

# Using a Mental Workload Index as a Measure of Usability of a User Interface for Social Robotic Telepresence

Andrey Kiselev<sup>1</sup> and Amy Loutfi<sup>2</sup>

School of Science and Technology, Örebro University  
SE-701 82 Örebro, Sweden

<sup>1</sup>andrey.kiselev@oru.se

<sup>2</sup>amy.loutfi@oru.se

**Abstract**—This position paper reports on the use of mental workload analysis to measure the usability of a remote user’s interface in the context of social robotic telepresence. The paper discusses the importance of remote/pilot user’s interfaces for successful interaction and presents a study whereby a set of tools for evaluation are proposed. Preliminary experimental analysis is provided when evaluating a specific telepresence robot, called the Giraff.

## I. INTRODUCTION

Various interactions take place simultaneously when humans are communicating through a robotic telepresence system. These interactions include human-robot interaction between local user and communication device (i.e. remotely controlled robot), human-human interaction between two or more users, and not least human-computer interaction between the remote user and robot’s remote interface<sup>1</sup>. This paper is part of a study series which focuses particularly on the latter interaction and presents an ongoing project on design, implementation and evaluation of the user interface for the Giraff robotic telepresence system [8], [9], [11].

The approach taken in this work combines different techniques for measuring interface usability. Some methods used in this work are standard for evaluating usability of “office” applications, but others are normally used for drivers’ and pilots’ productivity assessment. The rationale for taking these measures into account comes from the fact that driving the robot is a secondary task for remote users while the primary is communication between remote and local users. Thus driving a robot should remain mentally and physically non-demanding. In this light, performance or quality of interaction will not be classified as a good result if the overall mental workload of the subject was also high.

## II. BACKGROUND

Giraff pilot’s interface allows remote users to establish connection with a Giraff robot, drive it and interact with local user(s) through the embedded video-conferencing system. Example of the driving screen can be found on Fig. 1.

Driving the robot can be done by using mouse, touchpad or

<sup>1</sup>By “local user” authors assume one who is physically located in the same environment with a robot. Thus, “remote user” is one who controls the robot from a remote location.

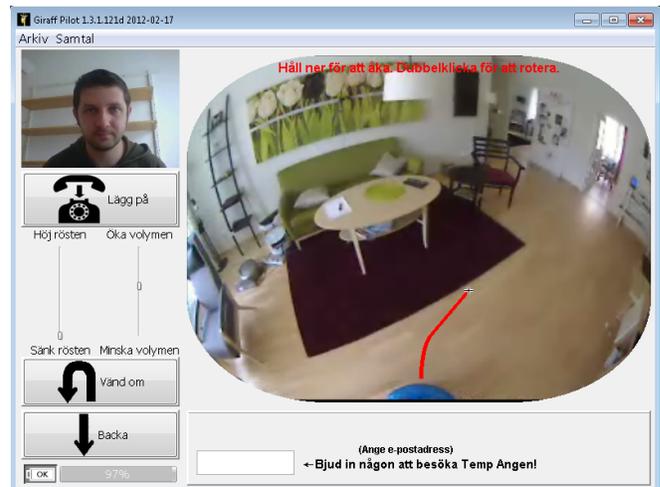


Fig. 1. A screenshot of the application’s main driving window.

any other standard pointing devices. The approximate tentative trajectory is drawn as a red line on a video panel. When left mouse button is pressed and held, the line transitions to green and the robot starts driving. The robot’s direction and speed are controlled by the orientation and length of the line respectively. The robot’s head tilt is controlled by dragging mouse pointer to the upper or lower parts of the video panel. Tilt can be adjusted at any point during driving.

It is also possible to drive the robot backwards by either using mouse or special button on the left panel. Alongside with it another button can be found which is used to rotate the robot 180 degrees counter-clockwise. Rotation can also be done by double-clicking a left mouse button on the left or right parts of the video panel. In this case the rotation angle is calculated according to the position of the mouse pointer on the panel. The left panel also contains controls (slide bars) for remote and local volume adjustment, call management button, battery information and local image.

