA.
- 5. Obermair, C., Reitberger, W., Meschtscherjakov, A., Lankes, M., and Tscheligi, M. 2008. perFrames: Persuasive Picture Frames for Proper Posture. In Proceedings of Persuasive 2008, Springer, Berlin, Heidelberg, 128-139.
- Reitberger, W., Tscheligi, M., de Ruyter, B., and Markopoulos, P. 2008. Surrounded by ambient persuasion. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 3989-3992.
- Reitberger, W., Ham, J., Weiss, A., Spahn, A., Meschtscherjakov, A., Nickel, P., Tscheligi, M., "The Ubiquitous Persuader" – Mechanisms, Applications and Ethical Dilemmas of Ambient Persuasion, Workshop, Adj. Proc. AmI2009, Salzburg.
- Reitberger W., Meschtscherjakov, A., Mirlacher, T., Scherndl, T., Huber, H., Pöhr, F., Mirnig, N., Tscheligi, M. A Persuasive Interactive Mannequin for Shop Windows. In Proc. Persuasive Technology 2009, New York, NY: ACM.

Prerequisites for Successful Measuring of Ambient Persuasive Technology

Harri Oinas-Kukkonen

University of Oulu Dept. Information Processing Science Rakentajantie 3, 90570 Oulu, Finland Harri.Oinas-Kukkonen@oulu.fi Tel: +358 8 553 1900

ABSTRACT

The area of ambient persuasive technology looks very promising in multiple ways. Yet, it also faces with challenges. Thus far the major weakness in research into ambient persuasive technology has been that the behavior change support systems have been described in such a coarse manner that it has been very difficult to demonstrate what actually caused the change, or to generalize the findings. This paper will discuss prerequisites for successful measuring of ambient persuasive technology. The O/C Matrix and the PSD Model are suggested as vehicles to better frame the research. These methodological research aids can be used in a variety of research and development settings ranging from studying full-fledged commercial systems to evaluating software specifications for systems under development.

Author Keywords

Persuasive technology, behavior change support systems, behavioral outcomes, behavioral change.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. Behavior change support systems.

INTRODUCTION

Ambient persuasive technology [20] has become an important research area and a wide variety of applications have been developed embodying it [1,3,4,5,6,17,18,19,22]. The key construct and object of study within this research area is a Behavior Change Support System [12]:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

A behavior change support system (BCSS) is an information system designed to form, alter or reinforce attitudes, behaviors or an act of complying without using deception, coercion or inducements. [12]

BCSSs are inherently transformative, deliberately attempting to cause a cognitive and/or an emotional change in the mental state of a user to transform the user's current state into another planned state. Empirical BCSS research provides a unique opportunity for quantifying measures for system success. This requires explicitly stating the aim of the system, how the success was to be measured, and the extent to which the system succeeded in achieving this measure. It has to be explicitly defined what really takes place through the software system to be able to demonstrate to what extent an outcome change is really due to the system, or a feature or a set of features in it. For this reason, sound ways of defining the systems and their goals clearly are needed. Otherwise, it will be difficult or perhaps even impossible to demonstrate any results from the BCSS or to translate lessons learned from the results into related problem and application domains. [12]

This paper will suggest a research process model for studying ambient persuasive technologies and applications. Two important steps preceding the actual measurement of behavioral change are to utilize the O/C Matrix [12] and the PSD Model [9] to better grasp the intent, structure, functionality, and content of the software system under investigation.

RESEARCH PROCESS MODEL

We suggest a five-step process model for carrying out research on behavior change support systems (Figure 1):

- 1. Select the theoretical basis for research [2,8].
- 2. Analyze the intent through the O/C Matrix [12].
- 3. Analyze the BCSS through the PSD model [9].
- 4. Measure the behavior change.
- 5. Explain the change through the theories, the O/C Matrix, and the PSD Model.



Figure 1. Experimental research model for BCSSs.

ANALYZING THE INTENT THROUGH THE O/C MATRIX

There are three types of behavioral changes, namely a change in an act of complying, a behavior change, or an attitude change, and these may be called respectively C-, B- and A-Change [12]. With a *C-Change*, the goal of the behavioral change is simply to make sure that the end-user complies with the requests of the system. The goal of systems supporting a *B-Change* is to elicit a more enduring change than simple compliance once or a few times. The goal of systems supporting an *A-Change* is to influence the end-users' attitudes rather than behavior only.

Three potential, successful voluntary outcomes are the formation, alteration, or reinforcement of attitudes, behaviors or complying; these may be respectively called as F-Outcome, A-Outcome, and R-Outcome [12].

| | C-Change | B-Change | A-Change |
|--------------------|------------------------------------|---------------------------------|----------------------------|
| F- | Forming an act | Forming a | Forming an attitude (F/A) |
| Out- | of complying | behavior | |
| come | (F/C) | (F/B) | |
| A- Out- come | Altering an act of complying (A/C) | Altering a behavior (A/B) | Altering an attitude (A/A) |
| R- | Reinforcing an | Reinforcing a | Reinforcing |
| Out- | act of complying | behavior | an attitude |
| come | (R/C) | (R/B) | (R/A) |

Table 1. O/C Matrix [12].

A matrix can be constructed from the intended outcomes and the types of change. See Table 1. When researching or developing a BCSS, it should be carefully considered which of these nine different goals the application is aimed at. The persuasion context may change dramatically when moving from one slot to another. Different persuasive goals and strategies will be needed for applications supporting different types of changes.

ANALYZING THE BCSS THROUGH THE PSD MODEL

The Persuasive Systems Design model [7,9], or more briefly the PSD, is the state of the art approach for researching and developing BCSSs. According to the PSD model, careful analysis of the persuasion context (the intent, event, and strategy of persuasion) is needed to discern opportune and/or inopportune moments for delivering the message(s). See Figure 2. The PSD model also defines software characteristics for BCSSs and describes them under four categories, namely primary task support, computer-human dialogue support, perceived system credibility, and social influence. (Many design aspects in developing BCSSs are general software design issues rather than specific to BCSSs only.) The design principles of the primary task category focus on supporting the carrying out of the user's primary activities. Design principles related to human-computer dialogue help move towards achieving the goal set for using the BCSS. The perceived system credibility design principles relate to how to design a system so that it is more believable and thereby more persuasive. The design principles in the social influence category describe how to design the system so that it motivates users by leveraging social influence.



Figure 2. PSD model [9].

Examples of using the PSD as a part of evaluating the success of persuasive technology can be found in [3,4,5,16,19]. Many types of research on software system features have been conducted. The most researched features
have been tailoring, tunneling, reduction, and selfmonitoring (representing the primary task category), suggestion (for supporting human-computer dialogue), surface credibility (in support of perceived system credibility), and social comparison, normative influence, and social learning (relating to social influence) [21].

CHALLENGES IN MEASURING THE BEHAVIOR CHANGE

Essential research issues that relate to the ultimate question, i.e. how to measure the change, include: How to measure C-, B- and A-Changes? What are the similarities and differences in measuring C-, B- and A-Changes? What are the connections between the different kinds of changes? How do the BCSSs developed for C-, B- and A-Changes differ from each other? Similar questions regarding measuring, connections and differences naturally should be tackled also regarding the F-, A- and R-Outcomes. Moreover, it should be understood how the experiments can be conducted in such a manner that it will be really possible to pinpoint a change to have been caused by a BCSS, or even more precisely, by a specific software feature in it. An additional important research issue relates to the changes that take place in user's goals or in the technological platform which the BCSS has been built upon in between starting and ending the measurement.

CONCLUSIONS AND DISCUSSION

In current research, there seems to be a tendency of describing the software systems and the persuasion context at too general a level. Black-box thinking of the software systems with no actual description of what was implemented and how may make the research results obsolete.

When describing a persuasive system, a very clear description of not only the application but also the technology context is needed. After all, in many cases much of the success or failure of an application can be attributed to the fluent navigation and smooth interaction arising from the technological infrastructure rather than to the design of the system. The O/C Matrix and the PSD Model may help in doing this more effectively.

In the future, designer bias and ethical issues will need more attention. Also the needs of underprivileged users such as the elderly and the children should be better addressed [10,11].

