Conversational Role Coordinator for Groups in Human-Robot Interaction

Karen Tatarian[1,2], Marine Chamoux[1], Amit Kumar Pandey[3], and Mohamed Chetouani[2]

[1] Software Department, SoftBank Robotics Europe, Paris, France
{name.surname}@softbankrobotics.com
[2] Sorbonne Universite, Institude for Intelligent Systems and robotics, CNRS UMR 7222, 75005 Paris, France
{name.surname}@sorbonne-universite.fr
[3] Hanson Robotics, Hong Kong, China SAR
amit.pandey@hansonrobotics.com

Abstract. With more social robots entering different industries such as educational systems, health-care facilities, and even airports, it is important to tackle problems that may hinder high quality interactions in a wild setting including group conversations. This paper introduces a footing framework that allows the robot to assign conversational roles, which include addressee/addresser, bystander, and overhearer, for multi-participants in an interaction. Accordingly, the robot adjusts its gaze pattern as a social cue that indicates that it has understood these roles. A pilot study was conducted to evaluate the effect of having this footing framework on the group affiliation and perceived social intelligence of a social robot.

Keywords: Social Intelligence · Proxemics · Footing · Situational Assessment.

1 Introduction

Among research on human-robot-interaction there is increasing interest in how robots should act in order to be perceived as socially intelligent. Social intelligence can be comprised of many factors, such as how the robot acts, speaks, and gestures. However, one element of robot social intelligence that has so far received little attention is proxemics. In human-human interaction, proxemics is defined as the study of the human use of space and its effects on the behavior, communication and social interactions [3]. A related concept is footing, which is used to describe the set-up of a conversation and the role of the participants in it [2]. For example, participants can be either a speaker, listener, or bystander (not actively participating).

In human robot interaction, a social robot is not an isolated agent but rather an active participant in social interactions and conversations. As such, it needs to be able to identify the different roles of participants. In addition, it should

be able to recognize its own role and adapt accordingly, whether as a speaker, listener, or bystander.

Accordingly, this paper presents a footing framework that coordinates the conversational roles for multi-party human robot interaction. The system takes into account the position, orientation, gaze, and speech of the participants and combines them with the governing theories of proxemics for situational assessment. The final output is the status of all participants around the robot in addition to the status of the robot itself. A study was conducted to better understand whether footing knowledge and status coordination for multi party interactions increase group affiliation and perceived social intelligence of a social robot.

2 Related Work

For a robot to be perceived as a socially-intelligent agent, it must be able to hold a successful social interaction, adapt to the social environment, and exhibit appropriate multi-modal behavior. For this reason, the robot is required to have the skills to understand and reason about the environment, not just from the perspective of the physical locations of the objects, but also from that of the 'mental' and 'physical' states of the human participants in the interaction [8].

There are three levels of *situation awareness*. The first level is perceiving the state of the elements found in the environment. The second level is building a goal the user wants to achieve based on the comprehension of the information inferred from the first level. Finally the third level is projecting on the future based on the perception and comprehension of the current situation [1]. Based on this model, a generalized perception system has been proposed aiming to achieve an effective and more natural multi-modal human robot interaction [8].

In addition, proxemics and footing have been an increasingly studied phenomena in human-agent interactions, where the agent can be either virtual or physical. For instance, for virtual agents, the interest of better understanding footing using nonverbal signals is highlighted in [7] and [9]. In [7], human interactions were studied to model how footing and the different roles of a multi-party conversation affect the behavior of the participants.

3 Background

This section introduces a few definitions needed to better understand the footing framework introduced in Section 4 and represented in Figure 1. In addition, the interpersonal distances, as introduced by Hall [4], suggest that the limit of the *personal distance*, where interaction with friends takes place, is around 1.2 meters and the limit of the *social distance*, which is the space for formal interactions, is around 3.6 meters. However, any distance beyond that is considered a *public space*.

