



Article

F-Formations for Social Interaction in Simulation Using Virtual Agents and Mobile Robotic Telepresence Systems

Sai Krishna Pathi ^{1,*} , Annica Kristoffersson ² , Andrey Kiselev ¹ and Amy Loutfi ¹

¹ Center for Applied Autonomous Sensor Systems (AASS), School of Science and Technology, Örebro University, 70182 Örebro, Sweden; andrey.kiselev@oru.se (A.K.); amy.loutfi@oru.se (A.L.)

² School of Innovation, Design and Engineering, Mälardalen University, 72123 Västerås, Sweden; annica.kristoffersson@mdh.se

* Correspondence: sai.krishna@oru.se; Tel.: +46-(0)19-30-35-91

Received: 7 August 2019; Accepted: 14 October 2019; Published: 17 October 2019



Abstract: F-formations are a set of possible patterns in which groups of people tend to spatially organize themselves while engaging in social interactions. In this paper, we study the behavior of teleoperators of mobile robotic telepresence systems to determine whether they adhere to spatial formations when navigating to groups. This work uses a simulated environment in which teleoperators are requested to navigate to different groups of virtual agents. The simulated environment represents a conference lobby scenario where multiple groups of Virtual Agents with varying group sizes are placed in different spatial formations. The task requires teleoperators to navigate a robot to join each group using an egocentric-perspective camera. In a second phase, teleoperators are allowed to evaluate their own performance by reviewing how they navigated the robot from an exocentric perspective. The two important outcomes from this study are, firstly, teleoperators inherently respect F-formations even when operating a mobile robotic telepresence system. Secondly, teleoperators prefer additional support in order to correctly navigate the robot into a preferred position that adheres to F-formations.

Keywords: telepresence; mobile robotic telepresence; F-formations; simulation; virtual agents; HRI

1. Introduction

Recently, mobile robotic telepresence (MRP) systems have been gaining importance in numerous applications ranging from elder care to office environments [1–4]. To allow social interactions through MRP, it is important to understand people’s dynamics and behaviors. This understanding can be achieved by studying human social signals, which include nonverbal behavioral cues such as facial expressions, body postures, gestures, and proxemics [5]. With a focus on the latter, proxemics is well-known for integrating robots into human environments. The proxemics concept, which was coined by Hall [6], is the study of humans’ interpersonal distances in co-present, face-to-face social interactions. These proxemics are used by people to spatially position themselves at specific distances with respect to others while interacting with them. This positioning is subconscious and is often performed depending on the intimacy between people [7]. While proxemics focuses on interpersonal distances between people, F-formation, proposed by Kendon [8], focuses on spatial and orientational relationship between people when interacting in groups. According to the original definition ([9], p. 209), “F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct and exclusive access”.

Work on spatial formations has also been conducted in the field of MRP systems [1,10,11]. In a number of works, a better quality of interaction between the teleoperator and local user has

been observed when the teleoperator spatially oriented the robot towards the users according to F-formations. However, there has been little research to determine whether teleoperators of MRP systems would place the robot in formations in order to join a social interaction. For this reason, the focus of our investigation is to determine whether: (1) teleoperators of MRP systems, in general, aim to follow F-formations when joining a social interaction, and (2) if they do not, whether this is due to limitations in the technology per se. To this end, we have designed an experiment where an MRP robot is navigated to groups in different formations.

MRP systems usually are a two-end setup: the robot is present at one end (local user site), and at the other end (remote user or operator site), there is a control interface that allows the operator to teleoperate the robot while interacting with the local user, as shown in Figure 1. In general, MRP systems focus on social interaction between local and remote users. Often, teleoperators engage in social interaction with a group of local users. In such scenarios, the question investigated in this work is whether teleoperators place the robot according to F-formations. Additionally, other scenarios where there are multiple groups with varying group sizes and different spatial formations are also considered. Particularly, two research questions are investigated:

R1 When joining social interactions, do teleoperators adhere to (respect) F-formations?

R2 For joining the groups, would teleoperators prefer an autonomous feature?

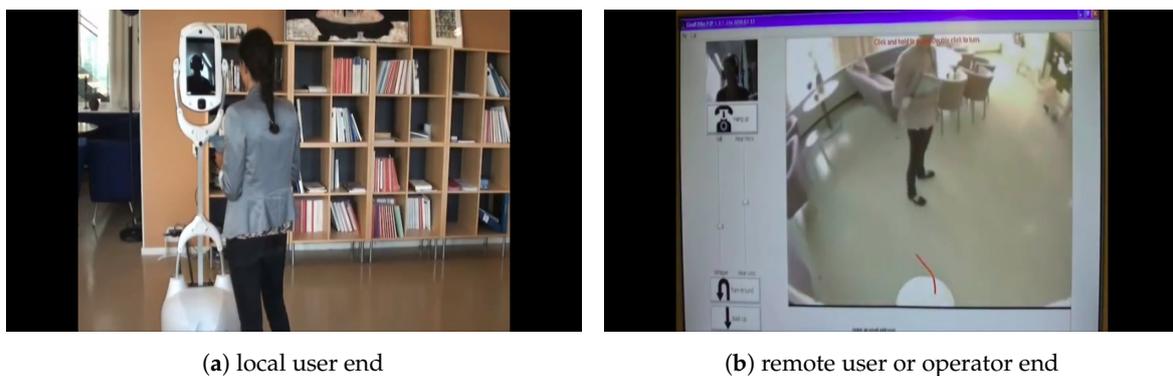


Figure 1. The Giraff robot's [12] two-end setup. (a) A local user is talking to the teleoperator on the screen. (b) Operator end with the graphical user interface (GUI) present on the computer.

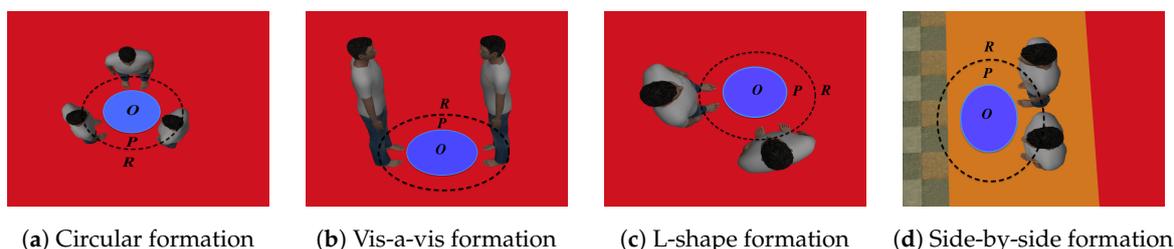
In this paper, our research is guided by **R1** & **R2**. To investigate the above research questions, we designed an experiment in a simulated environment where the various formations are represented by virtual agents (VAs). The experiment consists of two parts. In the first part, operators are asked to navigate and enter each group by navigating the robot in the simulated environment from an egocentric (first person) perspective. In the second part, operators view their performance from an exocentric (third person) perspective and comment on whether their intended performance matches the actual outcomes. The experimental setup was developed in a simulated environment in order to eliminate biases related to reaction of the group towards the robot. Preferential biases of the subjects also facilitate the detection of F-formations using precise known positions and orientations of each object in a simulated scene. Our long-term goal and motivation for performing this study is to verify whether MRP systems should eventually be equipped with features or tools to help operators adhere to F-formations while in operation. While the development of such tools is not the focus of this work, the study presented here is a first step to determine whether operators would need them.

