Towards Measuring Quality of Interaction in Mobile Robotic Telepresence using Sociometric Badges

Annica Kristoffersson 1, Silvia Coradeschi 1, Kerstin Severinson Eklundh 2, Amy Loutfi 1

1 AASS Center of Applied Autonomous Sensor Systems, Örebro University, Fakultetsgatan 1, 70182 Örebro, Sweden
2 School of Computer Science and Communication, KTH Royal Institute of Technology, 10044 Stockholm, Sweden

Keywords
Mobile Robotic Telepresence - MRP systems - F-formations - Spatial Formations - Presence - Telepresence - Sociometry - Quality of Interaction

1. Introduction

Mobile robotic telepresence (MRP) systems enable social interaction between two users who are separated by distance via a video conferencing system. MRP systems offer a “walking around” capability by combining teleoperation and telepresence. Typically, they are characterized by an LCD screen, a web camera, microphone and speakers mounted on a mobile robotic base. One example of a MRP system is Giraff, depicted in Fig. 1 (a). Application domains for MRP systems include elder care, hospitals, office environments and sick children attending classes. Currently, the field of MRP systems is rapidly expanding with an increasing number of commercial systems and research efforts [19]. To create MRP systems that are attractive to use, it is of interest to understand what promotes good interaction; both from the perspective of the pilot users who can maneuver a MRP system from distance and from the perspective of local users who interact with the person who is embodied in a MRP system. To exemplify, the pilot user’s difficulties in maneuvering the robotic unit may affect the social communication with the local user. To further complicate the picture, a range of different devices can be used to maneuver the robotic units. These include mobile phones, PCs and tablet computers.

Developers of MRP systems need to consider the domain in which they intend to deploy the robotic device, as well as needs of both local and pilot users. In this paper, the evaluations are performed with the Giraff MRP system, which targets the elder care domain. Elderly people may use MRP systems to meet virtually with a range of health care professionals without having to leave their homes. The health care professionals get a more nuanced picture of the elderly person’s living conditions as they can visit and move around in the elderly person’s home. However, many of these health-care professionals traditionally meet with their patients physically in a clinic environment, and many of them have never visited the elderly in their homes. The interaction between health care professionals and elderly that is enabled by the deployment of a MRP system is different from the social communication that takes place at the clinic. If MRP systems are to be successfully deployed in this domain, it is important to understand how the interaction is perceived and how it can be improved.

In an effort to develop an automatic means to measure quality of interaction between humans who communicate via a MRP system, this paper presents the results of a study in which we analyze the interaction between a researcher and 13 pilot users who work in health care, and who maneuvered the Giraff robot for the first time. Interactions occurred during showcasing of an apartment. Three kind of measures were used in the analysis: (1) subjective presence and ease of use questionnaires filled in after the experience, (2) a subjective assessment of spatial formations (how bodies orient themselves with respect to each other) and (3) sociometry, an objective measure of speech characteristics. The combined results of the study show that the quantitative analysis of the sociometric data correlates with a number of parameters gathered via qualitative analysis, e.g. different dimensions of presence and observed problems in maneuvering the robot.

In an effort to find more automated methods to assess the quality of interaction, the study in this paper used the same measures, with the addition of objective sociometric measures. Experimental results show that the quantitative analysis of the sociometric data correlates with a number of parameters gathered via qualitative analysis, e.g. different dimensions of presence and observed problems in maneuvering the robot.
upon which a methodology for measuring interaction quality in MRP systems can be obtained. This indicates that it may be possible to reduce the number of measures when assessing the quality of interaction. Experiments in the aforementioned application domains using additional MRP systems, a range of different devices for maneuvering the robots and a larger number of participants are needed to validate the results.

The article is organized as follows. Section 2 provides the reader with a background on the tools used to measure quality of interaction and previous work. Section 3 provides details on the experimental procedure, methods of data collection and the experimental expectations. Section 4 presents the results and discusses how they correspond to the experimental expectations. Section 5 summarizes the results and suggests future work.

2. Background

In this paper, three main tools (subjective presence questionnaires, subjective observations of spatial formations and objective sociometry) are used in an effort to develop an automatic means to measure quality of interaction when people communicate via MRP systems. These tools capture important characteristics of communication via MRP systems: social communication and mobility. Each of these tools have advantages and disadvantages [14], and it has been suggested that a promising direction for measuring presence, i.e., the sense of being there, is to aggregate a measure using both subjective and objective measures [13]. Following this thought, quality of interaction in mobile robotic telepresence may be an aggregated measure of the tools used in the experiment presented in this paper. To put the experiment in perspective to previous research, this section will provide theoretical details on the different tools used and results of two previous experiments with the Giraff MRP system [20, 21].

2.1. Presence

Presence is a multi-dimensional concept (para. 7 [7]). In literature a distinction is commonly made between social and spatial presence. When using presence questionnaires to evaluate the perceived presence in MRP systems, it was considered important to use questionnaires compatible with different media since the pilot users can use different devices to maneuver the robots. There are only a few validated questionnaires that were designed for cross-media compatibility, these include [3, 25, 26, 41]. However, only two of these were validated using video mediated communication systems [3, 26]. In our previous experiment [21], it was found that the Networked Minds Social Presence Inventory (Networked Minds) [3] and Temple Presence Inventory (TPI) [26] were suitable for use in a MRP system setting.

2.2. Spatial formations

In human-human interaction, there is a shared space in which the interaction occurs. In this space, we interact in different spatial configurations in terms of interpersonal distances [8], height and orientations [15]. How humans typically orient themselves with respect to one another was described by Adam Kendon using the F-formations system [15]. He distinguished three F-formations which have a shared interaction space, an O-space. The F-formations, which are depicted in Fig. 2 are: (1) vis-a-vis which is the spatial formation occurring when two individuals face each other while interacting, (2) L-shape which is when they stand perpendicularly to each other and discuss e.g. an object and (3) side-by-side, which occurs when they stand (or sit) beside each other facing in the same direction, for example while working on the same document in front of a computer. The spatial formations are created naturally while we interact, thus not much thought is put into it. The orientation of the lower part of the body has a dominant effect on the configuration of the spatial formation [15]. Physical aspects in an environment, e.g. interior design factors may discourage the creation of certain F-formations [29]. Further, embodied constraints can restrict what people can do and make some spatial configurations more probable than others [10]. For this reason, the form factors of the robot can impact the occurring spatial formations. To exemplify, the fact that the face of the pilot user is only provided in 2D can increase the need for a vis-a-vis formation. Elderly also reconfigured when not seeing the face of a person embodied in Giraff properly in [20]. This study is further detailed in Section 2.4.

2.3. Sociometry

Sociometry is an objective quantitative method to measure social relationships originating in the Latin ‘socius’ (social) and ‘metrum’ (measure). It was developed by the psychotherapist Jacob L. Moreno. He defined sociometry as “the mathematical study of psychological properties of populations, the experimental technique of and the results obtained by application of quantitative methods,” pp. 15-16 [29]. In 2004, Alex Pentland [35] constructed four measures that could possibly serve as predictive signals for social relationships: activity, engagement, emphasis and mirroring.

