Towards Measurement of Interaction Quality in Social Robotic Telepresence

Annica Kristoffersson¹, Kerstin Severinson Eklundh² and Amy Loutfi¹

Abstract—This paper presents tools for measuring the quality of interaction in social mobile robotic telepresence. The methodology is in part based on Adam Kendon’s theory of F-formations. The theory is based on observations of how bodies naturally orient themselves during interaction between people in real life settings. In addition, two presence questionnaires (Temple Presence Inventory and Networked Minds Social Presence Inventory), designed to measure the users’ perceptions of others and the environment when experienced through a communication medium were used. The perceived presence and ease of use are correlated to the spatial formations between the robot and an actor. Use of the tools is validated experimentally on a dataset consisting of interactions between an elder (actor) and 21 different users being trained in piloting a mobile robotic telepresence unit. The evaluation has shown that these tools are suitable for evaluating mobile robotic telepresence and also that correlations between the tools used exist. Further, also from a local user perspective, the spatial formations have affected the perceived comfort in an interaction.

I. INTRODUCTION

Mobile robotic telepresence (MRP) is a combination of teleoperation and telepresence offering a “walking around” capability. A MRP system is a video conferencing system through which the pilot of the system can move around in a remote environment and interact with its inhabitants (local user). The pilot interacts with the local user while at the same time navigating the robot via a computer. The local user experiences the interaction with the pilot via a video conferencing system mounted on a robotic base. Thus, usage of a MRP system includes many sorts of interaction simultaneously. Local users of the system are interacting with another human (HHI) but at the same time, they are interacting with a robot (human-robot interaction, HRI). The pilots of the MRP system are interacting with a computer system (HCI) but at the same time, they are interacting with another human (HHI) while being embodied in a robot which they cannot see themselves. Commercial examples of MRP systems are Giraff (Giraff Technologies [8]), QB (Anybot [23]), Texai (Willow Garage [24]) and VGo (Vgo Communications [25]).

In this paper, we discuss two experiments with a specific MRP system, the Giraff. In the first experiment, 21 alarm operators were trained in steering the Giraff by visiting an elder (actor). Regarding the first experiment, this paper presents an extended analysis of the results presented in [17] by studying the interaction between the pilot and the local user and describing the interaction using measures that relate to the spatial formations that occur between the local user and the pilot. To characterize spatial formations, we take inspiration from Kendon’s F-formations which provide a framework to describe the natural positioning/configuration of people when engaged in specific tasks. i.e. how people orient themselves with respect to each other. To characterize the levels of social and spatial presence, a questionnaire based on the Temple Presence Inventory (TPI) [20] and the Networked Minds Social Presence Inventory (Networked Minds) [5] was filled by the pilot user directly after the interaction took place. The TPI and the Networked Minds are two standard questionnaires to assess the perceived presence. The Networked Minds is designed to measure the users’ perceptions of others experienced through a communication medium [4] and the TPI is appropriate for use with most media and media content [20]. Both the TPI and Networked Minds have been applied in previous studies in HRI, e.g. [2], [1], [11]. We also present a preliminary analysis of the second experiment in which 10 elderly interacted with a pilot in the Giraff. In this experiment, the pilot intentionally chose one out of two different spatial formations in different steps of a scripted scenario. After the experiment, the elderly were retrospectively interviewed to assess how they perceived the interaction with the person piloting the Giraff and the Giraff itself. The motivation behind both experiments is that we aim to find objective metrics upon which to improve the design of MRP systems. To do this, an understanding of the factors which impact interaction quality is important. Particularly related to social robotic telepresence is the effect of mobility achieved via embodiment (and vice versa), and therefore an aspect to study is how mobility affects the quality of social interaction.

The article is organized as follows, in Section II a description of Kendon’s F-Formation System and related work is given. Section III details the experimental setups and data collection. Section IV presents our hypotheses. The results are presented and discussed in Section V. Finally, conclusions are drawn in Section VI.

¹This paper is a modified version of an article that will appear in International Journal of Social Robotics. If you use any of the material regarding the first experiment, please cite the original paper [18].