The rest of the paper is organized as follows: Section 2 provides background on F-formations and related works. Section 3 describes the simulation environment in detail, and Section 4 reports on the experiment procedure, the way data were collected, the metrics used for evaluation, and describes the subjects. Section 5 provides discussions and reports on the results, both qualitatively and quantitatively. Finally, Section 6 concludes the paper with future works.

2. Background and Related Works

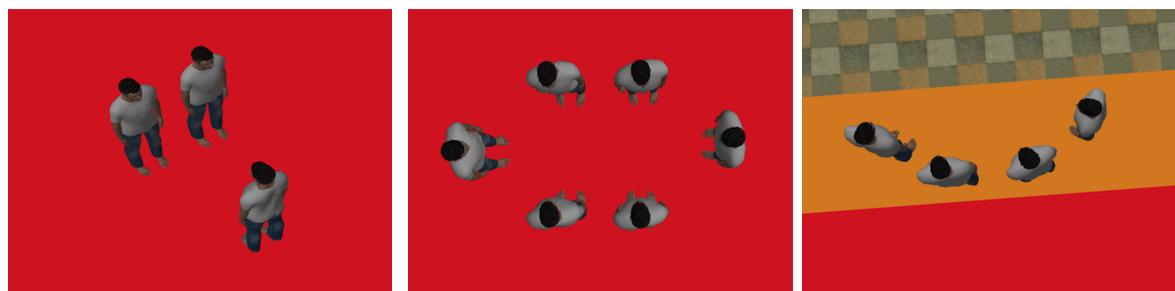
Since proxemics was introduced in 1966 [6], a number of works (e.g., [13–16]) have shown that proxemics play an important role not only in human–human interaction (HHI) but also in human–robot interaction (HRI). Most of the works consider dyadic scenarios in which a robot is approaching a human for a conversation (e.g., [13]), but normally, social interactions occur in groups of two or more people interacting with each other, which is the reason why detecting or studying groups is gaining importance.

F-formations are spatial patterns formed during face-to-face interactions between two or more people. Kendon [8] proposed different spatial arrangements depending on the number of people and the type of social interaction. There are four standard F-formations that are generally formed without any physical constraints and independently of particular situations. They are vis-a-vis, side-by-side, L-shape, and circular. The vis-a-vis formation is when two people are facing each other while interacting. The side-by-side formation is when two people stand close to each other and face the same direction while conversing. The L-shape formation is when two people face each other perpendicularly and are situated on the two edges of the letter “L”. The circular formation is when three or more people are conversing in a circle. The standard F-formations are shown in Figure 2. In addition, the authors from [17,18] proposed three formations that are formed in a spatially constrained environment. The triangle formation is when one person is facing two or more people while interacting, which could be observed at counters. The rectangle formation is formed in board meeting rooms or at dinner tables. The semi-circular formation is where three or more people are focusing on the same task while interacting with each other, which is mainly seen in front of a wall while watching information or a piece of art. These constraint-based formations are shown in Figure 3.



(a) Circular formation (b) Vis-a-vis formation (c) L-shape formation (d) Side-by-side formation

Figure 2. Kendon’s F-formations, these define three social spaces: O-space is the convex empty space between the people in a formation. P-space is the narrow strip on which people are standing while conversing, and R-space is the space beyond P-space.



(a) Triangle formation (b) Rectangle formation (c) Semi-circular formation

Figure 3. Constraint-based formations.

Formations are considered very useful in analyzing and increasing the quality of interaction in social interactions [1,10,11], and a number of works [19–25] have proposed different methods to detect F-formations automatically. The Hough voting strategy (density estimation) was used to locate the O-space (see Figure 2a) by considering each person’s position and head orientation in [19]. Another work [20] considered F-formation as a clustering problem and built a graph model where each node

is a person and the edges between them measures the affinity between pairs. It is an edge-weighted graph model based on body orientation and proximity to find the dominant set [21]. A method called graph-cuts for F-formation (GCFF) for groups in still images using proxemic information was proposed in [22], which introduced a new set of metrics using the idea of a tolerance threshold. In [24], the authors considered body orientation as the primary cue and proposed a joint learning approach to estimate the pose and F-formation for groups in videos. Frustum of attention to extract features from individuals and accordingly classify associates, singletons, and members of F-formations was used in [25]. Vascon et al. [23] developed a game-theoretic model embedding the social-psychological concept of an F-formation and the biological constraints of social attention. They generated a frustum based on the position and orientation of each person and computed affinity to extract the F-formation. These mentioned strategies were developed primarily by the computer vision and machine learning communities and often consider an exocentric perspective (e.g., camera in the ceiling) as opposed from the egocentric perspective of an embodied agent like a mobile robot or telepresence device.

Studies of formations have been considered in the HRI community [26–29]. The influence of F-formations during a task in which a child taught a robot was explored in [26]. However, the authors placed the robot and child according to Kendon’s F-formations prior to the interaction. The challenge of detecting the F-formations using a mobile robot based on lower body estimation, which is obtained by tracking the position and orientation of the people in the scene using an exocentric camera (overhead video data set), was explored in [27]. In our previous work [28], we explored F-formations and detected the F-formations based on the face orientation, and in [29] we proposed an optimal placement for a robot in social group interactions. There are also many works such as [3,16,30–33] that deal with proxemics, F-formations, and MRP systems, but not many works have studied the teleoperator’s placement in the group and whether they would place themselves according to F-formations. To our knowledge, this is the first study using simulation with multiple F-formation configurations and varying group sizes to study teleoperator performance in MRP systems.

3. Simulation Environment for MRP Systems

For the experiment, we created a simulated environment of a conference lobby consisting of a large hall with a red carpet, painted walls, windows, entrance door, round pub-style tables, two sofas, and VAs. Google SketchUp [34] was used to design the hall, interiors, and tables, then it was imported into the Gazebo simulator and VAs were placed. TurtleBot [35] was used as a telepresence robot. For teleoperating the robot, a gamepad controller was used. A snapshot of the simulation environment from different angles is shown in Figure 4. The environment was designed to emulate a case of a conference break during which people arrive at the conference lobby to have some coffee and snacks while interacting with other people.

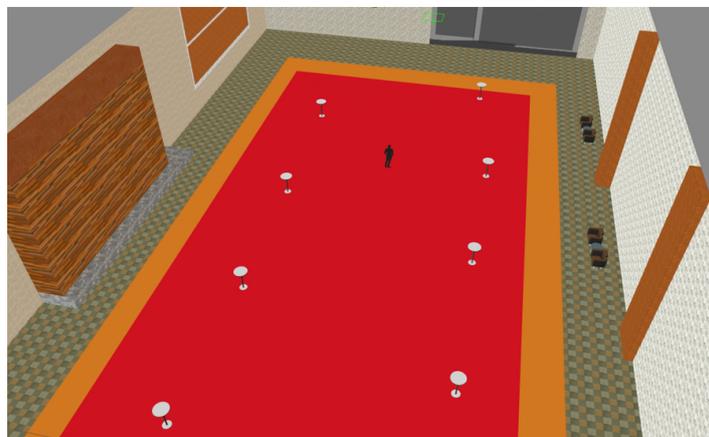


Figure 4. Simulation environment without any virtual agents (VAs).

