Making a Case for Spatial Prompting in Human-Robot Communication

Anders Green
KTH School of Computer Science
100 44 STOCKHOLM
green@csc.kth.se

Helge Hüttenrauch
KTH School of Computer Science
100 44 STOCKHOLM
hehu@csc.kth.se

ABSTRACT

In this paper we present an analysis of a set of examples of how verbal and non-verbal behavior of a service robot influence users' way of positioning themselves during interaction, using concepts from theories of non-verbal behavior. Based on the analysis we propose a design case where a robot utilizes a (naïve) *spatial prompting* strategy to influence the spatial positioning and communicative behavior of the user.

INTRODUCTION

A design requirement of a personal service robot is that it should be configured and provided with work tasks by the user in an interactive and intuitive way. These robots are intended to provide service tasks in the home, possibly offering wide range of services. Typically they are envisioned to be equipped with multimodal spoken dialogue systems, to reduce the complexity in the user interface.

In this paper we argue that theories of spatial positioning need to be considered when developing the communicative system of the robot. Furthermore we present an empirical account of the way spatial behavior of robots influence human users. We also propose the term *spatial prompting*, which refers to active strategies of the robot that are intended to influence users to position themselves in a way that is advantageous for further communicative actions.

Positioning, as it has been approached as a research challenge for human-robot interaction, is considered as providing adaptive *physical* movements of the robot. A result of this is that the *communicative dimension* of positioning typically has been ignored in systems that interactively position themselves in relation to their users. One requirement that is typically put forward is that the robot should position itself in a *socially appropriate manner* [1, 4]. The parameters that concern these approaches are typically derived from research on non-verbal behavior.

In robotics the problem of maintaining the robot localized and situated within a geometric representation of the world has been framed as the Simultaneous Localization and Mapping (SLAM) problem [13]. Recent advances in Human-Robot Interaction (HRI) have raised the interest in detecting and tracking the position of users during interaction. When the position of the user is known, the robot can plan how to position itself [1].

The research on spatial reasoning applied to robotics is well advanced but primarily focused on natural language understanding of spatial relations, providing for exchanges concerning locations of objects in the environment [3].

RESEARCH ON SPATIALITY IN COMMUNICATION

There are several research approaches for human-human that are relevant for spatial management between humans and robots. Hall [14] studied interpersonal distances and distinguished four different distances: *intimate (*0-1.5 ft), personal *(*1.5-4 ft), *social* 4-12 ft, and *public* (> 12 ft). These distances vary both with respect to the current activity and cultural factors. Another dimension that is relevant to spatiality is the concept of *territoriality*, according to Sack, i.e., "the attempt by an individual or group to affect, influence, or control people, phenomena, and relationships, by delimiting and asserting control over a geographic area" [15].

Kendon [12] also studied the spatial configuration of the participants, using the term F-formations, for instance the *L-shape* which describes the relation when two participants have a common visual focus. The shared space, the so called *o-space*, or the *transactional space* is then located in front of the participants, and it is within this area that the interaction is conducted. Clark [5] refers to this space as the *workspace*, where perceptual co-presence is established between speakers [5, 10]. In this context, research on perception and especially visual perception plays an important part for maintaining common ground between participants [10, 8]. Gill [9] has investigated the communicative effects that participants achieve by using nonverbal behavior, focusing on the functional rather than the morphological perspective of nonverbal behavior. One such function is the category *focus* which is a meta-discursive function that signals a shift in the center of attention in the discussion, e.g., a shift in body posture with the same meaning as the utterance "I am going to focus on this spot".

Another, less obvious, but nevertheless important concept is Schegloff's notion of *body torque* [15], a state of the bodily configuration when two different body segments are oriented in different directions. According to Schegloff [15] Body torque "project change", i.e., when some part of the body is organized in an unstable way, the participants may predict that a change in posture is pending. For instance, when turning the head, this might predict a change of the general body orientation. During interaction, speakers monitor the action of others, interpreting purposeful actions that lead towards a common joint goal as compliance [10]. Human-robot interaction is situated in a physical context, where understanding and reference to actions of the human partner during interaction explicitly needs be taken into account. This makes research on *virtual* collaborative [6] environments interesting also in this context, since it is concerned with models that explicitly represent spatiality and reference.

CORPUS ANALYSIS OF SPATIAL MANAGEMENT

We have analyzed a video corpus, collected in a European project [7], containing transcribed data of about 20 user sessions, (approximately 20 minutes each) where a user talks to a robot and teaches it the names and locations of objects using a combination of gestures and speech.