The Giraff pilot’s interface is a combination of normal “office” application look-and-feel, which target users are supposed to be familiar with, and robot remote control functionalities. Thus a combination of usability assessment methods must be used in order to comprehensively evaluate

this interface. Particularly, in this work authors partially follow [1] in order to assess the Giraff pilot’s interface usability from the perspectives of efficiency and effectiveness of achieving goals and conduct mental workload analysis as a joint reflection of users’ satisfaction and performance.

### III. METHOD

The main application of the Giraff robot within the context of the ExCITE project [10] is serving as a movable communication device between elderly and remote visitors. The two tasks remote users usually perform are interacting with local user (e.g. elderly) and controlling robot’s behaviour.

Complete typical real interaction scenario can be divided into several stages and some typical pilots’ actions can be extracted. For instance, such typical actions include undocking, driving a robot (following a person, following a path), finding objects, docking. Some of the actions might be quite challenging depending on the pilots’ experience and technical limitations<sup>2</sup> of the platform.

One of the situations when pilots are usually faced difficulties is when they have to avoid collisions with some objects in the environment. This requires them to feel size of the robot and distance to obstacle. This problem comes from the mechanical design of the robot, which uses wide angle lens for capturing bigger scene. Another typical task which makes problems for pilots is connecting the robot to the docking station. Although the docking station is designed in such way that it is tolerant to some degree of robot misalignment, this still require precise controlling and good feeling of size and distance.

In the current experiment the performance of novice users’ in performing some typical tasks was measured. The measurements were done by analysing time spent by subjects to drive the robot between checkpoint (please, see the detailed description of the experiment in Section IV) and number of collisions made on each part of the path. The performance measurements were supplemented with the mental workload analysis, which was measured with the NASA TLX test [2], [3]. Although interactions with local users is a typical task for a Giraff system, pure driving performance is vital for pilots to successfully accomplish more sophisticated interaction tasks.

Additionally authors use a profiling questionnaire that collects demographical data such as age, gender, education, and usage experience with communication and electronic products (phone, computer, DVD, Skype, video games, cameras, and other). The education level was obtained according to the ISCED 2011 [4] in order to allow conducting further comparative experiments in other countries with different standards of education levels.

The questionnaire which is used in this experiment for UI

evaluation is the USE Questionnaire [5]. This questionnaire is selected among others (such as Computer System Usability Questionnaire [6]) because it allows to comprehensively evaluate the target system in terms of its usefulness, ease of use, ease of learning, and satisfaction. The Use Questionnaire is a widely used tool for UI evaluation and results can be easily correlated with other studies. One of the known issues with the USE Questionnaire (and with a number of others well known questionnaires) is that it suffers from “acquiescence bias” [7], which must be considered when analysing final results of the experiment. Not all the questions of the USE Questionnaire are applicable to the experiment. For instance, the “Usefulness” section cannot be considered as a valuable measure since subjects did not have any strong demand for using the Giraff in their daily life. The sixth question in the “Satisfaction” section (“I feel I need to have it.”) can not be used for final results for the same reason. All other questions are applicable to the experiment.

### IV. EXPERIMENT

The experiment was conducted in the “Ängen intelligent home” for elderly between 3-d and 4-th of May 2012.

Ten subjects participated in the experiment, six males and four females, average age is 40.7, SD 15.2. Subjects represent different user groups, have different exposure to technology, but none of them have prior experience with using Giraff pilot’s interface.

The 35-meter path was drawn in the apartment with bright blue dashed line with arrows. The path had several key points: docking station (DS), bedroom checkpoint (B), kitchen checkpoint (K), fridge checkpoint (F), goal (G). The scheme and a photo of the path can be found in Fig. 2. Subject start from the docking station, then they visit bedroom checkpoint and kitchen checkpoint. At the fridge checkpoint they have to read a task. The task for this experiment is to find a circle with number 1 inside somewhere on the floor in the living room. This is the goal checkpoint. Its position is the same for all subjects and its main role is to be a reference point for docking performance measurements.

The complete procedure of the experiment for each participant consists of several stages. First, each participant was shown a short introductory film about the Giraff system and pilot’s GUI. The total length of the film is 2:12 sec. Then each participant was given verbal instruction supplemented by a screenshot about how to drive Giraff, which controls should they use. After that the driving section began.