The conference hall was built without any roof in order to record the global or exocentric view of the experiment. In addition, the hall was large in order to accommodate a larger number of VAs and different F-formations (see Figure 5).

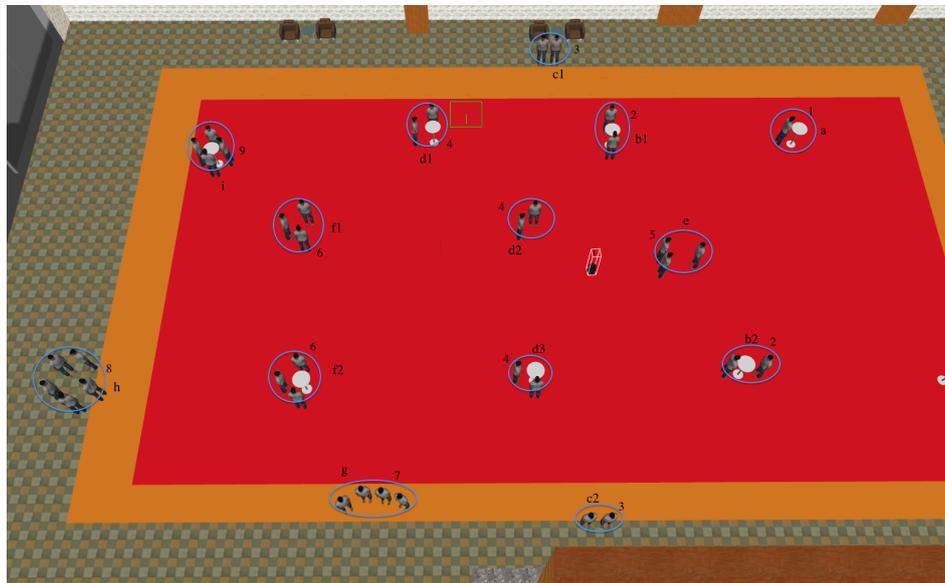


Figure 5. Simulation environment with VAs interacting in different formations and the robot navigating to join the groups. The numbers in the image represent formations as per the points listed in Section 3.

The scene was populated with 37 VAs that were positioned in 14 different formations (Figure 5). The number of formations was selected in order to capture the diverse formations mentioned by Kendon and Marshall while adding some redundancy as has been naturally observed in a conference lobby situation. Tables were placed in a few arbitrary formations. The model of the TurtleBot was modified by increasing the height of the camera to approximately 1.7 m, assuming that giving the teleoperator approximately the same height as a telepresence robot would provide a more realistic experience. The main reasons behind opting for the TurtleBot are that it has a symmetric base and a differential drive scheme, which is relatively easy to teleoperate. All the VAs have the same appearance and do not have any facial expressions or physical behaviors so that the teleoperator would not be biased based on the clothes, height, or gender, etc. The 14 different formations are shown in Figure 5, and the groups in Figure 5 are numbered as per the listed points below:

1. A virtual agent (VA) (singleton) was standing in front of a round table.
2. Two pairs of VAs were conversing in vis-a-vis formations with a table in between them.
3. Two pairs of VAs were standing in side-by-side formations without tables: one pair in front of the wall looking at the pallet wall and another one in front of a pair of sofas facing the whole hall and looking at the VAs around, while conversing.
4. Three pairs of VAs were conversing in L-shape formations: two pairs around a table and one pair without a table with different configurations.
5. Three VAs were conversing in a Triangular formation without a table.
6. Two groups of three VAs were conversing standing in circular formations: one group around a table and another group without a table.
7. Four VAs were conversing in a semi-circular formation without a table and facing the wall's texture.
8. Five VAs were standing in a circular formation without a table.
9. Four VAs were standing around a table in a circular formation.

The first 13 formations mentioned above had one or more free spots beside the VAs where the teleoperator could place the robot. For the last formation, there was no empty spot in the group for the teleoperator to place the robot.

The subjects were provided with an egocentric view of the simulated environment, i.e., the subjects used the robot's camera when teleoperating the robot, as shown in Figure 6.



Figure 6. Image showing VAs interacting with each other and a modified TurtleBot observing. The inside image shows the view from the robot's sensory system. This is the egocentric view seen by the subject while teleoperating a robot.

4. Procedure

The subjects were invited to a room where a laptop with a gamepad controller and an external monitor were installed. The subjects used the laptop's screen and the gamepad controller to teleoperate the robot. The researcher located in the room used the external monitor, which was also connected to the laptop, to record the exocentric view of the experiment. The subjects could not see the external monitor while teleoperating the robot. They navigated the robot using an egocentric view window, as shown in Figure 6.

To teleoperate the robot, the direction buttons (D-pads) were used for movement and rotation. The movement speed was set to 0.5 m/s. and the rotation speed was set to 1 rad/s.

The experimental procedure was divided into five steps:

1. **Information to subjects:** The researcher started the experiment by explaining the process (cover story) to the subject as follows: "This is a conference lobby and you are a staff member from the conference. You are supposed to interview all the people or groups in the lobby'." Then, after showing the egocentric view window, the researcher said, "This is the view from the robot's camera, which is used by you to observe the scene for the rest of the experiment, and you will

- be able to operate the robot from this controller. You have to interview all the people or groups in the lobby, and when you are done interviewing one of them, inform or notify me. After my consent, you could proceed to the next person or group and then interview them. In this case, interviewing means you place the robot in a spot that you think is suitable and let me know. We continue this process until you have visited all the people or groups”.
2. **Socio-demographic questionnaire:** The subject filled out the consent form and socio-demographic questionnaire.
 3. **Teleoperation practice:** The subject practiced teleoperating the robot in the simulated environment without any VAs (see Figure 4).
 4. **Teleoperation interaction:** The subject teleoperated the robot and conducted interviews with the VAs in Figure 5. This interaction was done by joining the groups, i.e., the subject placed the robot according to his/her convenience and informed the researcher that the robot had been placed and continued until he/she thought that all people or groups in the scenario had been visited.
 5. **Questions after experiment:** His/her recording from the exocentric view was shown to the subject, who then filled out the questionnaire “Questions after experiment”.

There was no risk of harm to the subjects in the study, and the study does not reveal any details about the subjects or which subject is teleoperating the robot. The experiment was conducted in English and the researcher explained questions to the subjects when needed. The two questionnaires are discussed in Section 4.1. All questions are listed in Appendixes A and B. The average time consumed was 4 min for practicing (step 3) and 20 min for interacting with the VAs (step 4); there was no time limit in either of the steps.

4.1. Data Collection

The data collected during the experiment include the following:

- Socio-demographic questionnaire;
- The video recordings of the exocentric view and the egocentric view in the simulated environment;
- Rosbag data acquisition mechanism used to record the length of the path;
- Questions after the experiment.

At step 2, the subjects were asked to fill out a socio-demographic questionnaire. The questionnaire consisted of questions about age, gender, and questions related to their experience with robots and different controllers for navigating a robot (i.e., joystick, game controller, and keyboard), how often they teleoperate a robot, and their technical expertise in navigating a robot. See Appendix A for further information on the socio-demographic questionnaire.

The length of the path and the path chosen by the subject to complete the experiment were recorded using the rosbag data acquisition mechanism. The exocentric view recording was analyzed in order to obtain the time taken by each person for a complete run, the number of times they ran into obstacles, and finally, the placement of the robot in a group. This recording was also used to evaluate the formations made by the teleoperator.