By viewing the video corpus we identified and analyzed instances where the robot movements or verbal actions appear to influence the actions of the user. The examples reflect three different ways in which the robot actively influences the user to act:

- Primary verbal: by using a spoken command
- Primary non-verbal: by movements
- Multi-modal: using movement as trigger for a verbally specified (or grounded action)

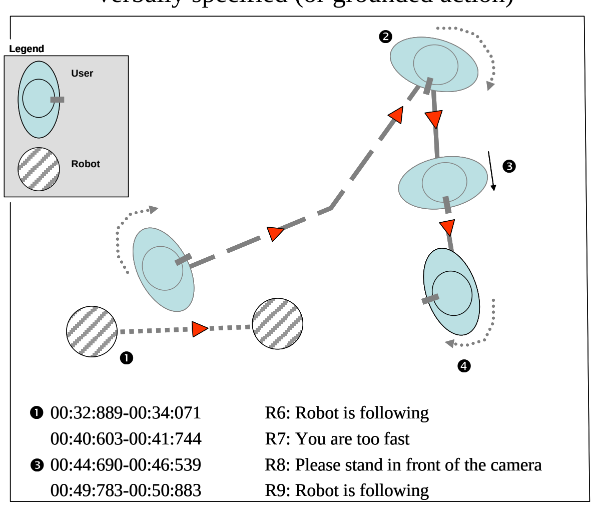


Figure 1: The user and the robot in a follow sequence. The dotted lines show the trajectory of the robot and user.

Primary verbal influence (Example 1)

The example in Figure 1 describes how the events unfold as the user has commanded the robot to follow, after acknowledging this in (❶) the robot starts moving. The robot follows and the user follow the paths depicted in Figure 1. In the second phase of the follow sequence the user has moved to a position that the robot considers too far away. Then the robot says "You are too fast", which triggers the behavior in (❷): the user turns towards what can be characterized as the robot's transactional space (according to Kendon [12]) or workspace (according to Clark [5]). Then the user starts moving slightly towards the robot workspace. When the robot speaks "Please stand in front of the camera", the user quickly moves in front of the robot (❸-❹), something which may be seen as a Give-turn Body Move (according to Gill [9]) that may be seen as a display of the users willingness to interact [11]. In other words, the user displays her attention towards the robot. The visual attention is aimed at the robot once the user has turned around (throughout phases ❷-❹).

The first example (Figure 1) illustrates how verbal action directives influence the physical actions of the user. There are instances of this type of example in each of the 20 sessions that the corpus covers.

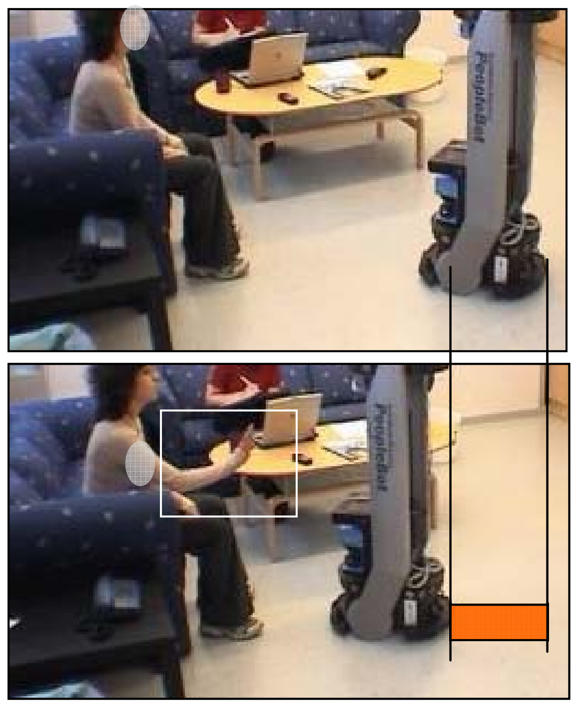


Figure 2: An example of a non-verbal action triggering user action, i.e., a stop gesture (the square in the lower image). The field in the second image, with lines pointing to the first image, illustrates the movement of the robot platform before the gesture is displayed.