The activity in an interaction is the fraction of time for which a person is speaking [6, 16] and can be assessed by analyzing the speaking time. The activity can also be estimated by the amount of turn-takings between speakers in a conversation, where higher turn-taking indicates higher levels of activity [36]. How engaging an interaction is for the participants in an interaction can be estimated by assessing the number of overlaps or silent sequences in the interaction. If the participants are taking over each other, it is a sign of high engagement. In contrast, long pauses in an interaction signify a low engagement [6]. Curran and Pentland [6] defined emphasis as the variation in speech prosody (pitch and volume). A dynamic voice will have a higher emphasis score than one which is less dynamic. Further, high speech energy (the sum of the absolute speech signal’s amplitude over a time period) is an approximation of the excitement [11]. The level of mirroring is higher when there are many short interjections, for example “oh” or other single words such as “ok”. Mirroring can influence the smoothness in an interaction and mutual liking [4].

One means for performing objective automatic assessments of sociometric data uses sociometric badges from Sociometric Solutions [39]. These are wearable sensors worn around the neck. Included in the badge are a 3D accelerometer, two microphones, a bi-directional infrared transceiver and a bluetooth transceiver. The badge can sample locations in 3D, as well as recording the speech features of the carrier, whether another badge is in a face-to-face cone and whether other badges are within range. Although the claim is that the badge can be used to measure face-to-face interaction, there can be no guarantee that there is actually one occurring as the infrared transceiver used for the assessment is worn around the neck, while the wearer’s head can still be turned around. The sampled data stored in each badge can be automatically analyzed by running pre-defined scripts from within the computer software that accompanies the badges.

Publications using the badges cover: executive training [32], healthcare [30], management [40, 42, 43], organizational design and engineering [33, 34], real-time feedback [16, 17], science [36] and comparisons between media for video conferencing [38]. Sociometric solutions also points out a reference paper [31] that presents the develop-
opment of the badges and how they have been used to measure different metrics. A comparison of satellite-hub communications through different video conferencing systems was performed. In one system, a static display was used. The other systems utilized displays located at the hub which turned when the remote user moved either the mouse pointer or the head [38]. Six groups of three participants experienced interaction with a remote user through each of the three types of video conferencing systems. In total, five sociometric badges were used for each interaction, one for each participant, one for the system itself and one for the remote user. The participants felt more engaged in the conversation and responded more accurately to deictic prompts while interacting through a system with a turn-able display. The remote user also ranked the interactions through the non-static interfaces as better for communication. Further, data from the badges indicated that the conversation was more active and that the speakers were more attentive, engaged and excited while interacting through the systems with turn-able displays. However, participants perceived a lower sense of the remote user was “present” in conditions with turn-able displays than in the static display condition.

2.4. Prior experiments on quality of interaction using Giraff

A handful of works on human-robot interaction (HRI) deal directly with Kendon’s F-Formations (e.g. [12, 20, 21, 23, 37, 44]). In our previous studies [20, 21] we have studied how the interaction is perceived from the perspectives of pilot users and elderly local users (we insist on adopting the term pilot user in our review paper [19]) experiencing the Giraff system for the first time. In [21], we studied what spatial formations the novice pilot users used in relation to an actor (person) in four different steps of a scripted scenario. Two of the steps were movements between different rooms, thus, we introduced two more spatial formations into our concept of spatial formations; ahead and follow. Further, it was found that some of the pilot users’ behaviour deviated from our expectations which were based on what Kendon had observed as being the natural behaviour in a human-human interaction. They did not form the vis-a-vis F-formation while discussing a medical issue with the actor who had an embodied constraint by being confined to a wheelchair. Instead, they oriented the robot towards a wall as depicted in Fig. 3. We call this a look-away formation. There may be different reasons for this behaviour, for example that the pilot user perceived the local user’s environment via Giraff’s wide-angle lens and that the participants were having problems turning the robot. Accordingly, there were six different spatial formations occurring in the experiment, three of them were Kendon F-formations, but three other formations, all lacking an O-space, also occurred: ahead, follow and look-away, see Fig. 4.

The spatial formations occurring between the pilot user and the actor were correlated with the perceived presence in several dimensions and the perceived ease of use. The presence questionnaire was a Swedish translation based on Networked Minds [3] and TPI [26]. Correlating the spatial formations that occurred while discussing a medical matter with the perceived ease of use, it was found that the pilot users who were observed doing the expected vis-a-vis F-formation were more likely to respond that it was easier to undock the robot and to make u-turns with it. This is supported by the observation that some pilot users...
who were observed using a side-by-side F-formation when finding an object instead of the expected F-formation L-shape, were seen to have difficulties maneuvering the robot. The pilot users observed doing the side-by-side formation also experienced a lower Comprehension (Comprehension is a dimension in Networked Minds that assesses the ability to understand the other’s intentions, thoughts etc. [3]). Overall, we found that the presence questionnaire was suitable for use in a MRP systems setting. However, we saw a need to improve the dimensions Attentional engagement and Co-presence since the answer to several questions had to be excluded in the analysis due to the participants misinterpreting the questions [21]. It should be noted that also Adalgeirsson and Brezeal (2010) [1] found that the Networked Minds dimension “Co-presence” included ambiguous questions that caused confusion.

Following up on [21], an experiment was performed from the perspective of elderly local users. The main objective was to assess whether the use of spatial formations that deviated from the accepted F-formations, for example the look-away formation observed in [21], would be perceived as acceptable by elderly people. In the subsequent experiment, [20], ten elderly people were invited to attend a guided tour in a showcase apartment. After the tour, the elderly participants watched video recordings of their own interaction with a guide embodied in Giraff and responded to questions regarding the interaction. It was found that there was a mis-match in between how the pilot users oriented the robot in [21] and the spatial formations that were acceptable for the elderly. For example, it was important that the robot was oriented towards them in a vis-a-vis F-formation while interacting. However, the importance of interacting in this configuration decreased during the interaction. The combined results indicate that it is important to consider the perspectives of both the pilot user and local user while interpreting quality of interaction measures of MRP systems.

Although [21] demonstrated that questionnaires assessing ease of use and presence can be used along with observations of F-formations when measuring the quality of interaction and that a vis-a-vis F-formation was considered important among elderly local users, we saw a need for more automatized methods to measure the quality of the interaction in order to support the pilot user in creating the best conditions for interaction. Therefore, a new experiment was designed in which the sociometry was also used as an objective tool to measure quality of interaction. The experimental procedures and our expectations will be described in the next section.

### 3. Experimental Procedures and Expectations

Thirteen participants were invited to a session in which they would use Giraff for the first time. There were nine women and four men with ages ranging from 20 to 58 years. The average age was \( \mu=39.55 \) years, \( \sigma=11.80 \). Table 1 shows how experienced the participants were with different technologies. Seven of them used video communication tools, and seven of them played computer/video games. Only the no answer group was the same for both questions. The participants were all working at an elderly residence in Örebro, Sweden. Assistant nurses, nurses and occupational therapists were among the participants. The laptop used in the experiment, equipped with a mouse and a headset, was situated in an apartment located near to the “Ången” smart home environment, in which the Giraff robot was situated. Ången is open to public visits upon request and contains a rich set of robots and other assistive tools that can be used in real homes. In this section, we will provide information about Giraff, the experimental procedures, the type of data collected and our expected results.