¹A. Kristoffersson and A. Loutfi are with AASS Center of Applied Autonomous Sensor Systems, rebro University, Sweden firstname.lastname at oru.se

²K. Severinson Eklundh is with the School of Computer Science and Communication, KTH Royal Institute of Technology, Sweden k.se at csc.kth.se
II. SPATIAL RELATIONSHIPS IN INTERACTION

Much of the work presented in this article is based on a theory on spatial handling with origins in HHI, Kendon’s F-Formation System. In this section, we first describe this theory. We then discuss how spatial and embodiment constraints influence the quality of interaction and spatial formations [12], [15], [16], [21]. Communication over a medium is different from HHI as several factors important for the level of intimacy [3] are distorted [10]. Up until this point, there are only a few HRI studies which make use of the F-formations, e.g. [13], [19], [26].

Kendon studies spatial relations that occur whenever two or more people engage in an interaction and his claim is that “a behaviour of any sort occurs in a three-dimensional world and any activity whatever requires space of some sort”, p. 1 [16]. This space must have physical properties, allow the organism to do what it needs to do and be differentiated from other spaces. According to Kendon, it is common in any setting for several individuals to be co-present, but how they orient and space themselves in relation to one another directly reflects how they may be involved with one another. Kendon’s F-formation system is a well known theory of spatial relationships. Kendon distinguished three F-formations by observations; (1) vis-a-vis, (2) L-shape and (3) side-by-side are depicted in Fig. 1. They arise when “two or more people sustain a spatial and orientational relationship, p.209” [15].

The vis-a-vis is an arrangement formed by two individuals who are facing each other, the L-shape is generated when two individuals are standing perpendicularly to each other facing an object and the side-by-side arrangement is when two participants stand close together looking in the same direction. The shared space created by the F-formation is called the O-space and is the overlap of transactional segment to which two or more people direct their attention and manipulate objects [15], [21]. The type of F-formation created can be influenced by environmental features such as obstacles and walls. The influence of these features on spatial patterns have yet to be studied in depth [16]. The F-formation spatial arrangement is reconfigured as participants in a group come and go, and change their positions and orientations [19]. However, the orientation of the lower part of the body has a dominant effect on the reconfiguration of the arrangement [15].

According to Marshall et al. [21], physical environments can limit and constrain opportunities for some shared activities, while encouraging others. They concluded that some features of the physical environment may work to discourage the creation of certain F-formations. Hornecker [12] has developed the concept of embodied constraints, these restrict what people can do based on the configurations of space and objects and make some behaviors more likely than others. Along these lines, physical embodiment via a MRP system can also constitute an embodied constraint in that the physical properties of the system restrict particular behaviors, e.g. the inability to move backwards in the environment.

The level of intimacy in HHI is an equilibrium “joint function of eye-contact, physical proximity, intimacy of topic, smiling, etc, p. 293” [3]. When using MRP systems, this would potentially mean that the experienced intimacy for the local user and pilot engaged in an interaction would be dependent on the above factors. A factor that has been pointed out as being of importance in a MRP system is the ability to adjust the height of the robot to increase the level of eye-contact [6], [14]. It should be noted that an interaction enabled by use of computer-mediated technologies, e.g. a video-conference, can not be equivalent to face-to-face situations according to Heath and Luff [10]. They write that gestures and other forms of body movement including gaze, which are systematically employed in face-to-face communication prove ineffective in a large part in video communication.

In the HRI domain, Kuzuoka et al. [19] examined how reconfiguration of F-formation arrangements occurred as an effect of a guiding robot following a predefined trajectory at a museum when rotating either its head, its upper body or the whole body to guide their listeners when talking about exhibits. The results are corresponding well with Kendon’s finding that the orientation of the lower body has a dominant effect on the automatic creations of F-formations. Yamaoka et al. [26] developed a model for how an information-presenting robot should appropriately adjust its position and found that robots that were constrained to an O-space were perceived as more comfortable than robots being close either to the listener or the object the robot was presenting. Hüttenrauch et al. [13] studied spatial formations in real home settings by deploying a service robot and instructing the users to guide the robot around in their own home and to teach the robot rooms, locations and objects. Each trial was video-taped for later analysis and revealed that the participants would lead the robot when passing narrow passages. The formation that occurs in the narrow passage indicates that one can also distinguish another spatial formation in HRI which Hüttenrauch calls follow in which the robot follows the user.