Primary non-verbal influence (Example 2)

An example of how the (non-verbal) movements of the robot platform can trigger actions of the user is depicted in Figure 2. In this case the action triggered is a gesture, but it could be another action. In the preceding sequence (not shown) the user has commanded the robot to follow (by saying “Follow me”). Then the user sits down and waits as the robot is approaching as seen in Figure 2 (upper image). During the approach the user raises the arm and displays a “Stop” gesture. It appears as if the robot comes too close; perhaps crosses the border between a *social* to an *intimate* distance, in terms of Hall [14] or triggers a behavioral reaction as the robot breaches a territorial border upheld by the user [15].

On the other hand we might interpret the raising of the hand in a “Stop” gesture as an indication to the robot that this is an advantageous position for the task at hand, i.e., the “Stop” gesture is displayed as part of a joint goal (according to [10]).

U: ok

U: now go to telephone

R: Going to telephone ❶

Figure 3: An example of how communicative actions and spatial configurations of the robot are interrelated in different modalities.

Multi-modal influence (Example 3)

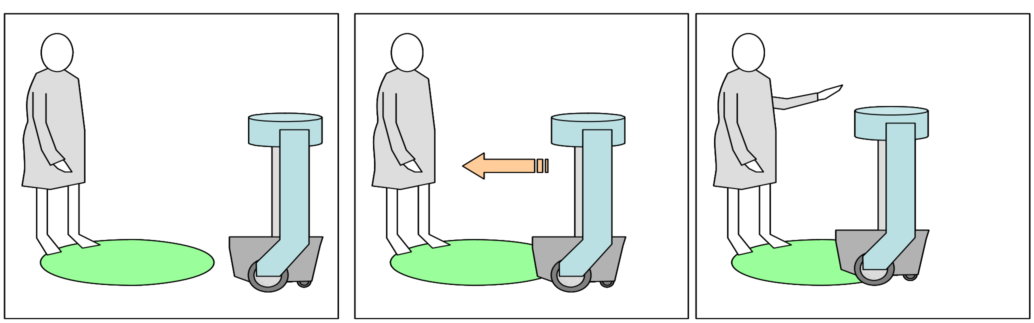
In the moment that passes before the example depicted in Figure 4 the user has acknowledged that the robot has completed the task of finding an object (by establishing a common reference to the object located in front of the robot). Then the user stoops and looks into the camera of the robot, while uttering the command: “now go to the telephone”. Then, when the robot confirms the request by saying “Going to telephone”, the user changes into an upright position ❶ (Figure 3).

Eye contact is maintained during the whole sequence. Our analysis of this is that the user is attempting to require (visual) attention on the part of the robot. We suggest that the moving camera of the robot provides a biomimetic display that makes the user assume a transactional space located in front of the robot. In terms of Body torque, the stooping is a temporary disconfiguration of a posture, and the change back to the more neutral posture in ❶ (Figure 3) is a return to what Schegloff calls a home position [15]. In our understanding, the torque, i.e., the stooping posture and the user’s attempt to require contact can be contributed to the spatial influence of the transactional space [12] and the displayed “eye” gaze of the camera.

A SPATIAL PROMPTING STRATEGY

In the analysis of the scenario we have found examples that suggest that the robot platform may influence the spatial behavior, i.e., posture, gaze direction and gesture displays (e.g., stop gesture). Typically a (multimodal) natural language user interface that is used for human-robot interaction is concerned with the aspects of spatiality that are encoded in language, such as referencing using spatial relations (e.g. “behind”, “beside”, “in front of X” etc) and deictic gestures [e.g. 3].

A system that encompassed a model for spatial influence could provide *spatial prompts* aimed to influence the spatial positioning of the user, for instance, to ensure an optimal configuration for further communicative behavior. An example of such a design is depicted in Figure 4. In terms of a dialogue system design, we can frame the display of Stop gesture in the second example (Figure 2) as a system



U: Follow me
R: Following

R: <Speed increase>

U-gesture: <STOP>

Figure 4: In the first image of the sequence, the robot is commanded to follow the user. As the user stops and the robot approaches the transactional space the robot increases its speed for a short moment. The speed increase is a <u>spatial prompt</u> intended to trigger a response by the user. A safety feature would stop the robot if no action is detected by the user.

goal, i.e., to reach a state in interaction where a Stop gesture has been displayed to provide an end-point in the robot's approach to a position. This requires an internal representation that considers the movement of the robot platform (and the user) together with the task state. In the follow episode in Example 2 the overall task is Follow, but as the user sits down the robot can be said to Approach the user. This could be taken into account in the system, for instance by *increasing* the speed of the robot, slightly in order to prompt the display of a gesture that shows the robot where to stop. It is obvious that a robot behavior like this raises some concerns in terms of acceptability [4], but the point of this example is merely to illustrate that spatial prompting is a possible strategy to actively influence the behavior of the users. We could also imagine an example when the robot prompts the user to tell when the robot is close enough, e.g., by saying: "say stop" while slowly approaching.