Even if many research efforts have already been conducted on behavior change support systems, we are still in early steps of research into measuring their outcomes and understanding how they relate to social media systems [13] and new forms of science [14,15].

REFERENCES

1. Harjumaa Marja & Oinas-Kukkonen Harri (2007) An Analysis of the Persuasiveness of Smoking Cessation Web Sites. The Second International Symposium on Medical Information and Communication Technology (ISMICT 2007), Oulu, Finland, December 11-13, 2007.

- 2. Harjumaa Marja & Oinas-Kukkonen Harri (2007) Persuasion Theories and IT Design. Lecture Notes in Computer Science, Vol. 4744, Springer-Verlag, Berlin Heidelberg, pp. 311-314.
- Harjumaa, M., Segerståhl, K., and Oinas-Kukkonen, H. (2009) Understanding Persuasive System Functionality in Practice: a Field Trial of Polar FT60. ACM International Conference Proceeding Series, Vol. 350, Proceedings of the Fourth International Conference on Persuasive Technology, Claremont, CA, USA, April 26-29, 2009.
- 4. Lehto Tuomas & Oinas-Kukkonen Harri (2009) The Persuasiveness of Web-Based Alcohol Interventions: A qualitative evaluation. IFIP Advances in Information and Communication Technology, Software Services for e-Business and e-Society, Vol. 305, pp. 316-327, 2009.
- 5. Lehto Tuomas & Oinas-Kukkonen Harri (2010) Persuasive Features in Six Weight Loss Websites: A Qualitative Evaluation. Lecture Notes for Computer Science, Persuasive, Vol. 6137, 2010.
- 6. Oinas-Kukkonen Harri & Chatterjee Samir (2009) Persuasive Technology: Introduction to the Special Section. Communications of the Association for Information Systems, Vol. 24, Article 23, pp. 395-398, March 2009.
- Oinas-Kukkonen Harri & Harjumaa Marja (2008) A Systematic Framework for Designing and Evaluating Persuasive Systems. Lecture Notes in Computer Science, Vol. 5033, pp. 164-176.
- 8. Oinas-Kukkonen Harri & Harjumaa Marja (2008) Towards Deeper Understanding of Persuasion in Software and Information Systems. The First International Conference on Advances in Human-Computer Interaction (ACHI '2008), Sainte Luce, Martinique, February 10-15, 2008, ISBN 978-0-7695-3086-4, pp. 200-205, Best Paper Award.
- 9. Oinas-Kukkonen Harri & Harjumaa Marja (2009) Persuasive Systems Design: Key Issues, Process Model, and System Features. Communications of the Association for Information Systems, Vol. 24, Article 28, pp. 485-500, March 2009.
- 10. Oinas-Kukkonen Harri & Kurki Heli (2009) Internet through the Eyes of 11-year Old Children: First-Hand Experiences from the Technological Environment the Children Live In. Human Technology: An Interdisciplinary Journal of Humans in ICT Environments, Vol. 5, No. 2, November 2009, pp. 146-162.
- 11. Oinas-Kukkonen Harri & Mantila Liisa (2009) "Lisa, Lisa, the Machine Says I Have Performed an Illegal Action. Should I tell the Police...?" A Survey and Observations of Inexperienced Elderly Internet Users,

Proceedings of the Southern Association for Information Systems Conference, Charleston, SC, USA, March 12-14, 2009, Georgia Southern University, pp. 146-151, Best Paper Nominee.

- Oinas-Kukkonen Harri (2010) Behavior Change Support Systems: Research Agenda and Future Directions. Lecture Notes for Computer Science, Persuasive, Vol. 6137, pp. 4-14, 2010.
- Oinas-Kukkonen Harri, Lyytinen Kalle & Yoo Youngjin (2010) Social Networks and Information Systems: Ongoing and Future Research Streams. Journal of the Association for Information Systems Vol 11, No. 2, February 2010, pp. 61-68.
- 14. Oinas-Kukkonen, H. (2010) Behavior Change Support Systems: The Next Frontier for Web Science. Proceedings of the Second International Web Science Conference (WebSci 10), Raleigh, NC, US, April 26-27, 2010
- 15. Oinas-Kukkonen, H. (2010) Discipline of Information Systems: A Natural Strategic Alliance for Web Science. Proceedings of the Second International Web Science Conference (WebSci 10), Raleigh, NC, US, April 26-27, 2010.
- 16. Räisänen Teppo, Lehto Tuomas & Oinas-Kukkonen Harri (2010) Practical Findings from Applying the PSD Model for Evaluating Software Design Specifications. Lecture Notes for Computer Science, Persuasive, Vol. 6137, 2010.
- 17. Räisänen Teppo, Oinas-Kukkonen Harri & Pahnila Seppo (2008) Finding Kairos in Quitting Smoking:

Smokers' Perceptions of Warning Pictures. Lecture Notes in Computer Science, Vol. 5033, pp. 263-266.

- Segerståhl Katarina & Oinas-Kukkonen Harri (2007) Distributed User Experience in Persuasive Technology Environments. *Lecture Notes in Computer Science*, Vol. 4744, Springer-Verlag, Berlin Heidelberg, pp. 80-91.
- 19. Segerståhl Katarina & Oinas-Kukkonen Harri (2010) Using the PSD Model for Evaluating a Persuasive Training Program: Lessons Learned. In: Poster Proceedings of PERSUASIVE 2010, Copenhagen, Denmark, June 7-10, 2010.
- 20. Segerståhl Katarina, Harjumaa Marja, Räisänen Teppo, Steiny Donald & Oinas-Kukkonen Harri (2008) Emerging Research Issues in Persuasive Ambient Intelligent Environments. In: Poster Proceedings of PERSUASIVE 2008, Oulu, Finland, June 2008, pp. 1-4.
- 21. Tørning, K. and Oinas-Kukkonen, H. (2009) Persuasive System Design: State of Art and Future Directions. ACM International Conference Proceeding Series, Vol. 350, Proceedings of the Fourth International Conference on Persuasive Technology, Claremont, CA, USA, April 26-29, 2009.
- 22. Wäljas Minna, Segerståhl Katarina, Väänänen-Vainio-Mattila Kaisa & Oinas-Kukkonen Harri (2010) Cross-Platform Service User Experience: A Field Study and an Initial Framework. In: Proceedings of the Twelfth International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '2010), Lisboa, Portugal, September 7-10, 2010.

Wildlife Surveillance Using GPS: From Movement Tracking to Behavior Recognition

Herbert Prins

Resource Ecology Group Wageningen University Wageningen, The Netherlands herbert.prins@wur.nl Andrea Kölzsch Netherlands Institute of Ecology Nieuwersluis, The Netherlands

a.koelzsch@nioo.knaw.nl

Arnold Bregt

Center for Geo-Information Wageningen University Wageningen, The Netherlands arnold.bregt@wur.nl

Tamme van der Wal AeroVision BV Naarden, The Netherlands tamme.vanderwal@aerovision.nl

ABSTRACT

Understanding the individual and social behavior of wild animals in the context of their habitat is imperative to protect and increase biodiversity. Our ability to observe individuals in the wild has increased exponentially with the availability of GPS. The increased miniaturization and prolonged operation time of GPS receivers and loggers allows us to consider new ways of analyzing and influencing animal behavior. This workshop brings together researchers, practitioners and technology providers discussing requirements for wildlife movement tracking, data processing and analysis. We will present the state-ofthe-art, exchange practical experiences and discuss future challenges.

Author Keywords

GPS, movement tracking, behavior recognition, feedback, intervention, wildlife management, software systems.

INTRODUCTION

Around the globe, initiatives are undertaken to protect and where possible - increase biodiversity. The declaration of 2010 as the International Year of Biodiversity by the United Nations illustrates that governments support the importance of biodiversity as a prerequisite for science, economy and society. For example, 'ecological corridors' and 'ecoducts' are constructed to increase freedom of movement for wildlife, to prevent splitting of populations, and to optimize the integration of nature conservation, agriculture and Lucas Noldus Noldus Information Technology BV Wageningen, The Netherlands l.noldus@noldus.nl

recreation. However, this can also lead to unwanted confrontations between humans and animals (e.g. traffic incidents), damage to crops, and the spreading of diseases between wildlife and farm animals.

