# Measuring User Behavior in a Complex USAR Team Evaluation

N.J.J.M. Smets[1], M.A. Neerincx[1,2], R. Looije[1]

[1]TNO, Soesterberg, The Netherlands. Nanja.Smets;Mark.Neerincx;Rosemarijn.Looije@tno.nl; [2]Man-Machine Interaction Group, Delft University of Technology, Delft, The Netherlands

## Introduction

Performing urban search and rescue (USAR) is a stressful and high demand task. The rescuers can use an extra pair of eyes and ears in the field, especially in places where they themselves cannot go. A ground robot can support rescuers in their task. To develop such a ground robot high fidelity and realism testing is needed [1]. High fidelity evaluations are difficult to set up, because the circumstances need to be complex and truthful. In this paper we performed a complex team experiment in the USAR domain. We measured a large number of variables to assess participants' performance, task load and emotion. This paper focuses on the question whether the analyses tools (Observer and FaceReader [2]) and measures gave an adequate indication of and insight in rescuers' operations for a robot with two levels of automation (no versus partial).

## Methods

*Participants*. Ten firemen participated, three in a pilot and one participant canceled, so that 6 participants in total completed the evaluation (5 male and 1 female , average age of 42). The mean number of years the participants had a driver's license was 24. Three participants had experience with operating robots. And three participants played (first person) computer games.

*Procedure.* Participants first had to read a general instruction about the experiment and then fill out some questionnaires. After which the participant was trained to use the robot (see Figure 1). Training was first performed in line-of sight and later out of line-of-sight. When the participant was confident enough and the instructors were satisfied with their performance, they received instructions about the main evaluation. The main evaluation consisted of unit tasks (different tests of different level of abstraction) [3] and the tunnel scenario. The tunnel scenario consisted of a reconnaissance with a team: Unmanned Ground Vehicle (UGV) operator, Unmanned Aerial Vehicle (UAV) operator and the mission commander. The participant performed the role of UGV operator. The tunnel scenario took 30 – 40 minutes, the whole evaluation took 3 hours. The evaluation was between subjects, they either used the UGV without autonomy (tele-operation) or with autonomy (such as waypoint and speech navigation).

*Task.* The participants controlled the ground robot to perform reconnaissance after a tunnel accident. The robot was deployed to gather more information about the situation inside. The participants specifically had to look for cars in the tunnel, the lay out of the situation, victims (where, how many and where) and look for fire and dangerous substances, depicted by pictures of warning signs.

Figure 1. Left is a picture of the UGV, in the middle is a picture of the tunnel scenario on the right a participant is tele-operating the UGV.


*Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)*
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

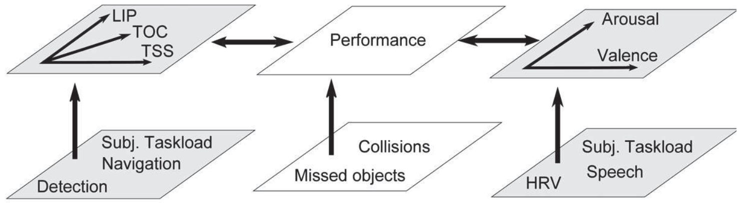



Figure 2. Relation of our measures and the task load (LIP Level of Information Processing; TOC Time Occupied; TSS, Task Set Switches), performance and emotion (arousal and valence) [reference].

*Measures*. We used the following measures, the relation between the measures and the cognitive task load, performance and emotional state are also depicted in Figure 2.

- Workload questionnaire. Every 2 minutes we asked the participant what his current workload was on a scale from 1 (none at all) to 5 (far too much). We also logged the reaction time to answer this question.
- Heart rate variability was measured with a belt around the chest.
- An observer rated the behavior of the participant. The behaviors that were rated are: communication with mission commander, situation report, tele-operating ground robot, operate robot by using speech commands, waypoint navigation (the last three behaviors are mutually exclusive). These behaviors were adapted from [4] to fit the available data. Novel is that also non mutually exclusive categories were used.
- Emotional state, the participant's face was recorded using a webcam and later analyzed using FaceReader.
- Performance, number of collisions: objective (observer) and subjective (by participant).