In the outlined material on spatial formations in HRI, we are missing studies on how the quality of interaction with others is perceived by both pilots and local users. We are interested in how the perceived quality varies depending on what spatial formations that are created during an interaction.
III. Method

In experiment 1, 21 alarm operators were invited to a training session during which they made a remote visit to an elder’s home via Giraff. The alarm operators respond to alarms coming from elderly who by pushing a button on a necklace get in direct contact with the service. The training session was a preparation for deploying a Giraff in a home with two elderly, one of which was using a wheelchair. This visit was their first use of Giraff and was presented as an opportunity to train on steering and using the Giraff. The training session served the purpose of collecting data via questionnaires and observations. The training session took place in a smart-home environment. The locality for the training sessions was chosen in order to simulate a close to real experience of use of the robot in a real home setting. This particular setup simulated an elder residence in which one elder(actor) was sitting in a wheelchair. A graphical overview of the remotely visited home can be found in Fig. 2.

There are constraints in the physical space. The kitchen is too small to accommodate both a wheelchair and a Giraff at the same time. The distance between the table and television set makes it difficult to fit the Giraff beside the wheelchair. The spatial constraints may discourage creation of F-formations and encourage other spatial formations. The alarm operators who came individually were placed in a room where a laptop equipped with a headset and a mouse was installed. Each training session began with informing the participants about the computer and its connected devices and then instructing them to make a visit to a remote home with the Giraff. They were instructed to interact with the elder as if it was a visit to a real home. Further, they were informed that they would be asked to fill in a questionnaire after completing the training session.

The use of an actor was essential to script the visit and to ensure that the interaction was as similar as possible between the visits from different alarm operators. However, two different actors were used during the training sessions.

The procedure outlined below was used for each visit in experiment 1. Here, we use pilot to denote the alarm operator, elder to denote the actor in the smart home and researcher to denote the researcher who sat beside the pilot and provided technical support in case of difficulties (e.g. the pilot cannot find the docking station or does not know which buttons to press on the Giraff Pilot application). Numbers in parentheses, e.g. S1 denote situations in the scripted visit.

1) The researcher instructed the pilot to start the Giraff application, log on to the Giraff server and to connect to the Giraff that was facing the wall.
2) Once connected, the pilot was instructed to undock the Giraff from the docking station by pushing the buttons **Backward** and **Turn**. The pilot was asked to locate the elder. The pilot would find the elder in the bed. No guidance in where to find the elder was given.
3) When the pilot had found the elder, the elder moved over to the wheelchair and asked the pilot to follow to the kitchen (S1).
4) While in the kitchen, the elder started a discussion about a medical issue (S2).
5) The elder then asked for help to find the remote control for the television set. The pilot and the elder moved to the living room (S3) to find it. The pilot would tell the elderly that it was lying on the floor in between the sofa table and the television set (S4).
6) After the pilot had found the remote control, an artificially triggered alarm rang in the bedroom. Depending on the pilot’s response, the elder found an appropriate means to conclude the conversation and asked the pilot to return to the docking station.
7) The pilot returned to the docking station and disconnected from the Giraff with help from the researcher if necessary.

For most pilots, the script took in between 200 s to 300 s measuring from when the pilots had undocked and turned the Giraff until the moment when elderly said bye.

In experiment 2, 10 elderly were invited to a guiding tour in a showcase environment. The elderly arrived individually and were met with a researcher who presented the Giraff. The elder was asked to sit down in the sofa in the living room and was offered coffee. A second experimenter then connected to the Giraff and began to interact with the elder. In this experiment, a scripted scenario was followed. However, the pilot chose one of two spatial formations in each step.