Examples like these show how we can turn empirical observations into design proposals. In terms of Body moves [9], the increase in speed in the example in Figure 4, would put Focus on the transactional space and an obligation on the user to react and specify where the robot should be positioned – before the safety feature of the robot stops it at a default distance. Then a prediction could be made, e.g., based on corpus data, so that the robot could provide a contribution that is relevant to the predicted task, e.g., "Show me an object", instead of "Stopped following" (as it is obvious to the user).

CONCLUSIONS

We have discussed a set of examples from our corpus of human-robot interactions arguing that verbal and non-verbal behavior of the robot actively influence users' spatial configuration during interaction. We also provide a scenario where a robot could utilize a spatial prompting strategy. In the future we aim to identify ways of spatially influence users by further analyze corpus data and validate the design proposal by implementing spatial prompting strategies to be tested on a robot platform.

ACKNOWLEDGEMENTS

The work described in this paper was conducted within the EU Integrated Project COGNIRON ('The Cognitive Robot Companion' - www.cogniron.org) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020

REFERENCES

1. Alami, R., Clodic, A., Montreuil, V., Sisbot, E. A., and Chatila, R. (2005). Task planning for human-robot interaction. In Proceedings of the 2005 Joint Conference on Smart Objects and Ambient intelligence: innovative Context-Aware Services: Usages and Technologies Grenoble, France.
2. Pacchierotti, E., Christensen, H., and Jensfelt, P., (2005). Embodied social interaction in hallway settings: a user study. In IEEE Workshop on Robot and Human Interactive Communication Ro-Man2006, Nashville, USA, pp. 164-171.
3. Skubic, M., Perzanowski, D., Blisard, S., Schultz, A., Adams, W., Bugajska, M. and Brock, D. (2004). Spatial language for human-robot dialogs, IEEE Transactions on Systems, Man and Cybernetics, Part C,, vol.34, no.2pp. 154- 167.
4. Walters, M.L., Dautenhahn, K., te Boekhorst, R., Koay K. L., Kaouri C., Woods, S., Nehaniv, C., Lee, D. and Werry, I. (2005) The Influence of Subjects' Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment. Proc. IEEE Ro-man 2005, pp. 347-352
5. Clark, H. H. & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. Journal of Memory and Language , 50(1), 62-81.
6. Gergle, D., Kraut, R. E., and Fussell, S. R. (2004). Action as language in a shared visual space. In *Proceedings of the 2004 ACM Conference on Computer Cooperative Work* (Chicago, Illinois, USA, November 06 - 10, 2004). CSCW '04. ACM Press, New York, NY, 487-496.
7. Green, A., Hüttenrauch, H., Severinson Eklundh, K. (2004). Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration In Proceedings of IEEE Ro-Man, September 20-22, 2004, Kurashiki, Japan.
8. Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience & Biobehavioral Reviews, 24 (6), 581–604.
9. Gill, S.P., Kawamori, M., Katagiri, Y., Shimojima, A. (2000) The Role of Body Moves in Dialogue. In RASK, April 2000, pp. 89-114.
10. Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press.
11. Allwood, J., Nivre, J., and Ahlsén E. (1992). On the Semantics and Pragmatics of Linguistic Feedback, Journal of Semantics 1992 9(1):1-26.
12. Kendon, Adam. (1990). Conducting interaction – Patterns of behavior in focused encounters. Studies in interactional sociolinguistics. Cambridge, NY, USA: Press syndicate of the University of Cambridge.
13. Smith, R., Self, M, and Cheeseman P (1986) Estimating uncertain spatial relationships in robotics. UAI 1986: 435-461
14. Hall, E.T. (1966). The Hidden Dimension: Man's Use of Space in Public and Private. The Bodley Head Ltd, London, UK.
15. Sack, R.D. 1(986). Human Territoriality: Its Theory and History. Cambridge: Cambridge University Press.
16. Schegloff, E.A., (1998). Body Torque. Social Research, 65:3, 535-596.