At step 5, the subjects were asked to fill out “Questions after experiment”, which contained questions that allow the subjects to provide their perspective on the experiment and the choices they made when joining the groups. See Appendix B for further information about the questionnaire. Finally, the subjects answered open questions: why they chose that particular spot and whether they would change their spot after observing the global view (the recorded exocentric view) for each group joined.

4.2. Method and Metrics to Evaluate our Experiment

To evaluate our experiment, a game-theoretic model [23] that considers the position and orientational information of people to detect groups in the scene was used .

This approach uses a quantitative model to detect groups by modeling the socio-psychological concept of F-formations and the biological constraints of social attention. A 2D frustum for each person in the scene is generated where the frustum is the field of view (FoV) and locus of attention for a given body orientation. Each frustum is modeled as a 2D (x and y position in the ground plane) Gaussian distribution in which each of the dimensions are generated separately and the parameter l corresponds to the variance of the Gaussian distribution centered around the location of the person. An example of a frustum can be seen in Figure 7. The Gaussian distribution is discussed in detail in [23].

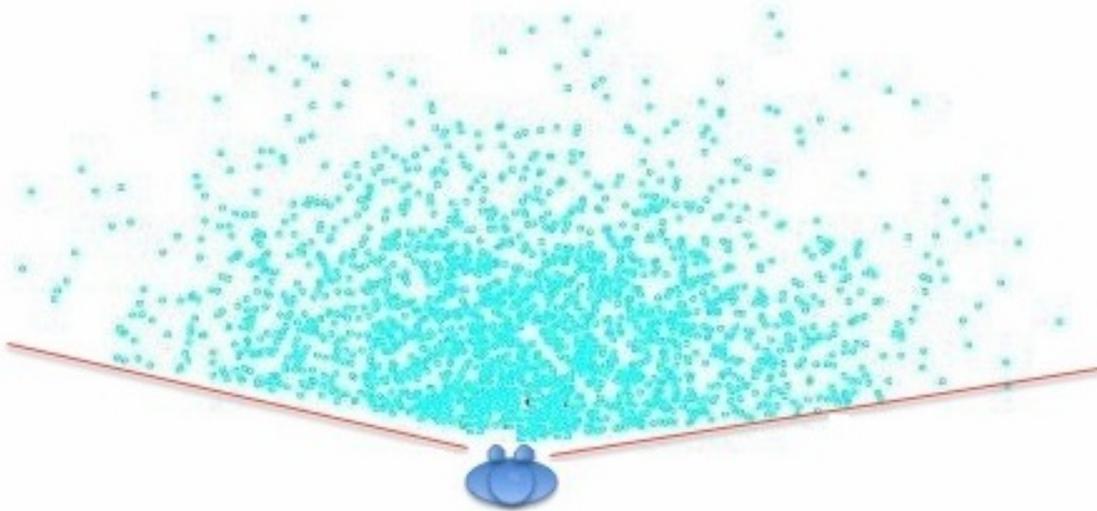


Figure 7. The frustum of a person. The particles (Gaussian samples) are generated to model the frustum of the person in his/her facing direction.

Given a person at $p(x,y)$ with head orientation θ , a sample $s(x,y)$ is inside the frustum if

$$\arccos\left(\frac{s \cdot f_L}{\|s\| * l}\right) \leq \frac{\alpha}{2}, \quad (1)$$

where $f_L = \{\cos(\theta) * l, \sin(\theta) * l\}$ is the line of symmetry of the frustum,

l = the variance of the Gaussian distribution,

α = aperture, and

θ = angle.

This sampling process is iterated keeping only the samples that fall within a pre-specified cone of aperture (160°).

Each person is modeled using their frustum represented as a 2D histogram. Secondly, a pairwise affinity matrix is computed for each person, and finally, F-formations are extracted using evolutionary stable strategy clusters. Please refer to [23] for more details.

In our work, position and orientation information of the VAs and robot is used. The approach generates a 2D frustum for each VAs and the robot in the group. Then, the affinity matrix is computed and the formations are extracted using a clustering strategy if the frustums of VAs overlap with the robot's frustum, i.e., the VAs and robot belong to the same group if their frustums overlap. The approach provides quantitative values and a resulting image that can be evaluated qualitatively, which are further discussed in Section 5.

Analyzing groups, [19,22–24] compute accuracy in terms of precision and recall. They consider a group as correctly estimated if at least $\lfloor T \cdot |G| \rfloor$ of their members are found by the grouping method and if no more than $1 - \lfloor T \cdot |G| \rfloor$ false members are identified, where $|G|$ is the cardinality of the labeled group G and $T \in [0, 1]$ is an arbitrary threshold, called the tolerance threshold. In particular, they focus on two interesting values of T : $2/3$ and 1 . $T = 2/3$ corresponds to finding at least $2/3$ of the members of a group, no more than $1/3$ false members. $T = 1$ means that a group is detected if all of the tracked members of a group are automatically labeled. With these metrics, they determine for each frame the correctly detected groups (true positives—TP), the misdetected groups (false negatives—FN), and the hallucinated groups (false positives—FP), with which they compute the standard pattern recognition metrics precision and recall:

$$precision = \frac{TP}{TP + FP}, \quad recall = \frac{TP}{TP + FN},$$

and the $F1$ score defined as the harmonic mean of precision and recall:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}.$$

4.3. Subjects

In total, 22 subjects participated in the study, among which 16 (72.7%) were male and 6 (27.3%) were female. Thirteen (59.1%) subjects were between 20 and 30, 8 (36.4%) subjects were between 30 and 40, and one was between 40 and 50. Thirteen subjects (59.1%) had prior experience with robots. Regarding experience with tools such as joysticks or game controllers, twelve subjects used them occasionally, five subjects never used them, three subjects had used them one time, and two subjects often used them, which can be seen in Figure 8. On average, the subjects rated their expertise in navigating a robot as $\mu = 4.05$ ($\sigma = 1.72$) on a 1–7 Likert scale where 1 = low and 7 = high. They also rated the difficulty in choosing a spot to place the robot in the group, which can be seen in Figure 9. The subjects were international researchers from different fields such as artificial intelligence, robotics, biology, statistics, human geography, and more. The main reasons for recruiting researchers for this study were their fluency in English and familiarity with the conference lobby scenario from their participation at international conferences. The whole experiment was conducted in English. No reward was given on completion of the experiment.

As mentioned before, familiarity of the conference lobby scenario was one of the criteria for recruiting researchers for this study. The population includes subjects from different age groups and different familiarity with robot teleoperation and controllers. Two age groups are underrepresented in the study: older adults and teenagers. Moreover, there were significantly fewer female subjects. These facts need to be taken into account when interpreting the results. Previous research [36] has shown that people with rich gaming experience can perform significantly better than people with less gaming experience in robot teleoperation scenarios. Familiarity with the conference lobby scenario might have also contributed to the results, and it might be expected that subjects unfamiliar with it would demonstrate different performance.