#### 3.1. Giraff

Giraff is a 163 cm tall MRP system that consists of a mobile robotic base on which an LCD screen, a web camera, speakers and a microphone are mounted on a pole. The video conferencing system used is the proprietary software, VSee. The camera uses a wide-angle lens that provides the pilot user with a fish-eye view of the robot's environment. The resolution is 320x240 pixels. The field of view is 100° vertically and 160° horizontally. The client interface can be installed on any PC and allows the pilot users of the system to move Giraff around in a local user's environment. The pilot steers the robot by pointing a cursor directly on the live video-feed and pressing the left mouse button. The robot’s speed corresponds to the length of a line that the user draws on the image, the longer the line the higher the speed. Fig. 1 (a) shows the current design of the Giraff robot and Fig. 1 (b) shows a snapshot of the version 1.4 client interface [21].

#### 3.2. Experimental Procedure

The experiment began with each participant watching a short (about 1.5 mins) instruction movie on how to maneuver Giraff. The participant was then asked to fill in a consent form and a sociodemographic

<table>
<thead>
<tr>
<th>Video communication</th>
<th>Computer/Video games</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>7</td>
</tr>
<tr>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>No answer</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1. The experience of video communication tools and/or computer/video gaming among the participants

---

1 The video is available at: [https://vimeo.com/66729107](https://vimeo.com/66729107)
form. After having completed the forms, the participant hung the sociometric badge around his own neck and then connected to Giraff in Ängen. When the participant had connected to Giraff, a researcher started to interact with the participant. She guided the participant around the apartment following a semi-scripted procedure described below. Fig. 4 provides a supporting graphical overview of the scenario.

1. The participant (P) connects to the Giraff robot and undocks it. The researcher (R) says “Hi, I am seated on the sofa” with the expectation that P will move Giraff towards R. Some guidance in driving is given if necessary.

2. In situation S1, R welcomes P and tells P that P will be guided around in the apartment. R informs P about that the research focus is on the communication through Giraff and asks P to interrupt and talk as much as P wants to. R repeats the commands that P can use to maneuver the robot and allows P to try them. R also talks about the paintings (O1) that improve the room’s acoustics.

3. R asks P to follow to the kitchen. The transition is called T1.

4. R stops in front of several objects on a table. R talks particularly about O2 and how it can be used but also about a few other devices available on the table. The situation is called S2.

5. R asks P to follow to the bedroom. The transition is called T2.

6. In the bedroom, R talks about an object on the bed (O3) and the lamp (O4) that lightens up automatically when a person rises from the bed. The situation is called S3.

7. R says that the Giraff driving experience will soon be over but that Giraff has to be docked into the docking station (DS) again before saying good bye. R wonders if P remembers where the docking station is. P is expected to take the lead in transition T3.

8. R guides P to the DS and they say good bye to each other before the call is ended.

As should be clear to the reader, S1-S3 are situations in which the interaction is “static” while the transitions T1-T3 are interactions occurring while the pilot users are maneuvering the robot between different locations. The objects referenced in the script are denoted O1-O4. After having completed the scripted scenario, the participant was asked to fill a questionnaire regarding the Giraff pilot application and perceived presence. Both questionnaires are further described below.

To summarize our experimental procedure, the participants were seated in an office, saw an instruction movie and, filled in consent and sociodemographics forms. They hung the sociometric badge around their neck and connected to Giraff. They used the robot to follow a researcher around in an apartment before being asked to go back to the docking station. Finally they filled in a questionnaire. In total, the procedure took up to an hour. So, while the number of participants were low, the duration of each interaction was long.

2 A deliberate choice was made to hang the badge around the neck of the participant rather than on Giraff. This setup ensured sampling the analogue voice rather than a transformed digital one. The researcher also wore a badge around her neck.

Figure 4. A graphical overview of the scripted scenario used in the experiment. C1-C3 are locations of cameras. S1-S3 are situations. T1-T3 are spatial transitions. DS is the docking station. The researcher (R) is denoted by a vertical rectangle and the participant (P), that is embodied in Giraff, is denoted by a circle. The arrows denote how R and P were expected to be spatially oriented in the situations S1-S3.

3.3. Data collection

Three surveillance cameras configured to capture most situations and transitions in the scripted scenario were used in the experiment. In Fig. 4, they are denoted C1-C3. The experimental scenario in this study was almost identical to the one used in [20] in which a researcher guided elderly people around in the Ängen showcase apartment while the researcher was embodied in Giraff. In the experiment, it was found that it was important for elderly local users to be able to see who they were interacting with. Further, during the interviews, complaints were raised that the pilot user was not always fully visible on the Giraff screen. It had also been found in [21] that there were correlations between spatial formations and perceived ease of use and also between spatial formations and perceived presence. Therefore, we were interested in the spatial formations occurring during T1-T3 and S2-S3 and in understanding how many pilot users who were fully visible on the Giraff screen.

Two different questionnaires were used in the experiment. The first questionnaire was a sociodemographic form. The second was more substantial and contained questions divided into eleven dimensions of presence based on Networked Minds [3] and TPI [26] as well as a section on the perceived ease of use. The questionnaire was to a large degree the same as the questionnaire used in [21] but with scale modifications for a few of the questions in the Networked Minds dimensions “Attentional Engagement” and “Co-presence” [2, 3]. The changes were based on misunderstandings that had occurred in the previous experiment [21] and in [1]. The revised dimensions are presented in Appendix A. Each of the remaining presence dimensions in the questionnaire, ex-
cept for the TPI dimension “Social Richness” included several questions that were to be answered on a Likert scale 1-7 where 7 = To a very high degree. Social Richness was assessed by asking the participants how they rated their experience, choosing from opposite pairs of words (e.g. insensitive or sensitive) on a semantic differential scale 1-7. The dimensions “Object Realism” and “Person Realism” originating from the TPI dimension “Perceptual Realism” [26, 27] are further explained in Appendix A.

The experimental setup included sociometric badges worn around the necks of the participant and the researcher. Although the badges contain a set of sensors and two microphones, the features used in our experiment were the microphones located on the back of each of the badges and one of the badge’s accelerometers. The back microphone sampled speech features at a frequency of 1 Hz. The sampled audio data for any period of real time can be automatically analyzed and be presented in terms of e.g. audio amplitude, current speaker at time unit t, turn takings and so on via the computer software Sociometric DataLab that comes with the badges. The majority of the speech features used in our analysis were inspired by [38] which compared the efficiency of static and movable screens in a satellite-hub interaction. Sirkin et al. found differences in the speech features (speaking time, speech energy, speech segment length and number of turns per second) in the configurations used in the experiment. Although our experiment was a one-to-one interaction via a MRP system and involved transitions in real space, we found Sirkin’s study inspiring. In an attempt to assess activity, engagement, excitement and mirroring, we chose to measure several speech features. Speaking time, speech energy, speech segment length, number of turns per second, silent segment length and moments of overlaps in speech were all recorded. These features are defined as follows:

**Speech energy** The sum of the absolute speech signal’s amplitude over a time period for the two or more people interacting.

**Number of turns per second** The number of turns over the time period prohibiting successor overlap. That is, no turn is counted if the speaker is interrupted by the other person, but continues talking.

**Speaking time** The fraction of time that each person is speaking.

**Speech segment length** The total number of speech segments and their maximum length, minimum length, average length and standard deviation. Number of segments of length one, two, three, four, five or six seconds, and those longer than six seconds were noted. The data are calculated per person involved in the interaction. For comparison reasons, the number of segments of different lengths is divided by the duration of each of the situations S2-S3 and transitions T1-T3.

**Silent segment length** Similar to speech segment length. For comparison reasons, the number of segments of different lengths are divided by the duration of each of the situations S2-S3 and transitions T1-T3.