Until recently, available technology (like radio tracking and satellite tracking techniques) suffered from limited temporal and spatial resolution, resulting in position data that are too coarse to determine how animals behave relative to the environment and each other. In other words: it tells you where the animal is but not what it is doing. Furthermore, the size and weight of transmitters and batteries made them unsuitable for small animals.

WORKSHOP OBJECTIVES

GPS technology allows moving objects to be tracked at small time intervals, and spatial resolution has improved dramatically over the years. However, analysis tools able to process large streams of GPS data and to convert spatial coordinates into behavioral parameters are still lacking. New approaches to extracting information and transforming it into knowledge have to be found. Furthermore, by adopting concepts from the field of artificial intelligence, GPS tracking can be extended with real-time data processing, interpretation and feedback, offering fascinating new opportunities for wildlife management and disease control. The objective of this workshop is to discuss the state-of-the-art, exchange practical experiences and define a first set of requirements towards a GPS-based behavior recognition and wildlife management system.

WORKSHOP PROGRAM

The workshop will consist of four technical topics:

- 1. **Measurement**: what to measure (location, time, physiological signals, ambient variables) and how
 - Possibilities of GPS relative to radio tracking and satellite tracking

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

- Logging versus real-time transmission of location data
- Temporal and spatial resolution: what can be achieved?
- GPS tracking under challenging conditions
- Integration of movement data with ambient parameters or physiological signals
- 2. **Analysis**: processing of position data and other variables into behavior recognition
 - Algorithms for location-based event recognition
 - Algorithms for behavior recognition and movement track segmentation
 - Integration of GPS data analysis and GIS systems
 - Software tools for analysis of movement patterns, individual behavior and interactions
- 3. Feedback & Intervention: closing the loop, from realtime processing to feedback, guidance and control
 - Real-time data reduction, analysis
 - · Feedback to the animal, guidance systems
 - · Feedback to the environment: control systems
- 4. **Hardware aspects**: energy issues, robustness, size (miniaturization), weight, etc.
 - Size and weight of GPS loggers
 - Battery life, duration of measurements
 - Impact of hardware on animal behavior and welfare

Each topic is important for GPS-based behavior analysis and guidance systems, but has individual characteristics and progress. Prior to the workshop a market survey is performed to gather information about current practices, experiences, applications, and requirements for future technology. The discussion of each topic will start with the analysis of the market survey. This is followed by a discussion on the state-of-the-art in this topic: where do we stand today, what is technically possible? Based on this the challenges based on user needs are defined: what lies ahead of us, which questions need to be answered, which technical hurdles must be taken? These are translated into the final step: product requirements for GPS based behavior measurement, analysis and feedback systems. What should a measurement, analysis of feedback system be able to do?



Figure 1. The workshop topics (hardware, measurement, analysis and feedback) are all discussed in 4 aspects: Demands, state-of-the-art, challenges and product requirements.

WORKSHOP SCOPE

Although the workshop is primarily targeted at wildlife, advances and lessons learned in animal husbandry and freeranging cattle among others are welcomed as that application domain has similar requirements towards the behavior analysis system.

WORKSHOP AGENDA

- Welcome and introduction (15 minutes)
- Topic 1: Measurement (45 minutes)
- Topic 2: Analysis (45 minutes)
- Demonstrations of hardware & software (30 minutes)
- Topic 3: Feedback & Intervention (45 minutes)
- Topic 4: Hardware aspects (45 minutes)
- Wrap-up: conclusions and follow-up activities (15 minutes)

TARGET AUDIENCE

This workshop brings together researchers, practitioners and technology providers discussing their needs for wildlife movement tracking, data processing and analysis. The organizers hope to initiate a lively discussion on the topics listed above. Attending this workshop will provide participants with a better understanding of the possibilities, limitations and future research directions of wildlife surveillance using GPS technology.

Measuring Consumer Behavior

René de Wijk

rene.dewijk@wur.nl

Jos Mojet jos.mojet@wur.nl

Wageningen UR Food & Biobased Research, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands

INTRODUCTION

Assessment of consumer behavior in specific situations, using observational and physiological methods, is becoming increasingly important in understanding conscious and unconscious consumer behavior. An increased understanding of consumer behavior may result in the development of improved consumer products and in more healthy dietary patterns. A growing number of techniques is available to assist researchers in measuring various aspects of consumer behavior such as walking patterns, product selection, meal composition, and eating/drinking. Due to advances in digital video, sensor technology and computer speed, complex measurements of behavior and physiology are now possible. Integration of these techniques allows multimodal measurements. With the growing number of techniques, the challenge for the researcher to choose the right solution becomes larger. Questions to be answered in order to select the right combination of solutions include:

- How do I upscale techniques that have proven themselves in the laboratory to real-life test situations such as supermarkets and restaurants? In the recently constructed 'Restaurant of the Future' in Wageningen (The Netherlands), observation of everyday behavior is combined with psychological, physiological and sensory measurements. Other research requires observation of consumers in other situations, such as hospitals, supermarkets, and school canteens.
- Which tools do I use for observation: video, physiological measurements, tracking, facial expressions analysis, head movement, eye-tracking or a combination of these methods?
- Measuring consumer behavior can result in an overload of acquired data. How do I select, analyze, transport, and store these data in an effective and safe way?

This workshop intends to bring together users and developers of measurement tools for consumer behavior, to discuss the state of the art, advancements, experiences, expectations and bottlenecks. As such it will provide a platform for exchanging information about the opportunities, challenges and needs in the fast developing area of consumer behavior research. It will be organized as a series of presentations, followed by group discussion.

Author Keywords

Eye tracking, dietary monitoring, facial expression analysis, observational analysis, people tracking.

PRESENTATIONS

Measuring Consumer Behavior in the Restaurant of the Future

René de Wijk (Consumer Science and Intelligent Systems, Wageningen UR, Wageningen, The Netherlands).

The Restaurant of the Future in Wageningen, the Netherlands, is an instrumented company lunch restaurant daily visited by 125 registered guests. Food choice behavior over repeated visits is monitored using various techniques including video, and is related to individual characteristics of the consumers. The presentations will illustrate the advantages of this approach as well as some of the technical pitfalls.

The Development of a System for Automatic Monitoring of Consumer Eating and Drinking Behavior

Oliver Amft (Wearable Computing Lab., ETH Zurich; Technical University Eindhoven, The Netherlands).

The recent uptake of activity recognition methods in various research fields demonstrates the maturity of basic sensor-based recognition of behaviour. While fundamental challenges in machine inference of context/situation awareness remain yet to be solved, systems that can describe dietary behaviour become feasible. Current trends in multimodal activity recognition and behaviour inference will be discussed. In particular, previous and ongoing work on Automatic Dietary Monitoring and related studies are presented to illustrate how relevant information can be recognized in daily life using unobtrusive sensors. I will address current research issues related to the diversity of activities and pattern variability. Finally, some new project

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.



Photographs from left to right: Automated people tracking, video observation of consumers in the Restaurant of the Future, automated facial expression recognition, eye tracking, automated dietary monitoring.

efforts are summarized, which will commence in the next months at the newly established ACTLab at TU Eindhoven.

Is Eye-Tracking an Effective Experimental Tool for Capturing Consumers Attention?

Svetlana Bialkova & Hans van Trijp (Marketing and Consumer Behavior Group, Wageningen University).

Understanding how consumers pay attention to information and extract meaning is crucially important in credence attributes such as health, and sustainability. Hence there is a need to develop appropriate measures for adequately capturing whether and how consumers pay attention to particular information. The current study discusses Eyetracking methodology as an effective experimental tool for capturing consumers attention in shopping situations and thus a promising approach for exploring consumer behavior.

Use of FaceReader for Measuring Facial Expressions in Real-Life Situations

Marten den Uyl (VicarVision bv, Amsterdam, The Netherlands).