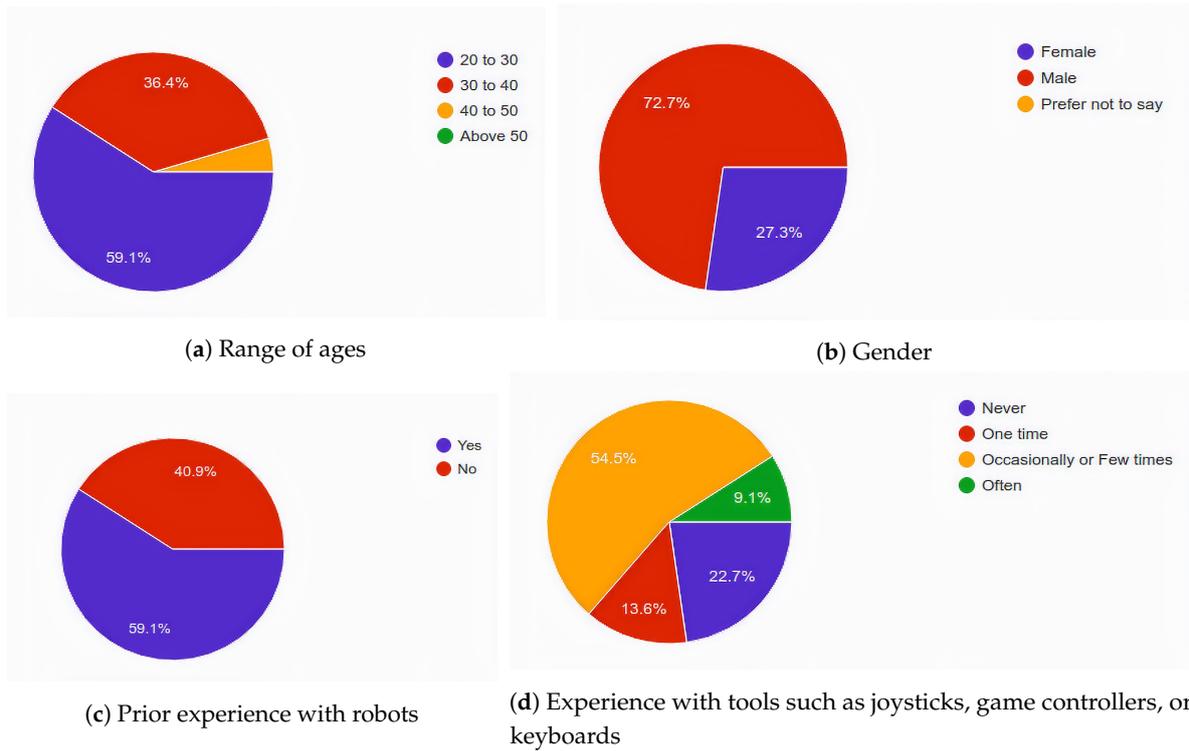


Figure 8. Information on subjects.

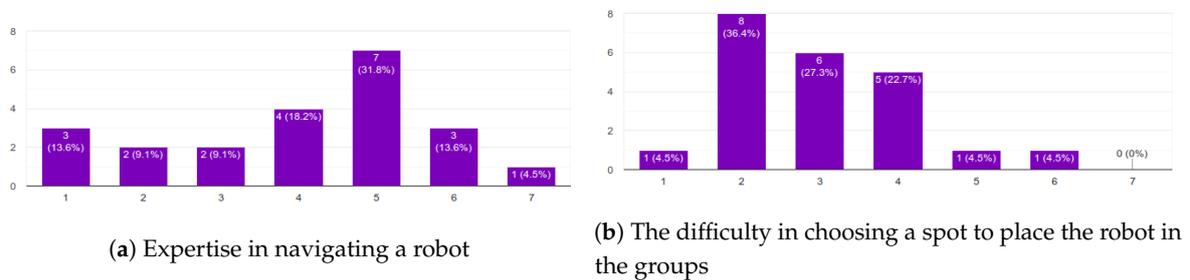


Figure 9. Subjects answered on a Likert scale of 1 to 7. 1 = easy or low and 7 = difficult or high.

5. Results and Discussions

F-formations are helpful in understanding social interaction and increasing the quality of interaction in both HHI and HRI. While investigating the research questions R1 & R2, we came across different types of results and observations, which are discussed in three parts as follows:

- Modified formations after the placement of the robot in the groups;
- F-formations while joining the social group interactions;
- Autonomous Feature.

5.1. Modified Formations after the Placement of the Robot in the Groups

Generally, in social group interactions, members of the group tend to place themselves in reachable space to others, which are called transactional segments [20], thus making formations. In the presented experiment, subjects joining the group perceive the present formations in which the VAs are standing and locate an empty spot in the group where they could place the robot. Considering this, we framed a question as Why did you place the robot in that particular position? (See Appendix B, Question 5.) Subjects could select multiple answers. For this, 15 (68%) subjects stated their reason as to see the

faces of the VAs, 12 (54%) subjects mentioned that it was an ideal position to place the robot, 4 subjects mentioned the upper body view, and 2 subjects mentioned that it was to see shoulders. This suggests that a greater number of people were interested in watching the faces of VAs when joining the group. In this process, the teleoperator used the egocentric view from the robot's RGB camera. This was intended to provide a viewpoint similar to real-life experience. Figure 9 shows that 20 (90%) subjects found it easy to select the spot. The placement of the robot in the group added a new person to the formation, and the present formations were modified. New existing formations came into the picture as follows:

- When a single VA was standing, then the teleoperator joined the dyadic interaction in a vis-a-vis formation, only one subject joined in an L-shape formation instead.
- When two VAs were standing in the vis-a-vis formation and the teleoperator joined them, then the formation became circular.
- When two VAs were standing in the L-shape formation and the teleoperator joined them, then the formation became circular.
- When two VAs were standing in the side-by-side formation in front of the wall, the teleoperator joined them sideways and the interaction still remained a side-by-side formation. This happened often, but in 4 cases, the teleoperator placed the robot in front of them, transforming the formation into a circular one.
- When two VAs were standing in the side-by-side formation facing the open space, then the teleoperator joined them from the front and the formation became a triangle.
- When three VAs were standing in the triangle formation and the teleoperator joined, then the formation became circular.
- When three VAs were standing in the circular formation and the teleoperator joined, the formation still remained circular.
- When four VAs were standing in the circular formation and there was no spot left, all teleoperators approached and placed the robot in between two VAs and the formation still remained circular.
- When four VAs were standing in the semi-circular formation in front of the wall, then the teleoperator approached them from either side and the formation still remained semi-circular. This happened often, but in 2 cases, the teleoperator placed the robot in front of them, transforming the formation into a circular one.
- When five VAs were standing in the *Circular formation* and the teleoperator joined, the formation still remained *Circular*.

There were cases where not all formations were visited due to the scale of the environment. While concentrating on joining the groups, some subjects could not keep track of which formations they had already joined, which resulted in a few unattended formations and a few formations attended twice. The unattended formations were removed from the analysis. In case of formations attended twice, only the first ones were considered for calculations. While joining the formations, only 7 subjects visited all of the 14 formations and 15 subjects visited 10 formations on average. Most of the subjects forgot to attend "d2" and "d3" formations (from Figures 5 and 10), which were placed in the middle of the scene. The similar appearance and the amount of VAs in the scene could have misled the subjects to believe that they had already visited those formations. In total, each subject visited 12 formations and missed out 2 formations, on average, during the experiment.