## Results

First we looked at the behavior of the participants during the scenario. See Figure 3 for an example of how the data can be visualized in Observer. Figure 2 shows that this participant communicated a lot with the mission commander, tele-operated the UGV by using the compass rose, speech (in the beginning of the evaluation) and touched the operator control unit a few times (the screen showing the different camera and laser images from the robot [5]). These views give us a good overview of the behavior of the participants in time.

Behavior analyses for the rated behaviors is shown in Table 1. The first three participants only used tele-operation to control the robot, this means no data is available for speech and waypoint navigation. Remarkable is that participant four spent a lot of time navigating by speech and less time communicating with the MC or using the compass rose. The other participants in the automated condition (participants 4, 5 and 6) communicated more with the MC than the participants in the manual condition (1,2 and 3), this can be related to the usage of the situation report which was less for the participants in the automated condition. When the manual navigation is heavily used (participants 5 and 6), they also collided more.

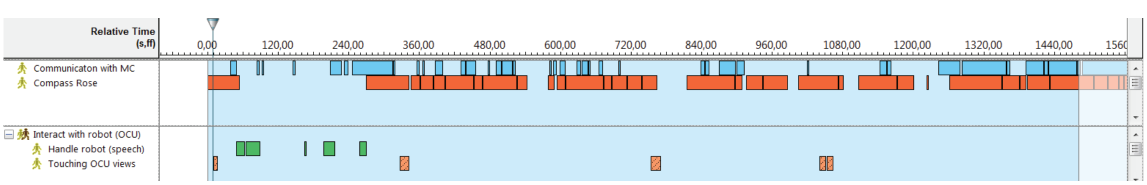



Figure 3. Screenshot of Observer with visualized rating data from the observers.

Table 1. The number and duration (seconds) of participants' behaviors (standardized data, range 0-1) and the number of UGV collisions (objective and subjective).

| Partic. | Communication with MC | Compass Rose | Speech navigation | Waypoint navigation | Mapping/ situation report | Performance # of collisions |
|---|---|---|---|---|---|---|
| 1 | 40, 0.52 | 41, 0.65 | - | - | 12, 0.23 | 10, 4 |
| 2 | 45, 0.33 | 50, 0.82 | - | - | 13, 0.12 | 7, 3 |
| 3 | 40, 0.17 | 53, 0.62 | - | - | 12, 0.15 | 5, 2 |
| 4 | 37, 0.45 | 17, 0.29 | 33, 0.31 | 3, 0.02 | 6, 0.07 | 6, 0 |
| 5 | 68, 0.51 | 87, 0.78 | 7, 0.08 | 3, 0.01 | 1, 0.01 | 19, Many |
| 6 | 68, 0.50 | 64, 0.73 | 0, 0 | 10, 0.08 | 0, 0 | 17, 4 |

We also asked the participant how much effort certain tasks took (see Table 2). The tasks related to the automated conditions took little effort and using the compass rose took more effort. We successfully collected heart rate variability data. Because of the number of participants we could not test for significance, but rest data may differ from event data (such as finding a victim), see Figure 4. Due to the limited number of participants and the realistic scenario that caused events to be presented in a different order for each participant we were not able to link experienced workload to the heart rate data. The emotional state of the participants during the tunnel scenario is shown in Table 3. The face recognition did not work optimal because of the lighting conditions and the participant did not always face the camera, for instance when talking to the mission commander. Expressions of participant 6 were well-recognized because of her make up.

## Conclusions

The detailed behaviors of the participants showed us what they were doing and for how long, thus giving us detailed understanding of human-robot teamwork and how certain behaviors interact with each other. The emotional data are interesting, but due to the nature of the task where the participant does not still in front of the camera (when talking to the MC) and the lighting conditions are not optimal, there were too much missing data. When the environmental conditions are better and with an eye tracker, we will probably get more useful data of the operator's emotion. With respect to heart variability data, the method of measuring should be improved. More extensive measuring in rest is needed, this way we expect a good baseline and better differences when events occur. The scenario needs to be improved, events should be more stressful to have a better effect on heart rate variability. For instance not only finding a victim, but also hearing or being able to interact with a victim.