FaceReading, the automated analysis of facial expressions, seems a very promising new technology for assessment of consumer behavior and experience, since the assessment is unobtrusive and continuous and may give direct insight in conscious and unconscious affective response of subjects. However there problems and limitations associated with the technology: lighting conditions and camera distance and angle are important to obtain the high quality video recordings required; chewing motions and eating behavior causing occlusions –hand for mouth- may interfere with emotion assessment; and subjects may vary considerably in their tendency to 'facialize' i.e. show affect in the face. In the presentation experiences with FaceReading in real life situations and open challenges for further development of the technology will be discussed.

Inside Consumer Experience: Studying Consumer Behavior in the Field

Leanne Loijens (Noldus Information Technology bv, Wageningen, The Netherlands).

The aim of the Inside Consumer Experience project is to develop mobile tools and services to measure food selection and consumption in real-life contexts like restaurants, shops, elderly homes or festivals. The results of a number of pilot studies will be presented in which new techniques have been tried out. The advantages of the tested techniques relative to other methods will be discussed and also the hurdles that have to be taken to make the techniques widely applicable.

AUDIENCE

This workshop aims to bring together researchers involved in consumer behavior research, as well as developers of technology and tools for measuring consumer behavior.

Workshop on Teaching a Course on Measuring Behaviour: Philosophy, Concepts, Experiments, and Resources

Richard E. Brown rebrown@dal.ca **Timothy P. O'Leary**

Psychology Department and Neuroscience Institute, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1

ABSTRACT

The purpose of this workshop is to discuss the teaching of undergraduate and graduate courses in "Measuring Behaviour". The authors will discuss their experiences and share their course outline and then discuss (1) possible topics to be covered in such a course, (2) textbooks and readings, (3) laboratory (practical) projects, (4) equipment needed, (5) ethics for testing animal and human subjects, (6) grading, and (7) student evaluations.

Author Keywords

Teaching, lectures, laboratories, equipment, resources.

INTRODUCTION

How should a course in "Measuring Behaviour" be designed? This workshop will discuss the requirements for teaching a laboratory course on Measuring Behaviour and develop ideas for future courses. We have developed such a course, which has one 90-min. lecture and one 90-min. laboratory class per week for a 13 week term for 24 third year Psychology and Neuroscience students. In the first part of the workshop, I will give an overview of the organization of this course, the topics covered in the lectures and the laboratory projects. In the second part, the floor will be open for discussion of other people's experiences with teaching such a class.

PART 1: COURSE OVERVIEW

The lecture portion of the course was designed to discuss the issues involved in understanding the importance of measuring behaviour, from fruitfly courtship behaviour to human facial expression and social behaviour. The laboratory component was designed to give practical experience in conducting behavioural research.

Lecture Classes

We used the textbook "Measuring Behaviour", 3rd edition, by Paul Martin and Patrick Bateson, plus a number of journal articles and course notes that we prepared ourselves. The lectures covered the following topics: (1) What is behaviour? (2) Qualitative description of behaviour; (3) Quantitative methods for describing behaviour; (4) Advanced quantitative methods; (5) Why measure behaviour? Theory and hypothesis testing; (6) The functional analysis of behaviour: experimental design and statistics; (7) Ethical issues in experimentation with animals and humans; animal health and welfare; (8) How can you measure your own behaviour? (9) Analyzing bouts of behaviour; (10) Measuring social behaviour; (11) Measuring facial expressions: can behaviour be used to detect lying? (12) Measuring behavioural development; (13) Errors in measuring behaviour and the use of automated equipment to measure behaviour.

Lectures involved the use of regular powerpoint (overhead) presentations, plus the use of demonstrations and videos from the internet. These included examples of fruitfly courtship behaviour, human facial expressions, magician's tricks, and the social behaviour of children. Some of these will be described. All course materials were available on the class website which students could access. Students wrote two exams to assess their comprehension of the lecture material.

Laboratory Projects

The laboratories enabled students to use the techniques discussed in class in four different projects: (1) Qualitative and quantitative description of mouse home cage behaviour; (2) Qualitative and quantitative description of mouse behaviour in the open-field and elevated plus maze: as measured by students and automated apparatus; (3) Sequential analysis of grooming behaviour in stressed and non-stressed mice, and (4) Observing the behaviour of pedestrians at cross-walks (Independent project). A poster on the independent project is presented to illustrate the independent project (Brown, O'Leary and Allen, 2010).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. For any other use, please contact the Measuring Behavior secretariat: info@measuringbehavior.org.

For the laboratory projects, we used a video camera, video playback system, computerized behavioural scoring program, and a computerized video-tracking system. Students wrote and were evaluated on four lab reports. Other laboratory projects that we are developing include measuring Siamese fighting fish display behaviour and gait analysis in humans.

PART 2: DISCUSSION ON TEACHING CLASSES

The workshop will examine new ideas for (1) classroom presentations, demonstrations and short videos, (2) in-class experiments, (3) topics for laboratory projects; (4) the equipment necessary to teach such a class, and (5) the

development of on-line resources for lecture and laboratory projects. It is expected that the workshop participants will give examples from their own experiences, discuss projects which "work" and those that do not "work" with such a class, and discuss ideas for the development of web-based resources.

REFERENCES

1. Richard E. Brown, Timothy P. O'Leary, Stephanie L. Allen 2010. Who said pedestrian means dull: Teaching methods in the measurement of behaviour through observational studies of pedestrian behaviour. *Proceedings of MB2010*, 392-393.

Acknowledgements

Scientific Program Committee

- Boris de Ruyter, Cognitive & Behavioral Insights User Experiences Department, Philips Research Europe, Eindhoven, The Netherlands
- Emilia Barakova, Faculty of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands
- Andrew Spink, Noldus Information Technology bv, Wageningen, The Netherlands
- Mathilde Bekker, Faculty of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands
- Egon van den Broek, Professor in Man-Machine Interaction, Technical University of Vienna, Vienna, Austria
- Stephen Campbell, Faculty of Education, Simon Fraser University, Burnaby, Canada
- Nicky Clayton, Department of Experimental Psychology, University of Cambridge, Cambridge, UK
- Wim Crusio, Research Director, French National Research Council, Talence, France
- Robert Gerlai, Department of Psychology, University of Toronto, Mississauga, Ontario, Canada
- Eco de Geus, Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
- Ilan Golani, Department of Zoology, Tel Aviv University, Tel Aviv, Israel
- Natalia Romero Herrera, Department of User Centered Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
- John Krantz, Department of Psychology, Hanover College, Hanover, IN, USA
- Effie Law, Inst.f.Techn.Informatik u.Kommunk.Netze, Swiss Federal Institute of Technology, Zurich, Switzerland
- Fabio Paternò, Laboratory on Human Interfaces in Information Systems, Institute of Information Science and Technologies, 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
- Ahmed Seffah, Department of Computer Science, Concordia University, Montreal, Canada
- Berry Spruijt, Behavioral Biology Group, Department of Biology, Utrecht University, Utrecht, The Netherlands
- Oliver Stiedl, Behavioral and Cognitive Neuroscience Group CNCR, Institute of Neurosciences FALW, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
- Manfred Tscheligi, ICT&S Center, University of Salzburg, Salzburg, Austria
- Louise Vet, Netherlands Institute of Ecology (NIOO-KNAW), Nieuwersluis, and Laboratory of Entomology, Wageningen University, Wageningen, The Netherlands

Local Organizing Committee

- Natasja Bogers, Chair, Local Organizing Committee
- Andrew Spink, Co-chair, Program Committee
- Melvin de Bruijn, Finances
- Sabine van der Meijde, Conference secretariat