Transactional Segment: In [11], the transactional segment is defined as the half circle around the forward direction of the person with a radius of 2 meter. This identifies an object in that segment as the person's implicit attentional target.

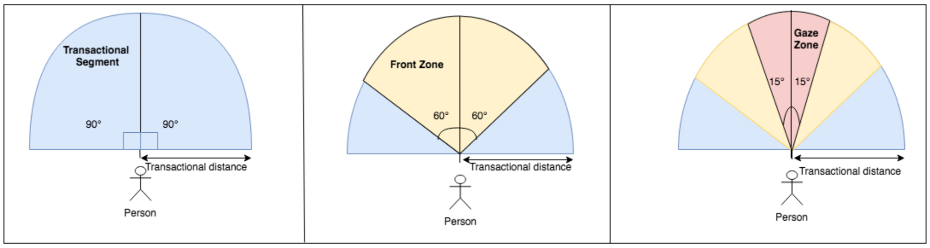

Fig. 1. Representation of the definitions of Transactional Segment, Front Zone, and Gaze Zone respectively.

F-formation: When Kendon in [6] studied organizational patterns of social encounters, he defined the term F-formation, which is when two or more people adjust their spatial and orientation relationship in order to have equal, direct, and exclusive access. Furthermore, Schegloff [10] also linked proxemics to the intentions of the participants.
Front Zone: In estimating the participation status of a person based on observations of human interaction, [11] defined the *front zone* as the area across an angle of 120 degree from the front of the person. Moreover, it was concluded that there is a perceived obligation to participate in a conversation when people are in each other's front zones.
Gaze Zone: When two people fall in each other's *gaze zones* thus having their gazes meet, there is now an obligation for the participation in a conversation [11].

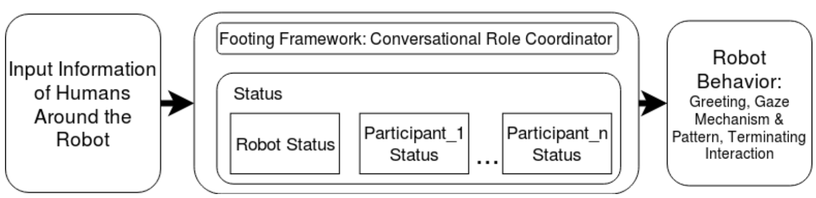

Fig. 2. Overall system

4 Footing Framework

The inputs needed for the framework are the distance between the robot and each participant, the orientation of each participant with respect to the robot, the angle between the robot and each participant, and the attention of each

participant around the robot (whether they are looking straight at robot or to the sides or up or down).