1) The pilot undocked, navigated the Giraff and stopped in front of the elder and said welcome. The pilot chose either a vis-a-vis or a look-away spatial formation, see Fig. 4.³
2) The pilot asked the elder to follow to the kitchen where a number of items on and around a table were shown and explained to the elder. These items are part of the showcase apartment and included a chair for assisting elder and a robot arm. In here the pilot either faced the items on the table (L-shape) or the wall. (look-away)
3) The pilot asked the elder to follow to the bedroom that contained sensors spread around the room. The

³The look-away formation is further described in Section IV.
pilot either faced the elder (vis-a-vis) or the bed when describing the content of the room.
4) The pilot drove back to the living room and asked the elder to sit down. Now, a set of paintings, a Roomba and a medicine dosett were shown. Again, the pilot chose either a vis-a-vis or a look-away formation.
5) The pilot said goodbye and returned to the docking station. The first researcher returned.

A. Data Collection

Throughout experiment 1, ten permanently installed web cameras positioned at different locations in the ceiling of the smart-home recorded each of the participant’s training sessions. The cameras were configured to capture most parts of the apartment from different angles. A snapshot from one of the videos is found in Fig. 3. The figure shows the Giraff and the elder in the wheelchair from several angles allowing analysis of different parameters, e.g. formations. Video recordings enable repeated and detailed access to the conduct and interaction of participants, and, more specifically, the interplay of talk, bodily, and material conduct [22]. Recording of video data is an ethical issue. It is argued that when being filmed, people inevitably react to the camera - rendering the data is unreliable [9]. In this experiment however, the participants were unaware of the fact that the Giraff was being recorded on video so they did not react to being filmed. The video recordings were taken without sound and they do not reveal who is currently embodied in the MRP system. However, the choice of camera configuration came with a sacrifice. The video recordings were not able to capture facial expressions from the pilot or sound from the interaction between the pilot and the elder. The sacrifice limits the possibility to fully understand the interaction and its effects on chosen spatial formations.

Upon completing the training, the participants were asked to fill in the questionnaire. Two of the sections in the questionnaire assessed the perceived social and spatial presence. Each dimension of perceived presence in the questionnaire consisted of several questions that were to be answered on a likert scale 1-7 where 1 = not at all and 7 = to a very high degree except for the questions in the TPI dimension Social richness. Social richness was assessed by asking the pilots to rate their experience in opposite couples (for example whether they perceived the experience to be Insensitive or Sensitive) on a scale 1-7. The dimensions Object realism and Person realism originate from the TPI dimension Perceptual realism. The original dimension contained more questions about modalities not available in a MRP system. Thus, a subset of questions from the perceptual realism dimension was used, see Appendix A.

In Experiment 2, three cameras were used to video capture the experiment. The movies recorded were primarily used in order to perform a voice-recorded retrospective interview with the elderly after having completed the scripted scenario. A retrospective interview technique was used immediately after the elderly were guided. The elderly watched a movie for each step of the guiding tour and were asked to comment to it by responding to a set of questions regarding a number of parameters e.g. the spatial formations.

IV. HYPOTHESES

The participants in experiment 1 were faced with a multiple of novelties, such as using a video conferencing system to interact with somebody, steering a robot, meeting a new person and a new environment. These factors, and the fact that video conferencing systems cause a distortion in perception made us expect that some pilots would not turn the Giraff in a vis-a-vis formation while interacting with the elder. Two of the situations included in the script enforced a movement of the Giraff together with the elder (S1 and S3) and it was expected that these would lead to situations in which the pilot user would either follow or go ahead of the elder. As already discussed in Section III, the size of the kitchen would potentially limit the possible spatial formations because it could not accommodate a wheelchair and a Giraff at the same time. In combination with the fact that the elder would discuss a medical issue (S2) with the user of the Giraff we expected that the user would form a vis-a-vis formation at an appropriate distance to the elder. Another space constraining factor in the apartment was the distance between the table and the television set. We expected that it would be difficult to fit the Giraff and the wheelchair beside each other at this location and therefore expected that a L-shape would occur in S4. Thus six patterns of formations were foreseen to occur during the training sessions as shown in Fig. 4. Three of which being F-formations as defined by Kendon and three of which being assumed based on the above presented assumptions. To conclude, the spatial formations were expected to vary throughout the scenario based on the different situations and changes in available space in the scripted scenario.
In our previous analysis [17], we analyzed the perceived presence and ease of use during the training session. It was found that the presence questionnaire was suitable for use in this HRI setting. According to Kendon, as discussed in Section II, it is common to be co-present when interacting with others. That is, how people orient and space themselves in relation to one another is a reflection of how they may be involved with one another. Thus it was hypothesized that:

[H1] Relations between chosen formations and perceived presence would exist.

We further expected that the pilots perceiving the MRP system as being easier to use would orient themselves in suitable formations to a higher degree than the ones not orienting themselves in suitable formations.

[H2] Relations between chosen formations and experienced ease of use would exist.

In experiment 2, we expected that the elderly would feel more comfortable interacting via the Giraff if the pilot chose the same spatial formations as would naturally occur in real life situations.

V. RESULTS

A. Subjects

The users invited to experiment 1 were alarm operators. The average age of the 21 users was $\mu_{\text{age}} = 42.19$, $\sigma_{\text{age}} = 10.34$. Only two of the alarm operators were men, therefore no comparison between genders is done in this study. None of the users had previous experience of Skype or similar systems for communicating with or without video feed. On a likert scale 1-7 where 1 = not at all and 7 = to a very high degree, the experience of using such technologies was $\mu = 1.90$ and $\sigma = 1.67$. Thus there was a dual novelty for the participants in that they lacked experience of both videoconferencing technology and MRP systems.

In experiment 2, 3 men and 7 women in the age range 61-82 years old participated. All claimed they had a habit of using computers and three stated they were using video conferencing software.

B. Experiment 1 - Choice of formations based on situations and space

It was expected that the different situations S1-S4 in combination with space constraints, would result in the spatial formations as depicted in Fig. 4. To investigate this, the videos were analyzed in several steps. First, each movie was watched and notes were made on when the different steps in the script occurred and how the interaction between the elder and the pilot took place. A number of fields of information had to be filled for each video. Secondly, all of the 21 video recordings were re-watched and notes were made on how the formations fluctuated during the interactions. These notes were then converted to illustrative graphs showing the fluctuations between different formations.

To exemplify, two graphs showing the occurrences of formations between two different pilots and the elder are presented in Fig. 5 (Pilot 1-1) and Fig. 6 (Pilot 1-10). The x-axis shows the time having elapsed from when the pilot moved the Giraff forward after undocking in Step 2. The y-axis shows the different formations as defined in Fig. 4.

In Fig. 5, Pilot 1-1 found the elder after 55 s. The pilot formed a vis-a-vis formation while interacting with the elder in the bedroom. Thereafter, the pilot followed the elder to the kitchen (S1) after 78 s. Upon arrival to the kitchen after 100 s, the pilot chose a vis-a-vis formation while interacting with the elder in the kitchen (S2). This formation was kept throughout S2 with a short break during which the pilot turned the Giraff around 360 degrees. The pilot followed the elder from the kitchen towards the living room (S3) after 183 s. At 195 s, the pilot formed a quick vis-a-vis formation after which it continued to lead towards the remote control. The remote control was found (S4) in a side-by-side formation after 207 s. This formation lasted for 13 s after which a vis-a-vis formation was upheld for 12 s. The elder then went towards the bedroom followed by the pilot, a new vis-a-vis occurred at 240 s. This formation lasted for 6 s after which the elder said bye. In this example, the pilot did as expected in S1 and S2.

In Fig. 6, Pilot 1-10 formed a vis-a-vis formation while interacting with the elder in the bedroom. Thereafter, the pilot started following the elder to the kitchen after 43 s (S1) and arrived there after 62 s. The pilot chose a vis-a-vis formation during the interaction in the kitchen. After 97 s, the pilot chose to go ahead to the living room (S3) in order to find the remote control. The remote control was found after 116 s during a L-shape formation (S4). This formation lasted for 9 s after which a Vis-a-vis was upheld until 172 s when the elder said bye. In this example, the pilot did as expected in S1-S4.