Author Index

| Abel, Troy | |
|---------------------------|---------------------------------|
| Adriani, Walter | 145, 422 |
| Aerts, Nancy | 414 |
| Ahlstrom, Christer | 196, 310 |
| Allen, Colin | |
| Allen, Stephanie L. | |
| Altimus, Cara | |
| Anastasopoulou, Panagiota | 41 |
| Andreasson, Theresa | 407, 425 |
| Anguera, M. Teresa16, 177 | , 243, 294 |
| Aoki, Hirotaka | |
| Arabadzisz, Dimitrula | |
| Arabo, A | |
| Arad, Maya | 48 |
| Araya, Jaime | 439 |
| Arestova, Olga | |
| Avdesh, Avdesh | 155 |
| Backes, Heiko | |
| Baganz, Daniela | |
| Ballesteros, Carolina | 439 |
| Balsters, Martijn | |
| Balzer, Ullrich | |
| Barakova, Emilia | 6, 70, 77 |
| Bartlett, Marni | |
| Battaglia, Francesco P. | |
| Beerling, Wieteke | |
| Bellante. Antonio | |
| Benjamini. Yoav | 92 |
| Bernhaupt. Regina | |
| Bhat. Pavitra | |
| Bischof, Horst | |
| Bladel. Cees van | |
| Blum. Tobias | 64 |
| Bons, Danielle M.A. | 109 |
| Böriesson, Jenny | 407 |
| Bos. Kees van den | 114 |
| Bos, Ruud van den | 145 |
| Bouarfa, Loubna | 66 |
| Boughorbel Sabri | 349 |
| Bouwknecht I. Adriaan | 403 |
| Bowen, Michael T. | |
| Boxtel, Anton van | 104 |
| Braak Cajo IF ter | 171 |
| Brand Nico | 396 |
| Bredero Maarten Douwe | 477 |
| Breehaart Jeroen | 340 |
| Breat Arnold | /105 |
| Brendle Julia | 262 |
| Brockmann Carsten | |
| Broek Egon L van den | Q/ Q5 |
| Brobez Catherine | , , , , , , , , , , , , , , , , |
| Bronez, Caulerine | |

| Brown, Richard E. | 392, | 499 |
|--|--------|--|
| Bruder, Carmen | | 334 |
| Bruekers, Fons | | 349 |
| Buffa, Gaspare | | 427 |
| Buitelaar, Jan K. | | 109 |
| Bulten, Eric | | 396 |
| Bulushev, Evgeny D. | | 379 |
| Buscaino, Giuseppa | | 427 |
| Calaminus, Carsten | | 430 |
| Callum, Jeannie L. | | 297 |
| Camerino, Oleguer | | 177 |
| Cao, Yujia | | 453 |
| Cárdenas-López, Georgina | | 441 |
| Carey, James R. | | 301 |
| Casarrubea, Maurizio | | 133 |
| Castañer. Marta | 243. | 294 |
| Castela. Marta | 284. | 355 |
| Chen. Mengai | , | 155 |
| Chen. Wei | | .44 |
| Chew. Chelsea | | |
| Chignell Mark | | 297 |
| Ciullo Gianluca | | 281 |
| Clark Anne B | | 369 |
| Clemens Laura Emily | | 475 |
| Coca Patricia de | 11 | 13 |
| Connin Cillos | | , 15 010 |
| | | 212 |
| Courtemanche François | •••••• | 212 218 |
| Courtemanche, François Craependonck Hansfried V M M yan | 373 | 212 218 414 |
| Courtemanche, François Craenendonck, Hansfried V. M. M. van | .373, | 212 218 414 133 |
| Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe | .373, | 212 218 414 133 430 |
| Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav | .373, | 212 218 414 133 439 |
| Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Dangapoy, Zhargal | .373, | 212 218 414 133 439 66 |
| Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal | .373, | 212 218 414 133 439 66 174 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin | .373, | 212 218 414 133 439 66 174 7 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya Da La Vaca | .373, | 212 218 414 133 439 66 174 7 247 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita Da Lamos, Jakoh | 373, | 212 218 414 133 439 66 174 7 247 448 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob | .373, | 212 218 414 133 439 66 174 7 247 448 223 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob Deák, Gedeon O | .373, | 212 218 414 133 439 66 174 7 247 448 223 247 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob De Lemos, Jakob Deák, Gedeon O. | .373, | 212 218 414 133 439 66 174 7 247 448 223 247 141 |
| Coppil, Onles Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob De La Vega, Julita De Lemos, Jakob Deák, Gedeon O. | 373, | 212 218 414 133 439 66 174 227 448 223 247 141 368 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob De Lemos, Jakob Deák, Gedeon O Di Lio, Alessandra Di Stefano, Vincenzo | .373, | 212 218 414 133 439 66 174 7 247 448 223 247 141 368 427 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob Deák, Gedeon O. Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo | 373, | 212 218 414 133 439 66 174 7 247 448 223 247 141 368 427 465 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob De ák, Gedeon O. Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo Di Stefano, Vincenzo | 373, | 212 218 414 133 439 66 174 247 247 448 223 247 141 368 427 465 73 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De Barbaro, Kaya De La Vega, Julita De La Vega, Julita De La Vega, Julita De Lemos, Jakob Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo Dielenberg, Robert A Diggelen, Jurriaan van Dijk, Betsy van | 373, | 212 218 414 133 439 66 174 7 247 448 223 247 141 368 427 465 73 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De La Vega, Julita De Lemos, Jakob Deák, Gedeon O. Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo Dielenberg, Robert A. Diggelen, Jurriaan van Dijk, Betsy van | 373, | 212 218 414 133 439 66 174 227 247 448 223 247 141 368 427 465 73 453 61 |
| Coppil, Onles Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob Deák, Gedeon O. Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo Di Stefano, Vincenzo Dielenberg, Robert A. Diggelen, Jurriaan van Dijk, Betsy van Dijk, C. Niek van | 373, | 212 218 414 133 439 66 174 7 247 448 223 247 141 368 427 465 73 453 61 281 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob Deák, Gedeon O. Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo Di Stefano, Vincenzo Dielenberg, Robert A. Diggelen, Jurriaan van Dijk, Betsy van Dijk, C. Niek van Dinardo, Cinzia | 373, | 212 218 414 133 439 66 174 7 247 247 247 247 448 223 247 141 368 427 465 73 453 61 281 212 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De Lemos, Jakob De ák, Gedeon O. Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo Di elenberg, Robert A. Diggelen, Jurriaan van Dijk, Betsy van Dijk, C. Niek van Dinardo, Cinzia | 373, | 212 218 414 133 439 66 174 7 247 247 247 247 247 247 247 368 423 247 141 368 427 465 73 453 61 281 212 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De La Vega, Julita De Lemos, Jakob Deák, Gedeon O. Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo Dielenberg, Robert A. Diggelen, Jurriaan van Dijk, Betsy van Dijk, C. Niek van Dinardo, Cinzia Diverrez, Jean-Marc | 373, | 212 218 414 133 439 66 174 439 247 448 223 247 141 368 427 465 73 453 61 281 212 246 370 |
| Courtemanche, François | 373, | 212 218 414 133 439 66 174 227 247 448 223 247 141 368 427 465 73 453 61 281 212 466 370 419 |
| Courtemanche, François Courtemanche, François Craenendonck, Hansfried V. M. M. van Crescimanno, Giuseppe Curkovic, Tomislav Dankelman, Jenny Danzanov, Zhargal Dautenhahn, Kerstin De Barbaro, Kaya De La Vega, Julita De La Vega, Julita De Lemos, Jakob Deák, Gedeon O. Detrait, Eric Di Lio, Alessandra Di Stefano, Vincenzo Dielenberg, Robert A. Diggelen, Jurriaan van Dijk, Betsy van Dijk, C. Niek van Dijk, C. Niek van Dinardo, Cinzia Diverrez, Jean-Marc Drinel, Katja van. Dirinkenburg, W.H.I.M. | 373, | 212 218 414 133 439 66 174 247 448 223 247 141 368 427 465 73 453 61 281 212 466 370 419 218 |