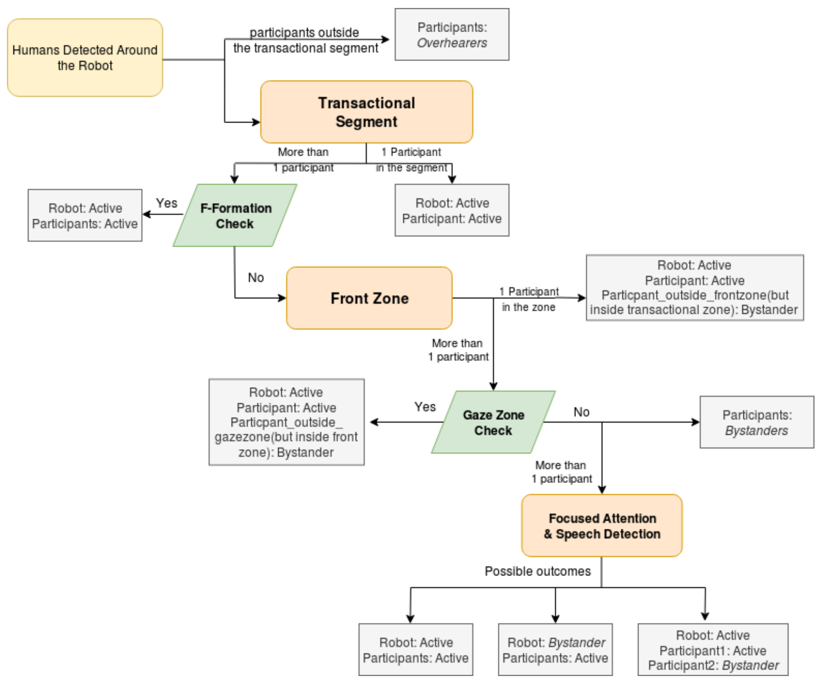

Fig. 3. Framework for Footing in Human-Robot Interaction

First, the transactional segment, defined for this framework, is a half circle in front of the robot with a radius of 1.6 meters. The radius length was adjusted (from the one defined in Section 3) based on the data collected from human interactions with the robot Pepper and in the intention of it being within the inner circle of the social distance defined by [4]. In addition, it is important to have the transactional segment within the field view of the robot. Thus, any participant within the transactional segment is considered to have the intentions to interact (whether as addresser/addressee or bystander) and any participant outside this segment has the role of an *overhearer*. Second when an f-formation is detected based on the position and orientation of the participants, they are directly given the status of *addresser/addressee*. In addition, if there are any participants not part of the f-formation detected but still within the transactional segment, they are given the *bystander* status. Third, for this framework, the *front zone* refers to the 120 degree fan-shaped area with a 1.6 meters in front of the robot. Any participant in this zone can either be an addresser/addressee or a bystander depending on the *gaze zone* priority, but the rest of the participants outside this area now get assigned the *bystander* status. Fourth, the *gaze zone* here is a 30 degree fan-shaped area with a 1.6 meter distance. The participants

who fall in this area get the priority of being assigned the *active* status, which is also referred to addresser/addressee, and all other participants within the transactional segment get the *bystander* status.
Finally, a few rules are added to the framework. First, if at least one participant is assigned the *active* status, the robot gets an *active* status as well. Second, if the direction of orientation of the participants is facing each other and not the robot, the robot is now assigned *bystander* status and must act accordingly. Third, if one participant's direction of orientation is facing the robot while the others do not and speech is detected, the participant facing the robot gains priority of being an addresser/addressee.

5 Evaluation Experiment

A pilot study of 16 participants (average age of 27) was conducted as a preliminary evaluation of the proposed framework for group interactions. The main purpose is to answer the following research question "Does proper footing knowledge and status coordination for multi party interactions increase group affiliation and perceived social intelligence of a social robot?". In this study, a group of two participants plus a robot (forming a triadic interaction) are engaged in a trivia game with the robot that lasted on average 15 minutes. Once the interaction was over, the participants were asked to answer an ALMERE questionnaire. The ALMERE questionnaire was designed to assess the acceptance of assistive social agent technology [5]. The questionnaire includes the following constructs: anxiety (ANX), attitude towards technology (ATT), perceived adaptability (PAD), perceived enjoyment (PENJ), perceived sociability (PS), and social presence (SP). In addition, the positioning and orientation of the participants were also measured to analyze how the dynamics of the group interaction differs when the robot has the footing framework suggested. However, for this paper, the focus is on the results from the ALMERE questionnaire.

5.1 Study Design & Hypothesis

This study employs a 2×3 independent groups design. A behavior with the footing status framework (Condition1) and a control condition with out-of-the-box basic awareness (Condition 2) were manipulated as between groups independent variables. The participants were assigned randomly to the conditions. In addition, there are three situations that occur throughout the interaction: situation with two addressees, situation with one addressee and one bystander, and situation one bystander and one addressee. The situations are manipulated as within group independent variables. The order of the three independent variables in each condition are random to counterbalance. Perceived intelligence and sociability of the robot are measured as dependent variables.
We hypothesize that the robot with the footing framework for conversational roles coordinator will score higher on the ALMERE survey specially for the perceived sociability, social presence and perceived adaptability constructs.