During the driving section each subject had to drive through all checkpoints until they reach the fridge checkpoint. There they had to read their task (“Go to 1”) written on the fridge. Then they had to find the G checkpoint in the living room and dock robot from that point back to the docking station. Driving sections were filmed for further analysis. At each part time to approach checkpoint and number of collisions

<sup>2</sup>By “technical limitations” we assume those that are derived from the robot’s design, such as robot’s physical dimensions or camera resolutions, and also those which come from environment. For instance, low video quality or temporal inoperability might be caused by poor internet connection between the robot and remote users.

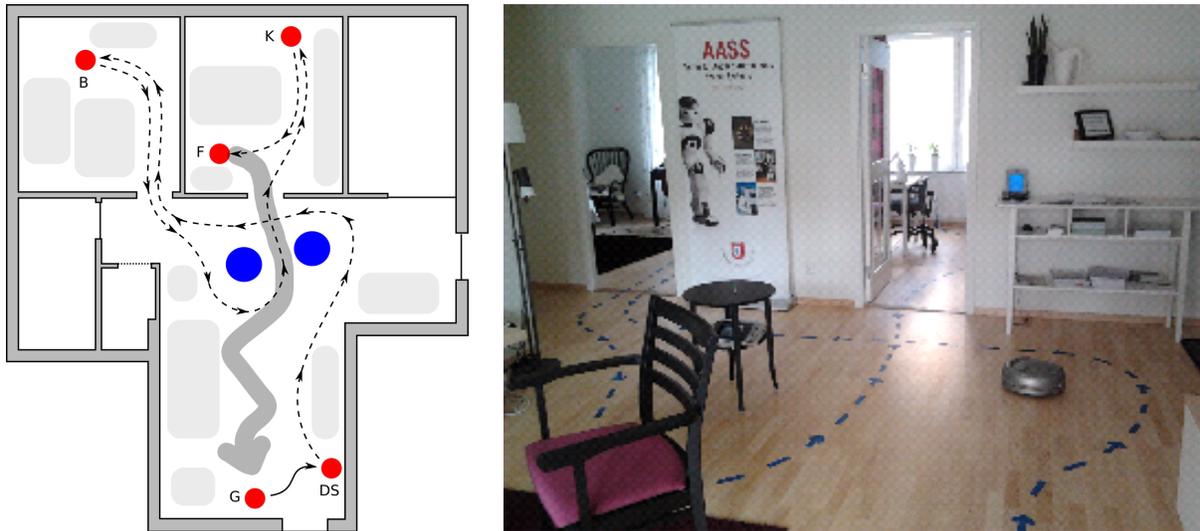


Fig. 2. *Left*: Outline of the Ängen apartment. Dashed line shows a path on the floor which subjects had to follow. Medium gray path - free driving which searching the object on the living room. Red circles - checkpoints: DS - docking station; B - bedroom checkpoint; K - kitchen checkpoint; F - fridge checkpoint; G - goal. Blue circles - artificial obstacles (coffee table, iRobot Roomba); light gray - other obstacles in the environment. *Right*: An example of the real environment.

were calculated.

At the final part of the experiment subject were asked to fill questionnaires: first NASA TLX, then profiling and finally the USE Questionnaire. All the questionnaires were administered through a web-page [12].

## V. RESULTS

### A. Performance

Results of the performance analysis along with NASA TLX score and average results of the USE questionnaire can be found in Tab. I<sup>3</sup>. Authors would like to refrain from providing any final conclusions based the results of this initial experiment.

### B. Observations and user reports

Observing user behaviours along with collecting user reports and opinions is an important step in UI evaluation. This subsection summarizes our findings, derived from video analysis and conversations with the participants. It is important to remember that these observations only derived from the current experiment, reported here. Correlating user reports across several studies is a subject for further investigation.

1) *Video resolution / quality*: It was clearly seen while setting up the experiment that when our task is written by pen or pencil it is simply can not be recognized by remote users. Authors had to use more contrast black marker and large font in order to make the task visible.

<sup>3</sup>The column "Confusion" shows whether or not subject was confused with undesirable tilt or moving-backwards robot's behaviours. TLX stands for the NASA TLX test score. USEQ stands for the average score of USE questionnaire (applicable questions only)

2) *Control over robots behaviour*: Two subjects, who have experience with computer games, reported that they would want to have more control over robot's behaviour and using keyboard seems to be more convenient for them. At the same time other participants reported that they are happy with current mouse-based control as it does not require any specific skills to control the robot.