| Economides Anastasios A | |
|---|--|
| | |
| Edelman, Nicholas | 149 |
| Egger, Veronika | |
| Egmond, Cees van | 68 |
| Ehrenstrasser, Lisa | |
| Ehrismann, Jasmin Simone | 475 |
| Eißfeldt, Hinnerk | 334 |
| Elzen, Stef van den | |
| Embrechts, Sofie | |
| Endepols, Heike | 430 |
| Fekas, Dimitris | 171 |
| Felismino, Marilia | . 284, 355 |
| Ferkingstad, John | |
| Feußner, Hubertus | 64 |
| Filiciotto, Francesco | |
| Fischer, Thomas | |
| Flinsenberg, Ingrid | |
| Flores-Plata, Lorena Alejandra | |
| Fonio, Ehud | 92 |
| Forster, Deborah | |
| Fouse, Adam S. | |
| Frette, Vidar | |
| Fujii, Naotaka | 7 |
| Gahr, Manfred | 58 |
| Garganta, Júlio | 16 |
| Garrote, Estibaliz | |
| Gelder, Beatrice de | |
| Gener, Thomas | .436,448 |
| Giardino Luciana | , |
| | 411 |
| Giuliani, Alessandro | 411 411 |
| Giuliani, Alessandro Gizatdinova, Yulia | 411 411 208 |
| Gialdino, Edelana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan | 411 411 208 92 |
| Gialdino, Edelana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin 185, 322, 326 | 411 411 208 92 459, 470 |
| Giardino, Euclaria Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonz Fuentes Daniel | 411 |
| Gialdino, Edelana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez Jsabel | 411 |
| Gialdino, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gonzalez, Isabel Gorbunov Roman | 411 208 92 459, 470 233 204 70, 77 |
| Giaidino, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf Rudolf | 411 411 208 92 459, 470 233 204 70, 77 430 |
| Giatuno, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Jaabel Gorbunov, Roman Graf, Rudolf | 411 411 208 92 459, 470 233 204 70, 77 430 334 |
| Giautino, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger Christoph | 411 411 208 208 204 233 204 70, 77 430 334 34 |
| Gialdino, Edelana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulvas Laszlo | 411 411 208 208 233 204 70, 77 430 334 436 34 |
| Gialdino, Edelana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes Hatice | 411 208 92 459, 470 233 204 70, 77 430 334 436 185 122 |
| Giadino, Edefana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Guesicilia Marco | 411 411 92 459, 470 233 204 70, 77 430 334 436 185 122 411 |
| Giaduno, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Jsabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco | 411 411 92 459, 470 233 204 70, 77 430 334 436 122 411 72 |
| Giatuno, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gutzan, Christian | 411 411 208 92 459, 470 233 204 70, 77 430 334 436 122 411 72 82 |
| Giadino, Ederana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gushin, Vadim Guzen, Christian | 411 411 208 233 204 70, 77 430 334 436 185 122 411 72 83 83 |
| Giadino, Edefana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gushin, Vadim Gutzen, Christian Gygax, Lorenz | 411 411 208 208 208 208 204 70, 77 430 334 436 122 411 72 411 72 83 83 84 84 |
| Giatumo, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gusciglio, Marco Gushin, Vadim Gutzen, Christian Guzen, Christian | 411 411 208 233 204 70, 77 430 334 436 122 411 72 83 83 83 83 |
| Giatumo, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gusciglio, Marco Gushin, Vadim Gutzen, Christian Guzen, Christian Guzen, Christian Guzen, Christian Guzen, Christian Guzen, Christian Guzen, Christian Guzen, Christian | 411 411 208 |
| Giaddino, Edefana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gusciglio, Marco Gushin, Vadim Gutzen, Christian Gustan, Christian Gustan, Christian Gustan, Christian Gustan, Christian Gustan, Christian Gutzen, Christian Gustan, Christian Gutzen, Christian Gustan, Christian Gustan, Martin Halasz, Paul | 411 411 92 459, 470 233 204 70, 77 430 334 436 122 411 72 83 403 403 403 404, 465 |
| Giaddino, Edefania Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gushin, Vadim Gutzen, Christian Gutzen, Christian Gutzen, Christian Gutzen, Christian Gutzen, Christian Gutzen, Christian Hales, Patrick de Haindl, Martin Halasz, Paul | 411 411 411 92 459, 470 233 204 70, 77 430 334 436 122 411 72 83 38 403 403 464, 465 297 |
| Giautino, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gushin, Vadim Gutzen, Christian Gutzen, Christian Gutzen, Christian Gutzen, Christian Haes, Patrick de Haindl, Martin Halasz, Paul Hamiliany, Allison | 411 411 208 92 459, 470 233 204 70, 77 430 334 436 185 122 411 72 83 403 403 419 .464, 465 297 489 |
| Giatuno, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gushin, Vadim Gutzen, Christian Gutzen, Christian Guzen, Christian Gutzen, Christian Gutzen, Christian Haes, Patrick de Haindl, Martin Halasz, Paul Halliday, Allison Ham, Jaap | 411 411 411 208 92 459, 470 233 204 70, 77 430 334 436 122 411 72 38 403 403 445 297 489 464, 465 297 |
| Giatumo, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gushin, Vadim Gutzen, Christian Gygax, Lorenz Haes, Patrick de Haindl, Martin Halasz, Paul Halliday, Allison Ham, Jaap Hambartsumyan, Henri | 411 411 411 208 208 208 208 203 204 70, 77 430 334 430 436 122 411 72 83 403 403 464, 465 297 489 464, 475 297 489 416 |
| Giatumo, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gushin, Vadim Gutzen, Christian Guzen, Christian Guzen, Christian Guzen, Christian Guzen, Christian Haes, Patrick de Haindl, Martin Halasz, Paul Halliday, Allison Ham, Jaap Hambartsumyan, Henri Hámornik, Balázs Péter Lasz | 411 411 208 233 204 70, 77 430 334 436 122 411 72 431 413 403 403 419 .464, 465 297 489 465, 416 36 36 |
| Giatumo, Euclana Giuliani, Alessandro Gizatdinova, Yulia Golani, Ilan Golz, Martin Golz, Martin Gonzalez, Isabel Gorbunov, Roman Graf, Rudolf Grasshoff, Dietrich Guger, Christoph Gulyas, Laszlo Gunes, Hatice Gusciglio, Marco Gusciglio, Marco Gushin, Vadim Gutzen, Christian Gutzen, Christian Gutzen, Christian Haes, Patrick de Haindl, Martin Halasz, Paul Halliday, Allison Ham, Jaap Hambartsumyan, Henri Hanon, Robert E. Hanon, Ritienne | 411 411 208 233 204 70, 77 430 334 436 185 122 411 72 83 403 419 444, 465 297 489 365, 416 36 36 36 |

| Haslwanter, Thomas | |
|--|--|
| Hasse, Catrin | |
| Hattar, Samer | |
| Hauber, Wolfgang | |
| Hawkins, Penny | |
| Heath, Richard | |
| Heer, Raymond C. de | |
| Heinze, Christian | 322, 326, 459, 470 |
| Hens, Koen | |
| Henttonen, Pentti | |
| Heyvaert, Mieke | |
| Hoeks, Kim | |
| Hoeksma, Marco | |
| Hollan, James D | |
| Hötzl, Thomas | |
| Hu, Jun | |
| Huo. Ming | |
| Imbimbo, Bruno | |
| Ingram, Donald | 301 |
| Ioakeimidis, Ioannis | 338 |
| Iorgovan Dusan | 212 |
| Itzhak Vossef | 137 |
| Jansen René F | 58 83 |
| Jauernig Oliver | 433 |
| Ihuang Hueihan | 1/10 |
| Jofre Toni | |
| Johnson Christing M | 2/7 |
| | |
| LONKOR LUCTOR L | 66 |
| Jonker, Pieter P | |
| Jonker, Pieter P Jonsson, Gudberg K | . 177, 243, 273, 303 |
| Jonker, Pieter P Jonsson, Gudberg K Józsa, Eszter | |
| Jonker, Pieter P Jonsson, Gudberg K Józsa, Eszter Juhász, Márta | |
| Jonker, Pieter P Jonsson, Gudberg K Józsa, Eszter Juhász, Márta Jyoti, Amar | |
| Jonker, Pieter P Jonsson, Gudberg K Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George | |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri | |
| Jonker, Pieter P Jonsson, Gudberg K Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia | |
| Jonker, Pieter P Jonsson, Gudberg K Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. | |
| Jonker, Pieter P Jonsson, Gudberg K Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald | |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten | 66 177, 243, 273, 303 215 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. | 66 177, 243, 273, 303 215 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. | 66 177, 243, 273, 303 215 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter | 66 177, 243, 273, 303 215 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. | |
| Jonker, Pieter P | |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kirkwood, Alfredo | 66 177, 243, 273, 303 215 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle | 66 177, 243, 273, 303 215 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle Kölzsch, Andrea | 66 177, 243, 273, 303 215 215 341, 345, 416 185 235 385 89 401 349 162 61 466 88 310 444 382 495 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle Kölzsch, Andrea | 66 177, 243, 273, 303 215 341, 345, 416 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle Kölzsch, Andrea Koolhaas, Jaap M. Koot, Susanne | 66 177, 243, 273, 303 215 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle Kölzsch, Andrea Koolhaas, Jaap M. Koot, Susanne Krahmer, Emiel | 66 177, 243, 273, 303 215 |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle Kölzsch, Andrea Koolhaas, Jaap M. Koot, Susanne Krahmer, Emiel Krajewski, Jarek | 66 177, 243, 273, 303 215 341, 345, 416 26 185 235 385 89 401 349 162 61 466 88 310 444 382 495 370 145 239 322, 326, 459, 470 |
| Jonker, Pieter P | |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle Kölzsch, Andrea Koolhaas, Jaap M. Koot, Susanne Krahmer, Emiel Krajewski, Jarek Krupenia, Stas | |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle Kölzsch, Andrea Koolhaas, Jaap M. Koot, Susanne Krahmer, Emiel Krajewski, Jarek Krupenia, Stas Kubová, Hana | |
| Jonker, Pieter P. Jonsson, Gudberg K. Józsa, Eszter Juhász, Márta Jyoti, Amar Kampis, George Kandalla, Mayuri Karpukhina, Anastasia Kas, Martien J.H. Kastberger, Gerald Kate, Warner ten Kenny, David A. Kerkhoffs, Gino M.M.J. Kettlewell, Peter Kim, Jeansok J. Kircher, Katja Kirkwood, Alfredo Kleppe, Gisle Kölzsch, Andrea Koolhaas, Jaap M. Koot, Susanne Krahmer, Emiel Krajewski, Jarek Kirdien, Heidi Krupenia, Stas Kubová, Hana Kuippes, Erik | |