5.2 Set-Up & Experiment

Each interaction required a group of two participants at a time to play a trivia game. The trivia game consists of general knowledge questions to which the players have to answer by either 'True' or 'False'. It is made up of three levels not based on difficulty but rather on whether it is one player or two players in order to simulate the three situations mention in subsection 5.1. The first level is for two players (two addresses) and the second and third levels are for one player where the participants have to switch turns among each other (one addressee and one bystander). During levels two and three, the non-playing participant is invited to observe the other participant play and was given the freedom to move around as he/she wants. At the beginning, the robot would greet the players and at the end of the game say goodbye and thank them for playing.
Moreover, the participants were divided randomly among condition 1 and condition 2 of the experiment. In condition 1, the robot uses the suggested framework in section 4 and adjusts it's gaze patterns accordingly to conform to the roles the participants have chosen based on their position and orientation. In condition 2, Pepper uses its basic awareness API, which is already embedded in it. In basic awareness, the robot studies its own surroundings and once a human is detected, the robot keeps tracking him/her. Basic awareness allows the robot to focus on only one human until another is detected.

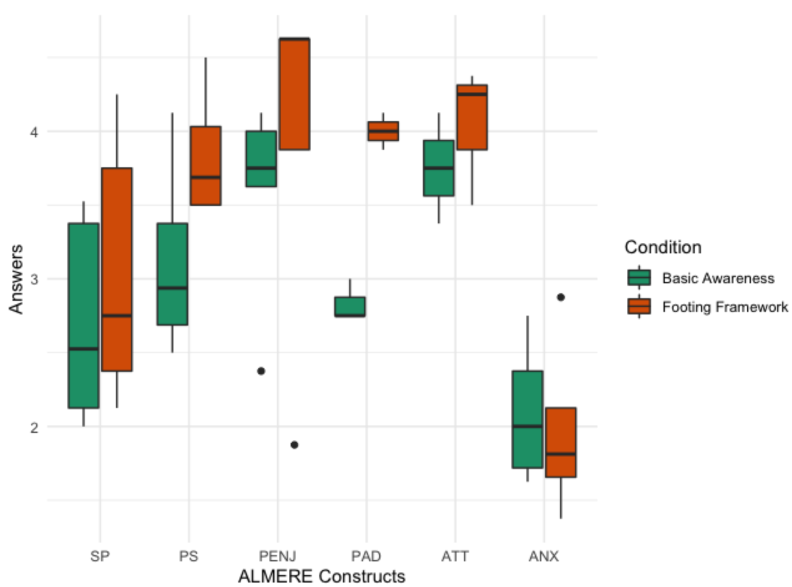

Fig. 4. Overall evaluation of ALMERE questionnaire

5.3 Results

After the trivia game was over, the participants were asked to fill the ALMERE survey. For the overall evaluation score, shown in Figure 4, a repeated measures ANOVA was conducted. A significant main effect was found ($p = 0.0394$) with the condition of the footing framework scoring higher on perceived sociability, social presence, and perceived adaptability. Therefore, the hypothesis is supported.

In addition, the participants were asked to rate an extra set of questions from

Extra Questions
Did you feel attended to by the robot?
Was the robot giving you enough attention?
Did you feel that the robot was looking at you?
Do you think the robot considered you as an important player?

Table 1. Extra questions for evaluation

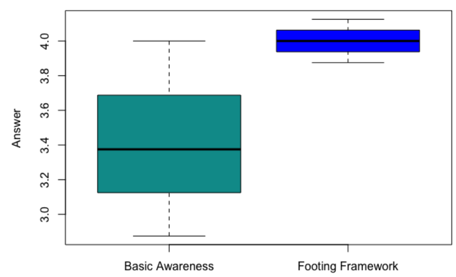

Fig. 5. Result of extra questions

1 (strongly disagree) to 5(strongly agree). The questions are found in Table 1 and Figure 5 shows the evaluation of the result. The robot with the footing framework condition scored much higher than the robot with basic awareness. This shows that the participants who interacted with the robot, which was able to adjust its gaze patterns to their conversational roles, felt more attended to by the robot.