3) *Pointing at objects of interest*: One of the most important observations shows that all subject tend to at least initially click at the point of interest (e.g. docking station or checkpoint) by mouse pointer when they start driving.

## VI. FUTURE WORKS

The main objectives of this initial experiment, which are a) to establish a general procedure for the Giraff pilot interface evaluation, and b) to provide a reference point for future interface evaluations, are achieved. The method provides useful information for interface refinement and will be used for further evaluations. For instance, it was clearly observed, that screen tilt functionality should be implemented in a different way. Nevertheless, the current user interface is easy to learn and use, which is clearly seen from the results of the USE questionnaire and supported by mental workload analysis and users' reports. Although the proposed method looks promising in principle, authors are interested in adding objective mental workload or either user satisfaction measurements into the current procedure if such measurement techniques are considered usable within the scope of the entire project. Also the USE questionnaire must be refined as well to overcome its known bias problem and applicability.

## ACKNOWLEDGMENT

This work has been done within the context of the ExCITE project which is supported by the EU under the

TABLE I  
PERFORMANCE MEASUREMENTS AND RESULTS OF THE NASA TLX.

Subject ID	Age	Gender	Performance, seconds						Collisions	Confusion	TLX	USEQ
			DS - B	B - K	K - F	F - G	G - DS	Overall				
1	21	male	86	92	54	49	60	341	1	No	37	95
2	20	male	100	86	38	43	19	286	0	No	16	86,8
3	27	female	81	74	20	126	36	337	5	Yes	55	64,1
4	50	male	211	99	34	138	25	507	3	Yes	61	84,9
5	42	male	123	70	27	51	28	299	2	No	48	94,2
6	47	male	74	74	30	58	37	273	2	No	47	73,1
7	67	female	183	127	55	62	168	595	4	Yes	62	78,5
8	55	female	108	88	124	44	30	394	1	Yes	21	59,3
9	35	male	137	90	32	66	53	378	1	No	49	78,2
10	43	female	114	110	35	79	72	410	1	Yes	68	64,9

Ambient Assisted Living Joint Programme (AAL-2009-2-125)

#### REFERENCES

- [1] ISO9241-11:1998(E) Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.
- [2] Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), 904-908. Human Factors and Ergonomics Society.
- [3] Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology*, 53(1), 61-86. Wiley Online Library.
- [4] International Standard Classification of Education 2011, UNESCO
- [5] Lund, A. M. (2001). Measuring Usability with the USE Questionnaire. *Usability Interface*, 8(2).
- [6] Lewis, J. R. (1995) IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7:1, 57-78.
- [7] Messick, Samuel; Jackson, Douglas N (1961). Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin*, Vol 58(4), Jul 1961, 299-304.
- [8] A. Kristoffersson, S. Coradeschi, K. Severinson Eklundh, and A. Loutfi (2011). Sense of Presence in a Robotic Telepresence Domain. *Universal Access in Human-Computer Interaction. Users Diversity*, pp. 479-487, 2011.
- [9] A. Kristoffersson, K. Severinson Eklundh, and A. Loutfi (2012). Measuring the Quality of Interaction in Mobile Robotic Telepresence: A Pilot's Perspective. *International Journal of Social Robotics*, 2012.
- [10] S. Coradeschi, A. Loutfi, A. Kristoffersson, S. Von Rump, A. Cesta, and G. Cortellessa. Towards a Methodology for Longitudinal Evaluation of Social Robotic Telepresence for Elderly. In *Proc. of Human-Robot Interaction Wksp. on Social Robotic Telepresence*, 2011.
- [11] A. Kristoffersson, S. Coradeschi, A. Loutfi, A. and K. Severinson-Eklundh (2011). An Exploratory Study of Health Professionals' Attitudes about Robotic Telepresence Technology. *Journal of Technology in Human Services*, 29:4, 263-283, Taylor & Francis.
- [12] Giraff Pilot User Interface Evaluation Questionnaire. Temporary located at: <http://hej-hej.tw1.ru/>