| Kuyck, Kris van | | 365 |
|-----------------------------|---------|-----|
| Kyung, Gyouhyung | | 318 |
| Lassi, Glenda | | 30 |
| Laufenberg, Tom | | 322 |
| Laviola, Gianni | | 145 |
| Laviola, Giovanni | | 422 |
| LeGates, Tara | 444, | 451 |
| Leiva, David | 162, | 230 |
| Lex, Björn | | 430 |
| Liedo, Pablo | | 301 |
| Lipp, Hans-Peter | | 368 |
| Littlewort, Gwen | | 250 |
| Liu, Hao | | 265 |
| Ljung, Elisabeth | | 407 |
| Lógó, Emma | | 215 |
| Lopes, António | | 177 |
| Lopez-Pigozzi, Diego | 436, | 448 |
| Lorenzini, Luca | | 411 |
| Luiz, Ricardo | 284, | 355 |
| Luyten, Laura | í | 365 |
| Maat. Andries ter | | 58 |
| Machin. Ian | | 426 |
| Mahieu Michel | | 373 |
| Maierbrugger, Gudrun | | 314 |
| Makeig Scott | | |
| Mäkinen Erno | | 208 |
| Man Singh Madhusudan | | 401 |
| Manolov Rumen | 158 | 230 |
| Mares Pavel | 150, | 399 |
| Marques Pedro | 284 | 355 |
| Martin-Iverson Mathew T | 204, | 155 |
| Martine Ralph N | ••••• | 155 |
| Mary Christina | ••••• | 133 |
| Maria Pryon | | 204 |
| Masus, Diyan | ••••• | 204 |
| Matthews, Gloria | ••••• | 394 |
| Maurer, Michael | | 401 |
| McGregor, Iain S. | 464, | 465 |
| Meert, Theo F. | | 403 |
| Megens, Anton A.H.P. | | 414 |
| Mellace, Marco | ••••• | 171 |
| Mertens, Liesbeth | •••• | 409 |
| Meschtscherjakov, Alexander | 287, | 489 |
| Meynen, Michel | | 403 |
| Meza, Rosemary | | 250 |
| Mies, Günter | ••••• | 430 |
| Miguel, Catarina | ••••• | 243 |
| Mihelj, Matjaz | | 269 |
| Miheyeva, Natalya | | 463 |
| Miklosi, Adam | | 185 |
| Mikulecká, Anna | 399, | 446 |
| Millonig, Alexandra | | 314 |
| Mirlacher, Thomas | | 352 |
| Mitchell, Malcolm | | 466 |
| Moiseeva, Anastasia | | 291 |
| Mojet, Jos | | 497 |

| Mondal, Alinda | |
|---------------------------------------|---------------|
| Mooij-van Malsen, Annetrude (J.G.) de | |
| Moran, Nance | |
| Morey, Gaspar | |
| Morice, Amy | |
| Moridis, Christos N | |
| Muehlemann, Thomas | |
| Müller, Hans G. | |
| Munih, Marko | |
| Munsters, Carolien | |
| Navab, Nassir | 64 |
| Naylor, Alice | |
| Neerincx, Mark | 73 |
| Neuman, R | |
| Nguyen, Huu Phuc | 475 |
| Nicholson, Janet R. | |
| Nicolas, Céline | |
| Niculescu, Andreea | |
| Nijholt, Anton | 94, 95, 453 |
| Nolan, Patrick M. | |
| Noldus, Lucas | 4, 495 |
| Noortgate, Wim van den | 456 |
| Novak, Domen | |
| Nussbaum, Maury A | |
| Nuttin, Bart | |
| O'Leary, Timothy P. | 392, 499 |
| Oinas-Kukkonen, Harri | 491 |
| Ólafsdóttir, Íris | |
| Oldenrijk, Jakob van | 68 |
| Olensek, Andrej | |
| Olivier, Berend | |
| Onghena, Patrick | 158, 456 |
| Panteleeva, Sofia | 174 |
| Pantic, Maja | |
| Papadopoulos, Nikos | |
| Parada, Francisco J. | |
| Pemberton, Darrel J. | 372, 409, 414 |
| Pendergrast, Jacob | |
| Perez-Mendez, Lorena | 436, 448 |
| Phan, Linda | |
| Pieneman, Anton W. | |
| Pitcher, Thomas E. | |
| Plano. Andrea | |
| Platt. Bettina | 24.26 |
| Plyusning Irene Z | 379 |
| Poggio Tomaso | 149 |
| Ponten Henrik | 425 |
| Poolman Rudolf W | |
| Portal Esteban | 00 173 175 |
| Potier C | |
| Prins Herbert | 40U |
| Prom Tan Sandring | |
| Trom rep, Sanurine | |
| Province, Juan I | |
| Nacymaekers, Leen J. | |
| каізаню, кооре | |

| Rauterberg, Matthias | 44, 77, 265 |
|------------------------------|---------------|
| Ravaja, Niklas | 55 |
| Reefmann, Nadine | |
| Reilly, Judy | |
| Reitberger, Wolfgang | |
| Remmelink, Esther | |
| Reznikova, Zhanna | |
| Ribes Iñesta, Emilio | |
| Riedel, Gernot | |
| Riess, Olaf | |
| Rijk, Elisa | 68 |
| Rizzi, Luca | |
| Robles, Leopoldo | |
| Rommelse, Nanda N.J. | |
| Roof, Robin | |
| Ross, Megan R. | |
| Ross. Stephen R. | |
| Rossi, Alejandra | 226 |
| Roy V | 480 |
| Rüther Matthias | 401 |
| Runne Wouter C | |
| Ruyter Boris de | 1 489 |
| Ryck Marc De | |
| Rydén Markinhuhta Katarina | |
| Saarinan Mikaal | 407 |
| Sahli Hicham | 204 |
| Salminan Mikko | 204 |
| Sanchaz Vivas Maria Victoria | 136 118 |
| Santangelo Andrea | 430, 448 |
| Sandangero, Andrea | |
| Savensbergh, Geeft J.I. | 20 |
| Savino, Nicola | |
| Scharrow, Floor E | 01, 08 |
| Scheepers, Floor E. | |
| Schenka, Adoli | |
| Schenka, Corinne | |
| Schnieder, Sebastian | |
| Schnupp, Thomas | 326, 459, 470 |
| Schrom-Feiertag, Helmut | |
| Schulze, Katrin | |
| Schwarzer, Stephan | |
| Seljestad, Heidi | |
| Sellen, Katherine M | |
| Senechal, Sylvain | |
| Serre, Thomas | 149 |
| Serriello, Joseph | |
| Sierevelt, Inger N. | 61 |
| Snyder, Bradley | |
| Södersten, Per | |
| Solanas, Antonio | 158, 162, 230 |
| Solov'eva, Maria Yu. | |
| Sommer, David185, | 322, 326, 459 |
| Soós, Juliánna Katalin | |
| Sorbera, Filippina | 133 |
| Souchard, Laurent | |
| Spruijt, Berry M. | |