**Overlap** Number of seconds where both people are speaking simultaneously. For comparison reasons the time of overlap is divided by the duration of the interaction.

### 3.4 Experimental expectations

Experienced pilot users of the Texai MRP system in office environments have reported that it is difficult to focus on social interaction when moving [24]. It has also been shown that pilot users find the docking of Giraff a difficult procedure [8, 18]. The procedure is non-automatic and the pilot user is required to drive Giraff straight into the docking station in a proper speed. Accordingly, it was expected that interaction would be more active during T3 especially in terms of short speech segments while orienting the Giraff towards the docking station. It was thus expected that the speech energy, speaking time and number of turns per second in the interaction would be higher during the situations S2-S3 and transition T3 than during T1 and T2.

Due to the fact that it was the participants’ first experience of driving Giraff and the fact that the task was considerably different from the participants’ ordinary work tasks, we had no expectations around their engagement in the interaction. We rather expected that those who had more trouble driving the Giraff would need more assistance during the transitions. In other words, we expected that there would be a higher amount of speech energy, speaking time and number of turns per second in the interaction during transitions for the participants having trouble. Further, we expected that the sociometric data would correlate with how spatially and socially present the participants felt during the experiment. For example, it could be imagined that the participants needing a lot of assistance during the transitions T1-T3 would experience the robot’s environment as being more distant (less spatially present) than the participants not needing much help. Participants with problems with maneuvering the robot could for example bump in to obstacles, something they would not do if walking around in a real environment. Following the same thought, we expected that the sociometric measures: speech energy, speaking time and number of turns per second in the interaction during the static situations S2 and S3 would correlate with how socially present the participants felt. This could be expected since speaking time and number of turns per second are both measures of the activity [6, 36] in the interaction while speech energy is said to be an estimate of the excitement [11] all of which are predictive signals of communication between humans.

Similarly in the findings of our previous experiment [21], we expected that there would be relations between formations and perceived presence as well as between the formations and perceived ease of use. To summarize, we expected that:

- **Exp1** The interaction would be more active in terms of amount of speech energy, speaking time and turns during S2, S3 and T3.
- **Exp2** Problems in maneuvering the robot would be reflected in the sociometric data collected during transitions.
- **Exp3** The perceived spatial presence would be reflected in the sociometric data collected during transitions T1-T3.
- **Exp4** The perceived social presence would be reflected in the sociometric data collected during static situations S2-S3.
- **Exp5** Relations between chosen formations and perceived presence would exist.
- **Exp6** Relations between chosen formations and perceived ease of use would exist.

A key difference from our previous study [21] is that some of the participants in the present study had previous experience of interaction via video communication tools. Further, a few participants played computer/video games, a factor we did not include in the questionnaire used in [21]. Having experience of any of these tools is a possible signal for a higher computer experience in general. On this premise, we expected that participants who used video communication tools would be more comfortable with social communication via a MRP system and...
that this would be possible to assess via the perceived presence and
the sociometric measures. On the same premise, we expected that
the participants who played computer/video games would need to focus
less on the navigational aspects of the client interface, and would in-
stead focus on the communication with the researcher. Following our
assumption that the participants with experience of any or both of
the technological tools had a higher computer experience, we expected
that they would perceive the system as being easier to use than those
who had never used either technique previously. Accordingly, we expected that:

Exp7 Relations between experience of video communication tools
(and/or computer/video games) and sociometric measures
would exist.

Exp8 Relations between experience of video communication tools
(and/or computer/video games) and perceived presence would
exist.

Exp9 Relations between experience of video communication tools
(and/or computer/video games) and perceived ease of use
would exist.

4. Results

For each participant, the video recordings from the three different rooms
were analyzed in several steps. In the first iteration, a researcher made
the following annotations:

1. Annotation of time codes for the different occasions in the sce-
   nario in relation to the start of the video recording for each of
   the cameras C1-C3.

2. Annotation of the time codes in annotation 1 expressed in real
time (hh:mm:ss).

3. Annotation of the time codes in relation to when the participant
undocked the Giraff measured in seconds and the time spent in
each situation/transitions.

Further, the video recordings from C2 were used to assess whether or
not the participants were sitting right in front of the camera while driving
Giraff. Using the videos, a note was made about how well the partic-
ipants could be seen by the researcher, only eight of the participants
were seated directly in front of the laptop camera and therefore fully
visible to the local user on Giraff's LCD screen.

The duration of each situation is summarized in Table 2. There were
large variances in how much time participants spent in each of the situa-
tions S1-S3 and transitions T1-T3. However, it should be noted that
the participants were all novice users and thus required different amounts
of time to understand how to for example maneuver the robot. To ex-
emptly, some participants managed to move directly to the researcher
after the conversation started. Others needed more assistance in ma-
neuvring the robot towards her. Even though the researcher had re-
quested the participants to speak and interrupt, humans are more or
less talkative by nature. This could explain the varying duration of S2
and S3.

4.1. Choice of formations

Having completed the first iteration of the analysis in which the dura-
tion of each occasion in the experimental scenario was assessed, the
videos were analyzed again. This time, the spatial formations in situa-
tions S2-S3 and transitions T1-T3 were evaluated. The occurring spa-
tial formations were annotated according to those depicted in Fig. 2
with the addition of the look-away formation (see Fig. 3) which had
been observed in our previous study [21]. When determining the spa-
tial formation, the orientation of the lower body of the researcher and
the robot's screen were used. To validate the process of annotating the
occurring spatial formations, a set of reference snap shots depicting
each of the spatial formations were chosen from the movies. Fig. 5
provides a graphical summary of the different spatial formations while
Fig. 6 and 7 show the reference snap shots. A subset of the videos was
annotated by several researchers independently with a high agreement
on the annotations. In the previous study [21], the local user had an
embodied constraint in being confined to a wheelchair. In the current
experiment, the researcher could easily reconfigure the spatial forma-
tion depending on how the participants positioned themselves in S2
and S3. The fact that she could rotate the lower body easily led to
more frequent changes of the spatial configuration than in the previ-
ous experiment [21]. Thus, typically more than one spatial formation
occurred in each of the situations S2 and S3.

It was observed that 11 out of 13 participants followed the researcher
in transition T1. However, the researcher always entered the kitchen
before the participant. This suggests that the two participants who
showed a deviating behaviour (moving ahead of the researcher) may
have had problems with driving Giraff. All participants followed the
researcher in transition T2 and 8 out of 13 participants went ahead
of the researcher in transition T3.

In Table 2, the spatial formations that occurred in S2 are listed for
each of the participants. Ten participants interacted in the L-shape F-
formation during at least 90% of the duration of S2 while two interacted
mainly in the vis-a-vis F-formation. Worth noting here is that due to
one of the participants being shown different objects in the room, there
was no annotation done regarding occurring spatial formations. The
objects shown were positioned in such a way that the researcher was
out of range of camera C2.

The spatial formations that occurred in S3 are listed for each of the
participants in Table 4. All but one participant interacted in the vis-a-
vis F-formation during at least 90% of the duration of S3.