To be able to assess whether the pilots acted according to our expectations and whether the choice of formation correlated with perceived presence and ease of use, the
other seven pilots formed a formation while communicating in the kitchen in S2. The remote control to the television set in the living room were focusing on the interaction with the elder rather than in the questionnaire. Another reason may be that the pilots have been more difficult to navigate than what was expressed in (S1). Specifically F(1,19)=6.36, $\prec$ 0.05 ($\bar{x}_{\text{ahead}} = 6.11$, $SD_{\text{ahead}} = 0.23$ and $\bar{x}_{\text{follow}} = 5.78$, $SD_{\text{follow}} = 0.21$).

The pilots who chose the hypothesized L-shape in S4 experienced a higher Comprehension than the pilots who chose the side-by-side formation. Specifically F(1,19)=4.26, $\prec$ 0.05 and $L_{\text{L-shape}} = 5.15$, $SD_{L_{\text{L-shape}}} = 0.96$ among the 13 pilots choosing the hypothesized formation and $S_{\text{side-by-side}} = 4.33$, $SD_{S_{\text{side-by-side}}} = 0.75$ among the eight pilots who chose the side-by-side formation. See Appendix A for more details on the dimensions Attentional Engagement and Comprehension.

Considering that the period of time during which the pilot described the position of the remote control was short (e.g. 13 s for pilot 1-1 in a side-by-side formation and 9 s for pilot 1-10 in a L-shape formation), it is interesting that there exists a significant difference in perceived Comprehension. The numbers indicate that the pilots who formed the hypothesized L-shape were also able to better understand the elder’s intentions, thoughts etc. As pointed out in Section V-B, five of the pilot’s who chose the side-by-side formation in S4 were observed to steer the robot with more difficulty. It may be the case that they were not able to focus as much on the elder and thus not understand the elder’s intentions, thoughts etc. to the same degree.

The majority’s choice of formation when describing the position of the remote control and its relation to Comprehension is well in line with Kendon. He claims that “In conversations between just two persons, when the topic is disembodied, the arrangement tends to be L-shaped... Typically, when two people greet one another and then continue to talk together on some topic, they can be observed to begin with a face-to-face arrangement and then to shift to an L-arrangement as they move from salutation to talk” pp. 8-9. [16]

To summarize, as hypothesized in H1 there are correlations in between what spatial formations the pilots used and how spatially and socially present they felt in the environment and with the elder.

D. Experiment 1 - Relations between formations and perceived ease of use

The pilots were asked to respond to a number of questions such as “How was it to connect to Giraff?” on a likert scale 1-7 where 1 = very difficult and 7 = very easy. The users...
have responded 5 or higher in average on all questions asked regarding the perceived ease of use. One-way Anova analysis reveals that the perceived ease of use varies depending on the occurring spatial formations in S1 and S2 but not in S3 and S4.

The pilots who did as hypothesized in S1 answered significantly higher on how easy it was to start the Giraff application, F(1,19)=4.58, < 0.05. The mean value for the pilots choosing to follow was $x_{\text{follow}} = 6.67$, $SD_{\text{follow}} = 0.59$ while the value for the pilots choosing to go ahead was $x_{\text{ahead}} = 5.67$, $SD_{\text{ahead}} = 1.53$. It could be the case that pilots who found the interface easier to use were more focused on the task at hand (follow elder to the kitchen).