| Staaks, Georg | |
|--|---|
| Stassen, Laurents P.S | 66 |
| Steele, Andrew | 149 |
| Steffen, Hoy | |
| Stel, Marielle | 114 |
| Stiedl, Oliver | 58, 82, 83 |
| Stienen, Bernard M.C. | |
| Straus, Sharon E | |
| Stumpp, Jürgen | 41 |
| Šubrt, Martin | |
| Sunesson, Ylva | |
| Surakka, Veikko | |
| Suzuki, Satoshi | |
| Swerts, Marc | |
| Sydnes, Tone | |
| Talling, Janet | |
| Tan, CheeFai | 44 |
| Tarancev, Igor G. | |
| Telepak, Gregory | |
| Terzis, Vasileios | |
| Thorisson, Kristinn R. | |
| Thorngren, Marianne | |
| Thorsteinsson, Vilhjalmur | |
| Ticha, Katerina | |
| Timmermans, Harry | |
| Tomasson, Gunnar Gudni | |
| Torrents, Carlota | |
| Truong Khiet P | 00 |
| Theory, Theory is a second sec | |
| Tscheligi, Manfred | .287,489 |
| Tscheligi, Manfred Tucci, Valter | .287,489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. | .287, 489 30 59, 61 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl | .287, 489 30 59, 61 70 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio | .287, 489 30 59, 61 70 118 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke | .287, 489 30 59, 61 70 118 405 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate. Wim | 287, 489 59, 61 70 118 405 456 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle. Patrick | 287, 489 59, 61 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vansteenwegen, Debora | 287, 489 59, 61 118 405 456 277 365 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna | 287, 489 30 59, 61 70 118 405 456 277 365 394 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris | 287, 489 59, 61 70 118 405 456 365 394 394 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris Vargas, Ariel | 287, 489 59, 61 70 118 405 456 277 365 394 277 253 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris Vargas, Ariel Vén, Ildikó | 287, 489 59, 61 59, 61 118 405 456 277 365 394 277 253 341 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Varsteenwegen, Debora Varekamp, Chris Vargas, Ariel Vén, Ildikó Ver Donck, Luc | 287, 489 59, 61 59, 61 118 405 456 277 365 394 277 253 341 373 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris Vargas, Ariel Vén, Ildikó Ver Donck, Luc Verdile, Giuseppe | 287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Vardanyan, Anna Vargas, Ariel Vén, Ildikó Ver Donck, Luc Verdile, Giuseppe Verhelst, Werner | 287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Vardanyan, Anna Vargas, Ariel Vén, Ildikó Ver Donck, Luc Verhelst, Werner Verheven, Cees C.P.M. | 287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris Vargas, Ariel Vén, Ildikó Ver Donck, Luc Verdile, Giuseppe Verhelst, Werner Verheyen, Cees C.P.M Vetrov, Dmitry P. | 287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris Vargas, Ariel Vén, Ildikó Vén, Ildikó Ver Donck, Luc Verdile, Giuseppe Verhelst, Werner Verheyen, Cees C.P.M Vetrov, Dmitry P Vieira da Rocha. Heloísa | 287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris Vargas, Ariel Vén, Ildikó Ver Donck, Luc Verdile, Giuseppe Verhelst, Werner Verheyen, Cees C.P.M Vetrov, Dmitry P Vieira da Rocha, Heloísa | 287, 489 .287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Vardanyan, Anna Vargas, Ariel Vén, Ildikó Ver Donck, Luc Verdile, Giuseppe Verhelst, Werner Verheyen, Cees C.P.M Vetrov, Dmitry P Vieira da Rocha, Heloísa Vingerhoets, Ad | 287, 489 .287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vandewalle, Patrick Vardanyan, Anna Vardanyan, Anna Vargas, Ariel Verklas, Ariel Vén, Ildikó Ver Donck, Luc Verdile, Giuseppe Verhelst, Werner Verhelst, Werner Verheyen, Cees C.P.M Vetrov, Dmitry P Vieira da Rocha, Heloísa Vingerhoets, Ad Vinokhodova, Alla | 287, 489 30 59, 61 70 118 405 405 456 277 365 394 277 365 394 277 355 204 68 66 59, 61 70 59, 61 59, 61 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vandewalle, Patrick Vardanyan, Anna Varekamp, Chris Vargas, Ariel Vén, Ildikó Ver Donck, Luc Verhelst, Werner Verheyen, Cees C.P.M. Vetrov, Dmitry P. Vieira da Rocha, Heloísa Vingerhoets, Ad Vinokhodova, Alla Viranvi, Zsofia | 287, 489 .287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris Vargas, Ariel Vér, Ildikó Ver Donck, Luc Verdile, Giuseppe Verhelst, Werner Verheyen, Cees C.P.M. Vetrov, Dmitry P. Vieira da Rocha, Heloísa Vinciarelli, Alessandro Vinokhodova, Alla Vishnevskiv, Valera V. | 287, 489 |
| Tscheligi, Manfred | 287, 489 |
| Tscheligi, Manfred Tucci, Valter Tuijthof, Gabrielle J.M. Tuyls, Karl Valente, Fabio Valstar, Rineke Van den Noortgate, Wim Vandewalle, Patrick Vansteenwegen, Debora Vardanyan, Anna Varekamp, Chris Vargas, Ariel Vén, Ildikó Ver Donck, Luc Verdile, Giuseppe Verhelst, Werner Verheyen, Cees C.P.M. Vetrov, Dmitry P. Vieira da Rocha, Heloísa Vinciarelli, Alessandro Vingerhoets, Ad Vinokhodova, Alla Vishnevskiy, Valera V. Von Borries, Katinka Vos, Alexander G. | 287, 489 .287, 489 |
| Tscheligi, ManfredTucci, ValterTuijthof, Gabrielle J.M.Tuyls, KarlValente, FabioValstar, RinekeVan den Noortgate, WimVandewalle, PatrickVansteenwegen, DeboraVardanyan, AnnaVarekamp, ChrisVargas, ArielVén, IldikóVer Donck, LucVerdile, GiuseppeVerhelst, WernerVerheyen, Cees C.P.M.Viciar da Rocha, HeloísaVinciarelli, AlessandroVingerhoets, AdVinokhodova, AllaViranyi, ZsofiaVishnevskiy, Valera V.Von Borries, KatinkaVos, Alexander G.Vovnarovskava, Natalia | 287, 489 .287, 489 .30 59, 61 70 .118 .405 .456 .277 365 .394 .277 365 .394 373 155 204 68 66 59 |

| Vries, Han de | |
|-------------------------|--|
| Wal, Tamme van der | |
| Waldhör, Jürgen | |
| Wang, Jane-Ling | |
| Weber, Todd | |
| Wechsler, Beat | |
| Weerd, Heleen van de | |
| Weffers, Harold | |
| Weihmann, Frank | |
| Westerink, Joyce H.D.M. | |
| Wijk, René de | |
| Willems, Roland | |
| Wolf, Martin | |
| Wu, Haishan | |
| Yabe, Yoko | |
| Yang, Sunggu | |
| Yechiam, Eldad | |
| Yu, Libo | |
| Zaleska, Margaret | |
| Zandian, Modjtaba | |
| Zeilhofe, Hanns Ulrich | |
| Zhao, Guoying | |
| Ziherl, Jaka | |
| Zoratto, Francesca | |
| Zou, Sige | |