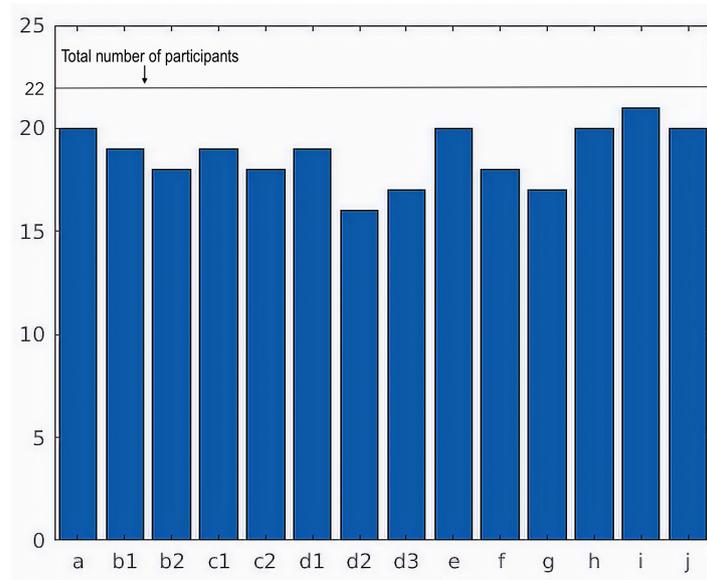


Figure 10. Groups histogram. In the image, each bar illustrates a formation (horizontal axis) and the number of subjects joining that formation (vertical axis). For example, the “a” represents the single VA standing (from Figure 5), 20 subjects joined this formation, and 2 subjects missed this formation.

5.2. F-Formations when Joining the Social Group Interactions

In order to verify **R1** (While joining the social interactions, do teleoperators adhere to (respect) F-formations?), the F-formations were analyzed after the robot was placed in the groups. The spatial information (X , Y) and orientation information (θ) of the VAs and the robot in the formations were extracted and fed into Vascon et al.’s algorithm [23]. This algorithm considers the position and orientation information of the VAs and robot and generates the 2D frustum for all of them. The algorithm computes the affinity matrix for every member of the group and then extracts formations using a clustering process. Finally, the algorithm provides a quantitative result (precision, recall, and F-measure) and qualitative result (visualization). In our study, when the VAs and robot information were provided to the algorithm, it produced results in terms of precision, recall, and F-measure scores each equal to 1 using the original values of the algorithm $l = 20$, aperture = 160, and Gaussian samples = 2000. This indicates that the robot was part of the formation when the teleoperator joined the group. The qualitative analysis resulted in the visualization showing the frustums of VAs and the robot, which shows that the robot was part of the formation with virtual agents, as seen in Figure 11. The values used in the algorithm for visualization in Figure 11 were $l = 1$, aperture = 6, and Gaussian samples = 200.

After the teleoperation session, the subjects were shown the exocentric view and interviewed by asking whether the placement of the robot is actually where they wanted to place it. All subjects selected a spot in the P-space around the group of VAs as an intended spot. They also mentioned that “this spot would be more suitable to talk to VAs in groups’.” While teleoperating the robot using an egocentric view, no subject exploited the O-space of the group and everyone placed the robot either in the P-space or in the R-space. When joining the interaction, 5 subjects placed the robot in R-space, i.e., stood a little farther away from the P-space. When the exocentric view was shown, all of the subjects, including these 5 subjects, selected a spot in P-space to join the group. When asked about the reason behind their placement, they mentioned that “they could not perceive the distance precisely through the egocentric view, and also the technical difficulties in controlling the robot were a contributing factor.” There was one situation where the subject had to move around the group and place the robot in the empty spot. Most of the subjects moved around and placed the robot, but 2 subjects did not. They placed the robot in between two VAs, even when there was not much space, and when asked about the reason for this placement while being shown the exocentric view, 1 subject mentioned that

“I did not see the empty spot from that position with an egocentric view”, and the other said “I got lazy to move around and place the robot in that empty spot”. Observation of the overall process suggests that subjects demonstrated a strong tendency to respect F-formations when joining the social group interactions.

From this analysis and by interviewing the subjects while showing the exocentric view, we conclude that teleoperators do respect F-formations when joining the groups.

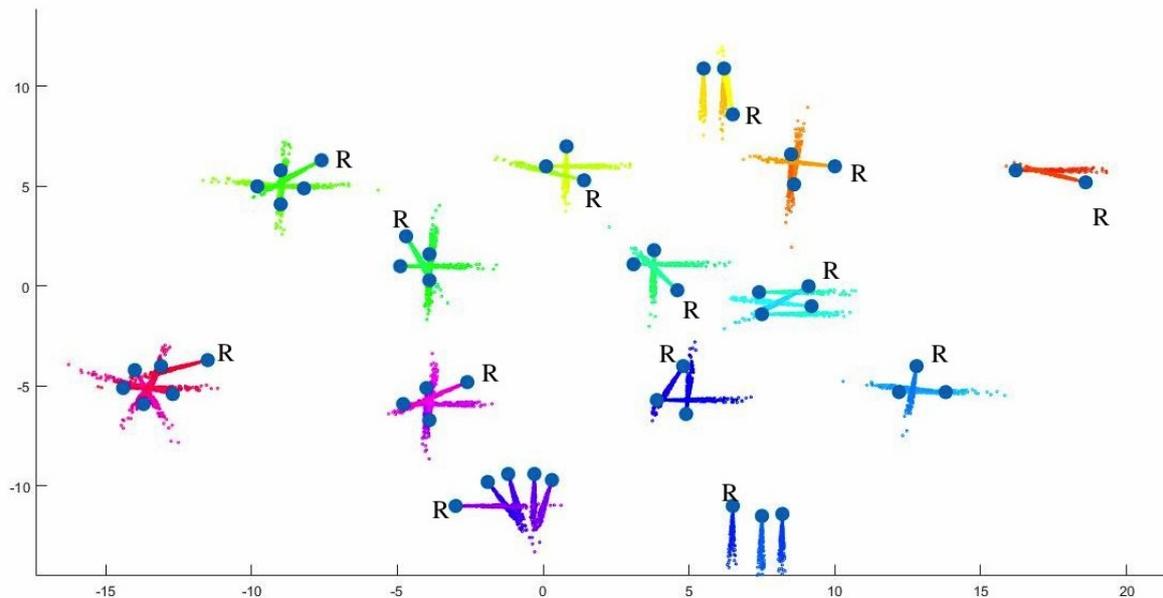


Figure 11. Analysis of F-formations using Vascon's method [23]. The image depicts the VAs and the robot in the scene. The data are taken from one of the subjects. In our model, the parameter values of $l = 20$, aperture = 160, and Gaussian samples = 2000 were used for F-formation analysis, and $l = 2$, aperture = 60 and Gaussian samples = 200 were used for the illustration.

5.3. Autonomous Feature

In the MRP domain, there are studies on autonomous features for telepresence robots [31,37]. These works are focused on “walk and talk” situations between an individual and a robot. The current study has demonstrated that although the subjects intend to adhere to F-formations, their ability to do so is limited by the robot. Therefore, the ability of a robot to join F-formations autonomously could be helpful in many scenarios, such as finding the most appropriate spot and decreasing the operator's workload. Considering this, subjects were asked whether they would prefer to have an autonomous navigation function that would place the robot into the group automatically. Of 22 subjects, 19 subjects (86.4%) responded affirmatively, and the remaining 3 subjects (13.6%) expressed concerns about the robot's technical ability to autonomously join formations. The 3 subjects' concerns are valid as there is a gap in the research developed and the ability to apply this to a mobile robot in real-time human environments [38]. At the same time, researchers are also working on this process of robots joining the social interaction [29,39] to interact with people or help them in real time with mobile robots. This implies that there is a need to develop such autonomous features that could integrate robots into human social environments.