<table>
<thead>
<tr>
<th>Situation</th>
<th>μ (s)</th>
<th>σ (s)</th>
<th>min (s)</th>
<th>max (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start conversation - S1</td>
<td>45.31</td>
<td>24.41</td>
<td>14</td>
<td>114</td>
</tr>
<tr>
<td>S1</td>
<td>292.54</td>
<td>159.46</td>
<td>158</td>
<td>772</td>
</tr>
<tr>
<td>T1</td>
<td>54.92</td>
<td>50.22</td>
<td>11</td>
<td>198</td>
</tr>
<tr>
<td>S2*</td>
<td>217.83</td>
<td>78.29</td>
<td>100</td>
<td>342</td>
</tr>
<tr>
<td>T2*</td>
<td>94.58</td>
<td>109.07</td>
<td>30</td>
<td>424</td>
</tr>
<tr>
<td>S3</td>
<td>151.69</td>
<td>44.28</td>
<td>97</td>
<td>219</td>
</tr>
<tr>
<td>T3</td>
<td>133.15</td>
<td>55.04</td>
<td>71</td>
<td>282</td>
</tr>
<tr>
<td>Docked - Ended conversation</td>
<td>28.00</td>
<td>14.83</td>
<td>9</td>
<td>54</td>
</tr>
<tr>
<td>Total*</td>
<td>1032.75</td>
<td>335.12</td>
<td>757</td>
<td>1836</td>
</tr>
</tbody>
</table>

Table 2. A summary of the duration of each of the situations and transitions
in the scripted scenario. The lines marked with a * exclude one partici-

Table 3. The spatial formations that occurred in S2 are listed for
each of the participants. Ten participants interacted in the L-shape F-
formation during at least 90% of the duration of S2 while two interacted
mainly in the vis-a-vis F-formation. Worth noting here is that due to
one of the participants being shown different objects in the room, there
was no annotation done regarding occurring spatial formations. The
objects shown were positioned in such a way that the researcher was
out of range of camera C2.

The spatial formations that occurred in S3 are listed for each of the
participants in Table 4. All but one participant interacted in the vis-a-
vis F-formation during at least 90% of the duration of S3.
Figure 5. The formations observed in the experiment. Formation 1-3 are F-formations: 1. Vis-a-vis, 2. L-shape, 3. Side-by-side, 4. Look-away, 5. Ahead and 6. Follow. The Giraff is grey and the local user is black. The dotted ovals denote the shared interaction space, O-space.

Table 3. The spatial formations that occurred in situation S2. The symbols indicate how much time was spent in each formation: -. 0-10%, -. 11-40%, 0. 41-59%, +. 60-89%, ++. 90-100%. 1. Vis-a-vis, 2. L-shape, 3. Side-by-side, 4. Look-away, 5. Ahead and 6. Follow. The participant marked with an * experienced a different scenario in S2 not captured by C2.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10*</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>--</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2. Sociometric measures

The experiment was made in a dynamic environment with novice pilot users of the system. In the previous experiment [21] in which users maneuvered Giraff while interacting with a local user confined to a wheelchair, we observed that they were focused on the task of maneuvering the robot rather than on communicating while performing transition tasks. Thus there was an interest in analyzing the speech features sampled during transitions and static situations separately. By doing this division, it was possible to manually calculate the speech features during each of the situations presented in the scripted scenario: start conversation - start S1, S1, T1, S2, T2, S3, T3 and docked - conversation ended. However, only the speech features sampled during transitions T1-T3 and situations S2-S3 as well as the entire interaction from start to end of the conversation were calculated. For each of the situations and transitions; the speech energy, number of turns per second, speaking time and speech segment length were calculated. For the whole scripted scenario; all of the a priori mentioned data but also silent segment length and overlap were calculated. For more information about the different measures, see Section 3.3.

Table 4. The spatial formations that occurred in situation S3. The symbols indicate how much time was spent in each formation: -. 0-10%, -. 11-40%, 0. 41-59%, +. 60-89%, ++. 90-100%. 1. Vis-a-vis, 2. L-shape, 3. Side-by-side, 4. Look-away, 5. Ahead and 6. Follow.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10*</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. The average, minimum and maximum speech energy between the participants and the researcher during T1-T3 and S2-S3.

<table>
<thead>
<tr>
<th>Situation</th>
<th>μ</th>
<th>σ</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0.0207</td>
<td>0.0092</td>
<td>0.0102</td>
<td>0.0455</td>
</tr>
<tr>
<td>S2</td>
<td>0.0210</td>
<td>0.0053</td>
<td>0.0151</td>
<td>0.0373</td>
</tr>
<tr>
<td>T2</td>
<td>0.0232</td>
<td>0.0112</td>
<td>0.0118</td>
<td>0.0527</td>
</tr>
<tr>
<td>S3</td>
<td>0.0220</td>
<td>0.0058</td>
<td>0.0156</td>
<td>0.0387</td>
</tr>
<tr>
<td>T3</td>
<td>0.0205</td>
<td>0.0041</td>
<td>0.0132</td>
<td>0.0259</td>
</tr>
</tbody>
</table>

The Sociometric Datalab Software interprets a 16-bit PCM (Pulse-code modulation) signal that can represent ~32.8k positive and ~32.8k negative values. The program exports the absolute value of the sound pressure on a -1 to +1 scale.

3 The Sociometric Datalab Software interprets a 16-bit PCM (Pulse-code modulation) signal that can represent ~32.8k positive and ~32.8k negative values. The program exports the absolute value of the sound pressure on a -1 to +1 scale.
Figure 6. Spatial Formations in kitchen (S2).

Figure 7. Spatial Formations in bedroom (S3).

Table 6. The average, minimum and maximum number of turns per second between the participants and the researcher during T1-T3 and S2-S3.

<table>
<thead>
<tr>
<th>Situation</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0.14</td>
<td>0.08</td>
<td>0</td>
<td>0.27</td>
</tr>
<tr>
<td>S2</td>
<td>0.16</td>
<td>0.03</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>T2</td>
<td>0.15</td>
<td>0.08</td>
<td>0.05</td>
<td>0.37</td>
</tr>
<tr>
<td>S3</td>
<td>0.17</td>
<td>0.03</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>T3</td>
<td>0.15</td>
<td>0.04</td>
<td>0.04</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Table 7. The average speaking time (fraction of time) for the researcher (R) and the participants (P) during T1-T3 and S2-S3. The number is given as a percentage of the total time in each situation.

<table>
<thead>
<tr>
<th>Situation</th>
<th>$\mu_R$</th>
<th>$\mu_P$</th>
<th>$\sigma_R$</th>
<th>$\sigma_P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0.33</td>
<td>0.25</td>
<td>0.15</td>
<td>0.20</td>
</tr>
<tr>
<td>S2</td>
<td>0.32</td>
<td>0.22</td>
<td>0.04</td>
<td>0.08</td>
</tr>
<tr>
<td>T2</td>
<td>0.30</td>
<td>0.24</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>S3</td>
<td>0.35</td>
<td>0.23</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>T3</td>
<td>0.31</td>
<td>0.23</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Table 8. A summary of the response regarding the perceived ease of use of the system. 1 = very difficult, 7 = very easy.