6 Discussion & Conclusion

In the results, we see that the robot for both conditions performed similarly in context of social presence and perceived enjoyment. These results are not surprising since it was the same robot with the same design and features. However, the robot with the footing framework scored higher in the perceived sociability section and showing that the robot with this condition had a more socially intelligent behavior. The robot with the footing framework also scored much higher on perceived adaptability indicating that the robot did adapt and conform to the change in the conversational roles throughout the interaction. Third, when looking at the attitude construct, once again the robot with the footing framework performed better with users rating a more positive feeling and attitude about the application of this robot. Finally, the users also rated that they were less anxious when interacting with the robot with the footing framework.

Moreover, to further gather qualitative insight, the participants were asked to

answer four extra questions shown in Table 1. The ANOVA results, in Figure 5, indicate a significant effect ($p = 0.045$) and the Footing framework scored higher. This gives an indication that in this condition, the users felt more attended to. This work is still in its early stages and with the pilot study results supported the preliminary hypothesis. The next steps include looking at the quantitative measures to better understand how the dynamics of the group interaction with this footing framework differs from one without. It is also interesting to see the effect of having additional social cues in the gaze mechanism, such as turn-taking and floor-holding, and/or proactive adjustment of the robot's position in the conversation. The suggested framework still needs to be optimized for it to be more adaptive and increase the quality of interaction further.

7 Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765955.

References

1. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society **37**, 32–64(33) (March 1995). https://doi.org/doi:10.1518/001872095779049543
2. Goffman, E.: Behavior in public places: notes on the social organization of gatherings. Social theory, Free Press of Glencoe (1963)
3. HALL, E.T.: A system for the notation of proxemic behavior1. American Anthropologist **65**(5), 1003–1026 (1963). https://doi.org/10.1525/aa.1963.65.5.02a00020
4. Hall, E., of Congress), C.P.C.L.: The Hidden Dimension. Anchor books, Doubleday (1966)
5. Heerink, M., Krose, B., Evers, V., Wielinga, B.: Assessing acceptance of assistive social agent technology by older adults: the almere model. I. J. Social Robotics **2**, 361–375 (12 2010)
6. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters. Studies in Interactional Sociolinguistics, Cambridge University Press (1990)
7. Lee, J., Marsella, S.C.: Modeling Side Participants and Bystanders: The Importance of Being a Laugh Track. In: Proceedings of the 11th Conference on Intelligent Virtual Agents. Reykjavik, Iceland (Sep 2011),
8. Pandey, A.K., Gelin, R., Alami, R., Viry, R., Buendia, A., Meertens, R., Chetouani, M., Devillers, L., Tahon, M., Filliat, D., Grenier, Y., Maazaoui, M., Kheddar, A., Lerasle, F., Fitte-Duval, L.: Romeo2 project: Humanoid robot assistant and companion for everyday life: I. situation assessment for social intelligence. CEUR Workshop Proceedings **1315** (11 2014)
9. Pejsa, T., Gleicher, M., Mutlu, B.: Who, me? how virtual agents can shape conversational footing in virtual reality. In: Beskow, J., Peters, C., Castellano, G., O'Sullivan, C., Leite, I., Kopp, S. (eds.) Intelligent Virtual Agents. pp. 347–359. Springer International Publishing, Cham (2017)
10. Schegloff, E.A.: Body torque. Social Research **65**(5), 536–596 (1998), http://www.jstor.org/stable/40971262
11. Shi, C., Shimada, M., Kanda, T., Ishiguro, H., Hagita, N.: Spatial formation model for initiating conversation. In: Robotics: Science and Systems (2011)