Finally, it is important to note that the selection of the robot, i.e., MRP system, and its FoV are important factors when analyzing the results. It is true that the FoV varies between MRP systems, i.e., different telepresence robots have different cameras with different FoVs. Some telepresence systems have a smaller FoV, and others have up to 150 degrees in FoV. It is indeed true that the selected FoV used in the simulation could have impacted the results obtained in this paper. In the current experiment, we used Turtlebot with a Kinect RGB camera, and we did not make any explicit changes to

the robot's FoV, i.e., the default settings were used. If a different robot with a different FoV was used in the experiment, a change might be noticed in the navigation pattern of the robot and the way subjects approached the groups. This is indeed worth investigating in future studies, especially studies whose aim is to validate the design of MRP systems. There is also a possibility that a slight change could have been observed in the selection of the spot in the group, i.e., slightly closer to or farther from the group. However, our main claim, i.e., subjects place the robot in the group adhering to F-formations, was the focus of this work. When subjects were interviewed while being shown the exocentric view, all subjects opted for a similar spot upon which to place the robot. This spot was in the P-space of the group. It should also be noted that in our experiment, we were not observing the navigation pattern or the way subjects approached the group, i.e., we observed only the final placement of the robot. In one of our previous studies [40], we observed that limiting the horizontal view increased the quality of interaction.

6. Conclusions and Future Works

In this paper, we conducted a study in order to investigate two research questions, **R1** & **R2**. From this study, using quantitative and qualitative results and by interviewing the subjects while showing the exocentric view, we found that teleoperators do respect F-formations when joining the groups and also prefer an autonomous feature to join the social interaction. We also observed that when the robot joined the groups, the formations present were modified into a new existing formation, which could be an interesting direction in future for social robots.

In the future, a few works could be planned based on this study. Firstly, we would develop an approach for mobile robots to join social interactions automatically, adhering to F-formations [29]. Secondly, an experiment with an equal number of males and females with similar ethnicities could be devised to observe how far they stand from the group or how will they join the group, i.e., to understand the difference between male and female placement of the robot to join the social group interaction.

Author Contributions: Conceptualization, S.K.P., A.K. (Andrey Kiselev), A.K. (Annica Kristoffersson), and A.L.; Data curation, S.K.P.; Formal analysis, S.K.P.; Investigation, S.K.P.; Methodology, S.K.P.; Resources, S.K.P.; Software, S.K.P.; Supervision, A.K. (Andrey Kiselev), A.K. (Annica Kristoffersson), and A.L.; Validation, S.K.P.; Visualization, S.K.P., A.K. (Andrey Kiselev), A.K. (Annica Kristoffersson), and A.L.; Writing—original draft, S.K.P.; Writing—review & editing, A.K. (Andrey Kiselev), A.K. (Annica Kristoffersson), and A.L.; lead of the project work, A.L.

Funding: Örebro University is funding the research through the Successful Ageing Programme. The statements made herein are solely the responsibility of the authors.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Appendix A. Socio-Demographic Questionnaire

1. What is your age? [20–30, 30–40, 40–50, Above 50]
2. What is your gender? [Female, Male, Prefer not to say]
3. What is your research area in robotics? [Interaction, Mechanics, Cognition, Artificial Intelligence, Perception, Control, Other (with blank space)]
4. Do you have any prior experience with robots? [Yes, No]
5. Did you ever operate a telepresence robot? [Never, One time, Occasionally or a few times, Often]
6. Did you ever use tools such as a joystick, game controller, or keyboard to drive a robot? [Never, One time, Occasionally or a few times, Often]
7. On a scale of 1–7, how do you rate your technical expertise to navigate a robot? [1 = low, 7 = high]

Appendix B. Questions after the Experiment

1. What do you think of the scene? [Bad, Ok, Good, Better, Best]
2. Did you see any flaws in the scene? [Yes, No, Not sure]

3. If yes, are they important to this particular experiment? [Important, Can be ignored, Will not be noticed by the user]
4. If important, mention the flaws. [blank space]
5. Why did you place the robot in that particular position? [I thought it was an ideal position to be placed; So that I could see the face/s of the person/s; So that I could see the shoulders of the person/s; So that I could see the upper part of the body; Other (with blank space)]
6. How hard was it to select the spot? [Likert scale: 0 = low, 7 = high]
7. How hard was it to drive the robot? [Likert scale: 0 = low, 7 = high]
8. Would you prefer to have a button to place the robot in groups or would you prefer to do it yourself? [Prefer a button, Prefer to place myself, Other (with blank space)]
9. If “other”, why? [blank space]
10. Would you prefer an autonomous feature that navigates around and places the robot in the group? [Yes, No, Other (with blank space)]
11. After seeing the global view, would you change your spot if given a chance? [Yes, No, Maybe]
12. If yes, why? [blank space]

References

1. Michaud, F.; Boissy, P.; Labonte, D.; Corriveau, H.; Grant, A.; Lauria, M.; Cloutier, R.; Roux, M.-A.; Iannuzzi, D.; Royer, M.P. Telepresence robot for home care assistance. In Proceedings of the AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics, Palo Alto, CA, USA, 26–28 March 2007; pp. 50–55.
2. Michaud, F.; Boissy, P.; Labonté, D.; Briere, S.; Perreault, K.; Corriveau, H.; Grant, A.; Lauria, M.; Cloutier, R.; Roux, M.-A.; et al. Exploratory design and evaluation of a homecare teleassistive mobile robotic system. *Mechatronics* **2010**, *20*, 751–766. [[CrossRef](#)]
3. Tsui, K.M.; Desai, M.; Yanco, H.A.; Uhlik, C. Telepresence robots roam the halls of my office building. In Proceedings of the 1st Workshop on Social Robotic Telepresence co-located with 6th ACM/IEEE international Conference on Human-Robot Interaction (HRI), Lausanne, Switzerland, 6–9 March 2011; pp. 58–59.
4. Kristoffersson, A.; Coradeschi, S.; Loutfi, A. A Review of Mobile Robotic Telepresence. *Adv. Hum.-Comput. Interact.* **2013**. [[CrossRef](#)]
5. Bailenson, J.N.; Blascovich, J.; Beall, A.C.; Loomis, J.M. Interpersonal distance in immersive virtual environments. *Personal. Soc. Psychol. Bull.* **2003**, *29*, 819–833. [[CrossRef](#)] [[PubMed](#)]
6. Hall, E.T. *The Hidden Dimension*; Doubleday: Garden City, NY, USA, 1966.
7. Cristani, M.; Paggetti, G.; Vinciarelli, A.; Bazzani, L.; Menegaz, G.; Murino, V. Towards computational proxemics: Inferring social relations from interpersonal distances. In Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Boston, MA, USA, 9–11 October 2011; pp. 290–297.
8. Kendon, A. Spacing and orientation in co-present interaction. In *Development of Multimodal Interfaces: Active Listening and Synchrony*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 1–15.
9. Kendon, A. *Conducting Interaction: Patterns of Behavior in Focused Encounters*; Studies in Interactional Linguistics; Cambridge University Press: Cambridge, UK, 1990.
10. Kristoffersson, A.; Eklundh, K.S.; Loutfi, A. Measuring the Quality of Interaction in Mobile Robotic Telepresence: A Pilot’s Perspective. *Int. J. Soc. Robot.* **2013**, *5*, 89–101. [[CrossRef](#)]
11. Vroon, J.; Joosse, M.; Lohse, M.; Kolkmeier, J.; Kim, J.; Truong, K.; Englebienne, G.; Heylen, D.; Evers, V. Dynamics of social positioning patterns in group-robot interactions. In Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, Kobe, Japan, 31 August–4 September 2015; pp. 394–399.
12. Giraff. Available online: <http://www.giraff.org/> (accessed on 15 October 2019).
13. Walters, M.L.; Dautenhahn, K.; Boekhorst, R.T.; Koay, K.L.; Syrdal, D.S.; Nehaniv, C.L. An empirical framework for human-robot proxemics. In Proceedings of the New Frontiers in Human-Robot Interaction, Edinburgh, Scotland, 6–9 April 2009.