<table>
<thead>
<tr>
<th>How was it to</th>
<th>μ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>start Giraff Application</td>
<td>6.46</td>
<td>0.63</td>
</tr>
<tr>
<td>connect to the Giraff</td>
<td>6.38</td>
<td>0.62</td>
</tr>
<tr>
<td>undock</td>
<td>6.00</td>
<td>1.18</td>
</tr>
<tr>
<td>make a u turn</td>
<td>5.85</td>
<td>1.10</td>
</tr>
<tr>
<td>find the person you met</td>
<td>5.77</td>
<td>0.89</td>
</tr>
<tr>
<td>stop</td>
<td>6.23</td>
<td>0.97</td>
</tr>
<tr>
<td>go backwards</td>
<td>5.75</td>
<td>1.01</td>
</tr>
<tr>
<td>look upwards</td>
<td>5.69</td>
<td>1.43</td>
</tr>
<tr>
<td>follow the person you met</td>
<td>5.62</td>
<td>1.15</td>
</tr>
<tr>
<td>go back to docking station</td>
<td>4.85</td>
<td>1.17</td>
</tr>
<tr>
<td>know you were docked</td>
<td>5.77</td>
<td>1.42</td>
</tr>
<tr>
<td>finish the call</td>
<td>6.42</td>
<td>0.76</td>
</tr>
<tr>
<td>hear what the person said</td>
<td>4.92</td>
<td>1.73</td>
</tr>
<tr>
<td>see the person you met</td>
<td>5.62</td>
<td>1.50</td>
</tr>
</tbody>
</table>

Table 9. A summary of the response regarding the perceived ease of use of the system. 0 = I do not agree at all, 4 = I fully agree.

<table>
<thead>
<tr>
<th>I was satisfied with how the robot behaved</th>
<th>μ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>It was easy to learn how to handle the robot</td>
<td>3.23</td>
<td>0.70</td>
</tr>
<tr>
<td>It was easy to maneuver with the robot</td>
<td>3.23</td>
<td>0.70</td>
</tr>
<tr>
<td>It is easy to make the robot do what I want</td>
<td>3.00</td>
<td>0.71</td>
</tr>
<tr>
<td>Concepts and language in the computer</td>
<td>3.50</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Table 10. Summary of response among the presence dimensions in the TPI. The Perceptual Realism combines Object Realism and Person Realism.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>μ</th>
<th>σ</th>
<th>a</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Presence</td>
<td>3.63</td>
<td>1.98</td>
<td>0.87</td>
<td>11</td>
</tr>
<tr>
<td>Social Presence - Passive Interpersonal</td>
<td>5.60</td>
<td>1.57</td>
<td>0.89</td>
<td>13</td>
</tr>
<tr>
<td>Engagement (Mental Immersion)</td>
<td>5.86</td>
<td>1.15</td>
<td>0.95</td>
<td>13</td>
</tr>
<tr>
<td>Object Realism</td>
<td>5.42</td>
<td>1.29</td>
<td>0.92</td>
<td>12</td>
</tr>
<tr>
<td>Person Realism</td>
<td>5.67</td>
<td>1.37</td>
<td>0.86</td>
<td>12</td>
</tr>
<tr>
<td>Perceptual Realism*</td>
<td>5.54</td>
<td>1.38</td>
<td>0.94</td>
<td>12</td>
</tr>
<tr>
<td>Social Richness</td>
<td>4.97</td>
<td>1.86</td>
<td>0.89</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 11. Summary of response among the social presence dimensions in the Networked Minds. Psycho-behavioural Interaction has four sub-dimensions; Attentional Engagement, Emotional Contagion, Comprehension and Perceived Behavioural Interdependence.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>μ</th>
<th>σ</th>
<th>α</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-presence</td>
<td>5.57</td>
<td>1.78</td>
<td>0.75</td>
<td>11</td>
</tr>
<tr>
<td>Perceived Psychological Engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Attentional Engagement</td>
<td>5.09</td>
<td>2.31</td>
<td>0.73</td>
<td>11</td>
</tr>
<tr>
<td>- Emotional Contagion</td>
<td>4.13</td>
<td>2.24</td>
<td>0.90</td>
<td>5</td>
</tr>
<tr>
<td>- Comprehension</td>
<td>5.97</td>
<td>1.06</td>
<td>0.65</td>
<td>11</td>
</tr>
<tr>
<td>Perceived Behavioural Interdependence</td>
<td>5.95</td>
<td>0.98</td>
<td>0.88</td>
<td>11</td>
</tr>
</tbody>
</table>

4.3. Perceived ease of use

The sociometric measures outlined in Section 2.3 and Section 3.3 are defined for measuring social communication between humans. The social communication that is enabled by MRP systems such as Giraff can be affected by the cognitive burden in maneuvering the robot while interacting with another person. For successful use of MRP systems in elder care, it is therefore important that health care professionals find the system easy to use. A summary of the perceived ease of use of the Giraff MRP system is given in Tables 8 and 9. For the questions included in Table 8, the participants were asked to respond to a number of questions such as “How easy was to make a u turn?” on a Likert scale 1-7 where 1 = very difficult and 7 = very easy. For the statements included in Table 9, the participants were asked to respond on a different Likert scale 0-4 where 0 = I do not agree at all and 4 = I fully agree. Further, they were asked to rate the smoothness of the movement of the robot on a Likert scale 0-4 where 0 = shaky and 4 = smooth. Their rating was $\bar{\mu} = 3.00$, $\sigma = 0.68$.

The response to the questionnaire on perceived ease of use indicate that the participants perceived the system as being easy to use. However, the impression of the researcher who guided the participants around was that this perceived ease of use was not always reflected in her observations. For example, there was one person with notable difficulties in maneuvering the Giraff throughout the scenario who gave high scores overall in the ease of use questionnaire section. This person needed 424 s to complete transition T2 while the average person needed 94.58 s.

4.4. Presence

Tables 10 and 11 summarize the participants’ perceived presence. However, given the small sample of participants and the fact that some of the participants chose not to respond to all questions, the calculated internal consistency Cronbach’s $\alpha$ for the different dimensions of presence. However, the results indicate that the revised scales in the Co-presence and Attentional Engagement dimensions have improved the understanding of the questions in both dimensions.

4.5. Outside factors’ effect on sociometry, ease of use and presence

The activity in an interaction can according to the theory presented in Section 2.3 be assessed by the speaking time (6) and by the number of turns per second (38). An interaction in this respect is between two people, the researcher and the participant and naturally they influence each other in the interaction. Hence, the sociometric measures which are analyzed separately for each sociometric badge used in an interaction can differ.
Even though the number of participants was low, an F-test analysis was performed to assess how the sociometric measures varied depending on if the participant was a user of video communication tools or not. Differences between the different groups of participants were found in both transitions and static situations. It was found that non-users had a higher speaking time during T2 than users. Both the researcher and the non-users of video communication systems were producing more short speech segments (length=1 or 2 s) during T1. The researcher continued to produce more short speech segments (length=1 or 2 s) during T2 when interacting with non-users. On the contrary, there is an indication that the interaction was “smoother” (see Section 2.3) in S2 with participants who used video communication tools. The researcher and the “users” produced a higher number of short speech segments (length=1 or 2 s) during T1. Compared to non-users, the users of video communication tools thought it was significantly easier to make a u-turn with Giraff, $F(1,9)=6.61$, $p<0.05$ ($\mu_{\text{not-user}}=6.29$, $\sigma_{\text{not-user}}=0.95$ and $\mu_{\text{user}}=7.75$, $\sigma_{\text{user}}=0.96$ respectively). They further perceived a higher “Social Richness” (“To what extent a medium is perceived as sociable, warm, sensitive, personal, or intimate when it is used to interact with other people”, $p=0.01$) $F(1,6)=15.47$, $\mu_{\text{non-user}}=5.71$, $\sigma_{\text{non-user}}=0.91$ and the duration of S3 was lower, $F(1,9)=5.41$, $p<0.05$ ($\mu_{\text{user}}=134.5$, $\sigma_{\text{user}}=43.78$ and $\mu_{\text{not-user}}=189.75$, $\sigma_{\text{non-user}}=23.51$ respectively).