There is a significant statistical correlation between the pilots’ choice of positioning in S2 and how easy the pilots thought it was to leave the docking station (F(1,19)=8.14, < 0.01) and to make a u-turn (F(1,19)=9.16, < 0.01). The pilots who chose vis-a-vis formation during the interaction with the elder perceived it as easier to both leave the docking station ($x_{\text{vis-a-vis}} = 6.79$, $SD_{\text{vis-a-vis}} = 0.43$ compared to $x_{\text{look-away}} = 5.57$, $SD_{\text{look-away}} = 1.51$) and to make a u-turn ($x_{\text{vis-a-vis}} = 6.21$, $SD_{\text{vis-a-vis}} = 0.58$ compared to $x_{\text{look-away}} = 4.71$, $SD_{\text{look-away}} = 1.70$). As previously discussed in Section V-B, one of the possible reasons for not positioning the Giraff in a vis-a-vis formation with the elder in S2 was that an extra effort was needed. The pilots having chosen the vis-a-vis formation also responds that they perceived it as both easier to leave the docking station and to make a u-turn with the Giraff. For them, steering the Giraff in to a vis-a-vis formation could have been considered as being less of an effort than for the ones having more trouble navigating the Giraff.

To summarize, as hypothesized in H2 there are correlations in between what spatial formations the pilots used and how easy to use they perceived the Giraff system to be.

E. Experiment 2 - Spatial formations from an elderly perspective

While only an initial analysis of video and voice-recording data has been performed it is clear that the importance of eye contact while interacting with a pilot in the Giraff is emphasized by most elderly. This is revealed not only in the retrospective interviews but also in the video data with some of the elderly experiencing the unnatural spatial formations while interacting with the pilot user.

[Ex 1] “[...] it should be turned towards me. The contact is needed.”

[Ex 2] “The eye contact was there, I think that part is important.”

[Ex 3] “I almost had to move myself so that I could see her,[...] I should see the one I talk to.”

A few elderly were concerned about not seeing the pilot during the movement between rooms during the home tour.

[Ex 4] “It felt a bit strange when she had turned towards the table in the kitchen. I was instructed to go there but it felt weird to see her from the back so to say.”

Some elderly chose to move themselves in order to re-configure the spatial formation to face the pilot (vis-a-vis). Apparent in the bedroom was also that some elderly changed the spatial formations depending on what item that was described in the room. To conclude, spatial formations is of relevance for the local users while interacting with a person piloting the Giraff. Worth noting is also that all elderly would have preferred that the pilot adjusted the height of the Giraff while interacting with the elderly when sitting in the sofa.

VI. Conclusions

In this paper, we have investigated tools for measuring the quality of interaction in social mobile robotic telepresence. Using Kendon’s theory of F-formations, that is how bodies naturally orient themselves, the interaction between an elder and mobile robotic telepresence system was monitored during a scripted scenario. In addition a questionnaire that assessed the perceived presence and ease of use was filled by the pilot users. Correlations between the dimensions measured in the questionnaire and the chosen spatial formations emerged. In a second experiment, it was found that spatial formations between the pilot and an elder were also of importance for a comfortable interaction from the perspective of an elder. This work has shown that tools such as F-formations, the Temple Presence Inventory and the Networked Minds Social Presence Inventory are useful for evaluating the quality of interaction in mobile robotic telepresence systems. The experiments showed that these tools are suitable for evaluating mobile robotic telepresence and also that the correlations found in experiment 1 can give important guidelines on how to better operate the robotic unit in order to increase the quality of interaction. The initial analysis of experiment 2 is that using the F-formations are important from the elderly local users point of view. For example, improving the interface in order to allow easier rotation of the robot in order to change spatial formation could lead to a higher perceived comprehension of the thoughts and intentions of the other and a higher quality in the interaction for the pilot as well as the local users. Further experimentation is necessary to better understand the correlations between spatial formations and interaction quality and future work will focus on collection of more experimental data.

APPENDIX A SUPPORTING DEFINITIONS ON PRESENCE

The Object Realism contained two questions:
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.

The Person Realism contained two questions:
1) The person you met looked like it would have done in reality.
2) The person you met sounded like it would have done in reality.

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” [4]. The co-presence as used in this study consists of only four questions:

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 of that x was there.
4) x was aware that I was there.

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 others level of attention towards them, p. 10” [4]. The Attentional engagement as used in this study only contains two questions:

1) I payed attention to x.
2) x payed attention to me.

Comprehension is the degree to which the user and the other understand their respective intentions, thoughts etc.

REFERENCES