14. Mead, R.; Mataric, M. Autonomous human-robot proxemics: Socially aware navigation based on interaction potential. *Auton. Robot.* **2017**, *41*, 1189–1201. [[CrossRef](#)]
15. Mumm, J.; Mutlu, B. Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In Proceedings of the 6th International Conference on Human-Robot Interaction, Lausanne, Switzerland, 6–9 March 2011; ACM: New York, NY, USA, 2011; pp. 331–338.
16. Walters, M.L.; Oskoei, M.A.; Syrdal, D.S.; Dautenhahn, K. Along-term human-robot proxemic study. In Proceedings of the 2011 RO-MAN, Atlanta, GA, USA, 31 July–3 August 2011; pp. 137–142.
17. Marshall, P.; Rogers, Y.; Pantidi, N. Using F-formations to analyse spatial patterns of interaction in physical environments. In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, Hangzhou, China, 19–23 March 2011; ACM: New York, NY, USA, 2011; pp. 445–454.
18. Serna, A.; Tong, L.; Tabard, A.; Pageaud, S.; George, S. F-formations and collaboration dynamics study for designing mobile collocation. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Florence, Italy, 6–9 September 2016; ACM: New York, NY, USA, 2016; pp. 1138–1141.
19. Cristani, M.; Bazzani, L.; Paggetti, G.; Fossati, A.; Tosato, D.; Bue, A.D.; Menegaz, G.; Murino, V.; Fossati, A.; del Bue, A. Social interaction discovery by statistical analysis of F-formations. In Proceedings of the British Machine Vision Conference (BMVC), Dundee, UK, 29 August–2 September 2011.
20. Hung, H.; Kröse, B. Detecting F-formations as dominant sets. In Proceedings of the 13th International Conference on Multimodal Interfaces, Alicante, Spain, 14–18 November 2011; pp. 231–238.
21. Pavan, M.; Pelillo, M. Dominant sets and pair wise clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 167–172. [[CrossRef](#)] [[PubMed](#)]
22. Setti, F.; Russell, C.; Bassetti, C.; Cristani, M. F-formation detection: Individuating free-standing conversational groups in images. *PLoS ONE* **2015**, *10*, 2015.
23. Vascon, S.; Mequanint, E.Z.; Cristani, M.; Hung, H.; Pelillo, M.; Murino, V. A game-theoretic probabilistic approach for detecting conversational groups. In Proceedings of the Asian Conference on Computer Vision, Singapore, 1–5 November 2014; Springer International Publishing: Cham, Switzerland, 2014; pp. 658–675.
24. Ricci, E.; Varadarajan, J.; Subramanian, R.; Buló, S.R.; Ahuja, N.; Lanz, O. Uncovering interactions and interactors: Joint estimation of head, body orientation and f-formations from surveillance videos. In Proceedings of the IEEE Computer Society, Washington, DC, USA, 7–23 December 2015; pp. 4660–4668.
25. Zhang, L.; Hung, H. Beyond F-formations: Determining Social Involvement in Free Standing Conversing Groups from Static Images. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
26. Johal, W.; Jacq, A.; Paiva, A.; Dillenbourg, P. Child-robot spatial arrangement in a learning by teaching activity. In Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016; pp. 533–538.
27. Vázquez, M.; Steinfeld, A.; Hudson, S.E. Parallel detection of conversational groups of free-standing people and tracking of their lower-body orientation. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 3010–3017.
28. Pathi, S.K.; Kiselev, A.; Loutfi, A. Estimating f-formations for mobile robotic telepresence. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; ACM: New York, NY, USA, 2017; pp. 255–256.
29. Pathi, S.K.; Kristofferson, A.; Kiselev, A.; Loutfi, A. Estimating Optimal Placement for a Robot in Social Group Interaction. In Proceedings of the 2019 28th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New Delhi, India, 14–18 October 2019.
30. Fiore, S.M.; Wiltshire, T.J.; Lobato, E.J.C.; Jentsch, F.G.; Huang, W.H.; Axelrod, B. Toward understanding social cues and signals in human-robot interaction: Effects of robot gaze and proxemic behavior. *Front. Psychol.* **2013**, *4*, 859. [[CrossRef](#)] [[PubMed](#)]
31. Desai, M.; Tsui, K.M.; Yanco, H.A.; Uhlik, C. Essential features of telepresence robots. In Proceedings of the 2011 IEEE Conference on Technologies for Practical Robot Applications, Woburn, MA, USA, 11–12 April 2011; pp. 15–20.

32. Vázquez, M.; Carter, E.J.; McDorman, B.; Forlizzi, J.; Steinfeld, A.; Hudson, S.E. Towards robot autonomy in group conversations: Understanding the effects of body orientation and gaze. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; ACM: New York, NY, USA, 2017; pp. 42–52.
33. Tsui, K.M.; Desai, M.; Yanco, H.A.; Uhlik, C. Exploring use cases for telepresence robots. In Proceedings of the 6th International Conference on Human-Robot Interaction, Lausanne, Switzerland, 6–9 March 2011; ACM: New York, NY, USA, 2011; pp. 11–18.
34. Sketchup. Available online: <https://www.sketchup.com/> (accessed on 15 October 2019).
35. Turtlebot. Available online: <https://www.turtlebot.com/turtlebot2/> (accessed on 15 October 2019).
36. Takayama, L.; Marder-Eppstein, E.; Harris, H.; Beer, J.M. Assisted driving of a mobile remote presence system: System design and controlled user evaluation. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 1883–1889.
37. Cosgun, A.; Florencio, D.A.; Christensen, H.I. Autonomous person following for telepresence robots. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 4335–4342.
38. Taylor, A.; Riek, L.D. Robot perception of human groups in the real world: State of the art. In *AAAI Fall Symposium Series*; AAAI Press: Palo Alto, CA, USA, 2016; p. 366.
39. Tseng, S.H.; Chao, Y.; Lin, C.; Fu, L.C. Service robots: System design for tracking people through data fusion and initiating interaction with the human group by inferring social situations. *Robot. Auton. Syst.* **2016**, *83*, 188–202. [[CrossRef](#)]
40. Kiselev, A.; Kristoffersson, A.; Loutfi, A. The effect of field of view on social interaction in mobile robotic telepresence systems. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, Bielefeld, Germany, 3–6 March 2014; ACM: New York, NY, USA, 2014; pp. 214–215.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).