A similar analysis was performed to assess how the sociometric measures varied depending on if the participant was playing computer/video games or not. The researcher produced more speech segments and had a higher speaking time during S2 when interacting with participants who played computer/video games. The participants who did not play games produced more short speech segments (length=2s) during T3.

The participants who had played computer/video games answered that it was easier to make the robot stop, $F(1,9)=5.26$, $p<0.05$ ($\mu_{\text{playing}}=6.7$, $\sigma_{\text{playing}}=0.49$ and $\mu_{\text{not-playing}}=5.50$, $\sigma_{\text{not-playing}}=1.29$ respectively on a Likert scale 1-7 where 1 = very difficult and 7 = very easy). They were also happier with how the robot behaved, $F(1,9)=8.69$, $p<0.05$ ($\mu_{\text{playing}}=3.57$, $\sigma_{\text{playing}}=0.54$ and $\mu_{\text{not-playing}}=2.50$, $\sigma_{\text{not-playing}}=0.55$ respectively on a Likert scale 0-4 where 0 = I do not agree at all and 4 = I do not agree at all). These results can explain why the participants who had not played games produced more short speech segments during T3.

To follow up on our expectations Exp2-Exp9, there were differences in the sociometric measures with respect to both usage of video communication tools and computer/video games. For both tools, the activity in the interaction was higher during transitions when the participants were non-users of these technologies. One of our expectations in this experiment, Exp2, was that problems in maneuvering the robot would be reflected in the sociometric data collected during transitions. As such, we believe that the higher activity in the interaction among the non-users could be due to a higher need for help using the robot and a lower computer experience in general. The interaction was “smoother” in S2 with users of video communication tools. The same effect could not be found among the users of computer/video-games. Differences in perceived presence existed only in the usage of video communication tools case. The users of these tools perceived a higher “Social Richness” than the non-users. Likely, the users were more habituated to the type of interaction that MRP systems encompass. On the other hand, there were more differences in the perceived ease of use with respect to usage of computer/video games. These users perceived the robot easier to use.

The eight participants whose face could be fully seen in the video feed on Giraff thought it was easier to start the Giraff client on the computer, $F(1,11)=8.44$, $p<0.05$ ($\mu_{\text{whole-face}}=6.6$, $\sigma_{\text{whole-face}}=0.46$ and $\mu_{\text{not-whole-face}}=6.00$, $\sigma_{\text{not-whole-face}}=0.71$ respectively) and to connect to the robot, $F(1,11)=13.28$, $p<0.01$ ($\mu_{\text{whole-face}}=6.75$, $\sigma_{\text{whole-face}}=0.48$ and $\mu_{\text{not-whole-face}}=5.80$, $\sigma_{\text{not-whole-face}}=0.45$ respectively). However, it had no noticeable effects on sociometric measures. Although this could be seen as negligible problem since it did not have a negative effect on the pilot users’ experience, there is a need to consider the issue from the perspective of the local user in normal usage. It was found in e.g. [24] that local users need to be able to identify who the pilot user is. Further, it was found in [22] that visitors at a museum hosting a guiding robot interacted longer when the face of the person in the robot was shown. Thus, a modification of the client interface that helps the pilot user to position him or herself in a way so that her whole face can be seen by the local user could improve the interaction from the perspective of a local user.4

4 In fact, the elderly people being guided around by a researcher embodied in a Giraff in [20] said it was important to be able to see the person. The evaluation was performed in the ExCITE project [2].

4.6. Relations between dominant formations and sociometry, perceived ease of use and presence

A series of F-tests were performed to investigate whether different dominant formations are associated with differences in the sociometric measures. For each of the situations T1, T3, S2 and S3 respectively. F-tests were made to assess whether the sociometric measures differed depending on the spatial formation. The sociometric data used in the comparison was sampled during the specific part of the scenario (T1, T3 and S2-S3) and the scenario as a whole. We are aware of the fact that running F-tests on such a low number of participants is questionable, therefor the following results should be seen as indications. Additional studies would be needed to confirm these findings.

To follow up on Exp2, it was investigated whether problems in maneuvering the robot were reflected in the sociometric data during transitions. No differences were found in T1 or T3. In an attempt to follow up on the expectation without the existence of different formations in T2, the F-tests were done using the spatial formation in T1. The results indicate that there were differences in the number of turns per second, the speech energy and the speaking time of both the participant and the researcher during T2. However, it should be noted that there were only two participants driving ahead in T1 and that one of those was loud during T2. Therefore, this may have caused the difference.

The duration of S2 was longer among the two participants who interacted mainly in a vis-a-vis formation than in the expected L-shape in S2, $F(1,10)=11.35$, $p<0.01$ ($\mu_{\text{vis}}=605.2$, $\sigma_{\text{vis}}=8.48$ and $\mu_{\text{L-shape}}=204.67$, $\sigma_{\text{L-shape}}=34.74$ respectively). The total speech energy, the participants’ speech energy as well as the number of long speech segments (length=6s) during S2 were also higher in the interactions with these participants. Due to the odd positioning and orientation of the two participants, the researcher had to make many minor adjustments of the spatial configuration when fetching the described objects, which affected the time needed. There were 8 and 14 adjustments of the spatial configuration for these two participants during S2. The maximum number of adjustments during the interactions with the remaining participants was four and in some interactions, no adjustment was required.

All but one participant interacted mainly in vis-a-vis formation in S3. Therefore, no analysis of correlations between spatial formations and the perceived ease of use, presence and sociometric measures was performed in S3.
To summarize what could be seen in terms of relations between spatial formations and sociometric measures, differences (although not so many) existed not only during the transitions but also during static situations. The interaction was more active during transition T2 when involving participants who deviated from the expected, behaviour in T1. Similarly, activity was higher during interactions in a vis-a-vis formation instead of the expected L-shape formation in the kitchen, S2.

The only relation found between spatial formations and perceived presence in the experiment was found in T1. The eleven participants who followed the researcher in T1 perceived a higher "spatial presence" (the extent to which a person feels as if she or other objects, people or environments have been transported [27]) than the ones who drove ahead of the researcher, $F(1,9)\approx7.25, p<.05 \ (m_{\text{follow}}=3.90, m_{\text{lead}}=1.027$ and $m_{\text{location}}=1, m_{\text{driver}}=1$ respectively). Hence, we found little support for Exp3 in this study.

The eleven participants who had followed the researcher in transition T1 needed less time in T2, $F(1,11)=5.31, p<.05 \ (m_{\text{follow}}=68.10s, m_{\text{lead}}=35.196$ and $m_{\text{location}}=227s, m_{\text{driver}}=278.60$ respectively). However, as can be seen from the $s$-values, the standard deviations were high. The eight participants who had moved ahead of the researcher in T3 thought it was easier to learn how to handle the robot, $F(1,11)=5.44, p<.05 \ (m_{\text{follow}}=3.75, m_{\text{lead}}=0.48$ and $m_{\text{location}}=3.00, m_{\text{driver}}=0.71$ respectively on a Likert scale 0-4 where 0 = I do not agree at all and 4 = I fully agree). They also perceived the robot’s movement as smoother, $F(1,11)=10.46, p<.01 \ (m_{\text{lead}}=3.38, m_{\text{location}}=0.52$ and $m_{\text{lead}}=2.40, m_{\text{driver}}=0.55$ respectively on a Likert scale 0-4 where 0 = shaky and 4 = smooth). Exp3 is supported by both the response on the questionnaire section on perceived ease of use and the high variance in time needed in the transition.