Table 2. Effort on tasks ranging from 1 (no effort) to 5 (a lot of effort).

| Task | Effort |
|---|---|
| Managing the speed and direction of the robot using the compass rose | 3,3 |
| Interacting with people (such as the mission commander) during the scenario | 2,6 |
| Mapping/ situation report | 2.3 |
| Managing the speed and direction of the robot using voice commands | 2,0 |
| Using waypoint navigation | 2,0 |
| Using voice commands to direct the robot | 1,8 |

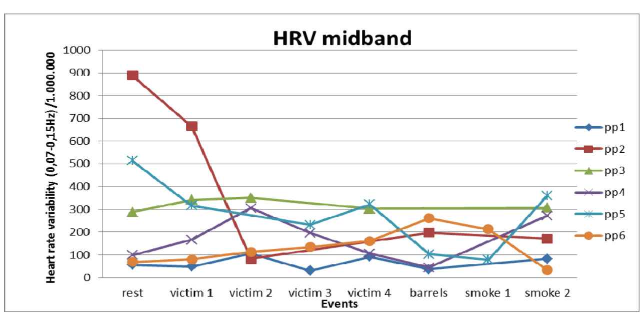
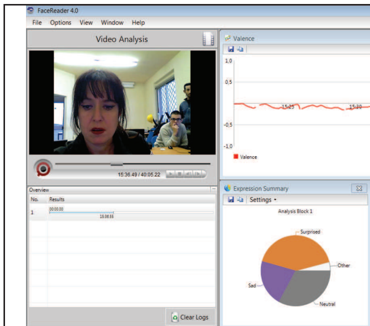



Figure 4. Heart rate variability of participants against the events in the scenario.

Table 3. Screenshot of FaceReader with mug shot of participant, valence chart with positive and negative dimensions and the momentary state in terms of Ekman's basic emotions (surprised, sad, neutral and other). The right part of table depicts the participant and the total number of times an emotion was recognized and how many seconds this emotion lasted in seconds during the tunnel scenario.




| Subject | Happy | Disgusted | Sad | Angry | Surprised | Neutral |
|---|---|---|---|---|---|---|
| 1 | 2<br>3.33s | 5<br>7.33s | 0 | 0 | 0 | 2<br>1.67s |
| 2 | 0 | 2<br>3.80s | 11<br>33.87s | 8<br>10.80s | 1<br>3.73s | 31<br>89.21s |
| 3 | 0 | 2<br>2.4s | 0 | 0 | 0 | 0 |
| 4 | 1<br>1.33s | 8<br>8.47s | 1<br>2.00 | 0 | 1<br>1.53s | 2<br>1.20s |
| 5 | 0 | 0 | 0 | 0 | 0 | 1<br>2.07s |
| 6 | 4<br>7.33s | 0 | 42<br>77.41s | 0 | 112<br>426.03 | 116<br>399.17s |

*Acknowledgements.* This paper describes research done under the EU-FP7 ICT247870 NIFTi project. For more about NIFTi, please visit http://www.nifti.eu.We would like to thank the fire fighters of FDDo (Dortmund, Germany) and of SFO (Montelibretti, Italy) for their support.

## References

1. Smets, N., Bradshaw, J., Diggelen van, J., Jonker, C., Neerincx, M., de Rijk, L., Senster, P., Sierhuis, M., ten Thije, J. (2010). Assessing human-agent teams for future space missions. *IEEE Intelligent systems* **25**(5), 46–53.

2. Noldus Information Technology. The Observer XT10, FaceReader 4.0. http://www.noldus.com/

3. Mioch, T., Smets, N.J.J.M., Neerincx, M.A. (submitted). Comparing Human-Robot Performances in Complex Situations by Means of Unit Task Tests. *Ro-Man. 21st IEEE International Symposium on Robot and Human Interactive Communication*, Paris, France, 9-13 September 2012.

4. Gómez, A.V. (2010). *Evolutionary design of human-robot interfaces for teaming humans and mobile robots in exploration missions*. Ph.D. dissertation, Universidad Politécnica de Madrid.

5. Larochelle, B. Kruijff, G. J. Smets, N. J. Mioch, T. and Groenewegen, P. (2011). Establishing human situation awareness using a multi-modal operator control unit in an urban search & rescue human-robot team. In *20th IEEE International Symposium on Robot and Human Interactive Communication. International Symposium on Robot and Human Interactive Communication* (RO-MAN-2011), Atlanta, GA, United States.