4.7. Presence and how it relates to sociometric measures and ease of use

A set of Pearson correlation tests were done to investigate whether there were correlations in between the different presence dimensions and the sociometric measures: speech energy, number of turns and speaking time during the transitions T1-T3 and the static situations S2-S3.

The correlations found for "spatial presence" are all related to transitions. The higher speech energy during transition T2, the lower spatial presence ($r=-.892, p<.017$). A similar trend was seen for "behavioural interdependence" (the degree to which actions of the user are dependent on actions of the other and vice versa) which was higher when the speech energy during S2 was high ($r=.892, p<.017$) but lower when the speech energy during transition T3 increased ($r=.923, p<.009$). To summarize, there is support for Exp3 and Exp4. Generally, the dimensions that involve social aspects (social presence - passive interpersonal, engagement (mental immersion), person realism, co-presence, comprehension and perceived behavioural interdependence) are positively correlated with the amount of activity during the static situations S2 and S3. Many of them (engagement (mental immersion), social richness, co-presence, emotional contagion, comprehension and perceived behavioural interdependence) are also negatively correlated with the activity during the transitions T1-T3. Further, the perceived spatial presence is also negatively correlated with the amount of activity during T2 and T3. Likely the participants who had more difficulties in maneuvering the device had a lower sense of being transported to the location of the robot which also likely affected their perceived social presence with the researcher situated there.

Table 12 summarizes how the experimental expectations hold. Due to the fact that the number of participants was low, it is difficult to generalize the results. However, the quantitative analysis of the sociometric data correlates with a number of parameters gathered in a qualitative analysis, e.g. different dimensions of presence and observed problems in maneuvering the robot. The implications of this form a basis upon which a methodology for measuring interaction quality in MRP systems can be obtained.

5. Discussion and Conclusions

The number of different mobile robotic telepresence (MRP) systems is rapidly increasing and several different application areas are foreseen. These include elder care, hospitals, office environments and sick children attending classes. In this paper, we have reported on the results from a study in which health care professionals maneuvered the Giraff MRP system for the first time while interacting with a researcher. The social communication with elderly enabled by a MRP system is different from the normal work processes of, for example, a nurse. For successful deployments of MRP systems in elder care, it is important that they provide an acceptable means of social communication.

We have demonstrated the use of several tools of different nature in an evaluation of the quality of interaction when using MRP systems. The tools include: (1) a subjective questionnaire that assesses the perceived presence and ease of use, (2) the Kendon F-formations sys-
Table 12. A summary on how the experiment expectations hold.

<table>
<thead>
<tr>
<th>Exp</th>
<th>The interaction would be more active in terms of amount of speech energy, speaking time and turns during S2, S3 and T3.</th>
<th>Na, but division of data allowed for finding correlations between spatial formations and sociometric data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp2</td>
<td>Problems in maneuvering the robot would be reflected in the sociometric data collected during transitions.</td>
<td>Na, there was no difference in the sociometric data in T1 or T3.</td>
</tr>
<tr>
<td>Exp3</td>
<td>The perceived spatial presence would be reflected in the sociometric data collected during transitions T1-T3.</td>
<td>Yes, it was negatively correlated with a higher activity in T2-T3.</td>
</tr>
<tr>
<td>Exp4</td>
<td>The perceived social presence would be reflected in the sociometric data collected during static situations S2-S3.</td>
<td>Yes, most dimensions were positively correlated with a high activity in S2-S3 but also negatively correlated with a higher activity in T2-T3.</td>
</tr>
<tr>
<td>Exp5</td>
<td>Relations between chosen formations and perceived presence would exist.</td>
<td>Na, only between the formation in T1 and spatial presence. However, the number of participants and the response rate on the questions were low.</td>
</tr>
<tr>
<td>Exp6</td>
<td>Relations between chosen formations and perceived ease of use would exist.</td>
<td>Yes, with respect to the questionnaire data and the time of T2.</td>
</tr>
<tr>
<td>Exp7</td>
<td>Relations between experience of video communication tools (and/or computer/video games) and sociometric data would exist.</td>
<td>Yes, there was a higher activity in the interaction with inexperienced users than with experienced users during transitions.</td>
</tr>
<tr>
<td>Exp8</td>
<td>Relations between experience of video communication tools (and/or computer/video games) and perceived presence would exist.</td>
<td>Yes, but only with respect to experience of video communication tools and social richness. The users with experience perceived a higher social richness than the inexperienced users.</td>
</tr>
<tr>
<td>Exp9</td>
<td>Relations between experience of video communication tools (and/or computer/video games) and perceived ease of use would exist.</td>
<td>Yes, the users with experience of playing games perceived the robot as being easier to use than the inexperienced users.</td>
</tr>
</tbody>
</table>
also be used for dynamically changing the client interface depending on the skills of the person maneuvering the robot. What is important for a novice learning to maneuver the system can be an annoyance for an experienced user. A person having problems maneuvering the system might need more automatic functionalities while an experienced driver may want more freedom in moving around. Here sociometric measures can be used to automatically adjust the client interface.

### Acknowledgments

This work has been supported by the EU under the Ambient Assisted Living Joint Program - ExQITE Project (AAL-2009-2-125). The authors would like to acknowledge Mia Lindström for support in finding participants for the experiment and Hadi Banaee for annotating recorded videos. Finally, we would also like to thank Socrimetic Solutions for the support in using the sociometric badges.

### Appendix

The level of co-presence “Is influenced by the degree to which the user and the agent appear to share an environment together,” p.5 [2]. The revised version of the Networked Minds Co-presence dimension as used in this study consists of eight questions where the first four are to be answered on a Likert scale 1-7 where 1 = Not at all and 7 = To a very high degree and the last four on a Likert scale 1-7 where 1 = I fully agree and 7 = I do not agree at all.

1. I felt that x and I were in the same place.
2. I believe that x felt as if we were in the same place.
3. I was aware that x was there.
4. x was aware that I was there.
5. I hardly noticed X.
6. X hardly noticed me.
7. I felt as if we were in different places rather than in the same place.
8. I think that x felt as if we were in different places rather than in the same place.

The attentional engagement “seek to measure the degree to which the users report attention to the other and the degree to which they perceive the other’s level of attention toward them,” p. 10 [2]. The revised version of the Networked Minds Attentional Engagement dimension as used in this study contains six questions where the first two are to be answered on a Likert scale 1-7 where 1 = Not at all and 7 = To a very high degree and the last four on a Likert scale 1-7 where 1 = I fully agree and 7 = I do not agree at all.

1. I paid attention to x.
2. x paid attention to me.
3. I was distracted from x when other things happened.
4. X was distracted from me when other things happened.
5. I tended to ignore x.
6. X tended to ignore me.

The object realism dimension originates from the TPI perceptual realism dimension [27]. It consists of two questions each that were to be answered on the Likert scale 1-7 where 1 = not at all and 7 = to a very high degree.

1. The objects you saw looked like they would have done in reality.
2. The objects you saw sounded like they would have done in reality.

Similarly, the person realism dimension also originates from the TPI perceptual realism dimension and consists of two questions using the same scale.

1. The person you met looked like she would have done in reality.
2. The person you met sounded like she would have done in reality.

### References


