Active Peripersonal Space for More Intuitive HRI

Patrick Holthaus
Applied Informatics
Faculty of Technology
Bielefeld University, Germany
Email: patrick.holthaus@uni-bielefeld.de

Sven Wachsmuth
Central Lab Facilities
Center of Excellence Cognitive Interaction Technology
Bielefeld University, Germany
Email: sven.wachsmuth@uni-bielefeld.de

Abstract— **In face-to-face interaction, humans coordinate actions in their surroundings with the help of a well structured spatial representation. For example on a dinner table, everybody exactly knows which objects belong to her and where she is allowed to grasp. To have robots, e.g. receptionists, act accordingly, we conducted an on-line survey about the expectations humans have while interacting with such a robot. Results indicate that humans attribute the robot handedness and an awareness of distance and territoriality in its own peripersonal space. In order to align a robot's behavior to these expectations, we have have developed a first spatial representation of the robots peripersonal space.**

I. Introduction

In this paper we have a closer look at the spatial interaction between a robot and a human partner that takes place on a desk. Receptionist robots, for example, share their working space with humans, while explaining routes towards offices or buildings. Also, other (not necessarily humanoid) robots can have a joint *interaction space* with their human partners when fulfilling tasks like cleaning or sorting objects on a table.

Especially in such non-laboratory environments, it is necessary for humanoid robots to be intuitive and to meet an interlocutor's expectations in order to assist people [1]. Therefore, a common goal for researchers in human-robot interaction (HRI) is to create robots that function without the need for special training [2].

One approach to facilitate this interaction is to model principles from human-human interaction, which is most intuitive and best researched, and test whether these principles can also be applied to HRI. We suggest to structure a robot's representation of space in a similar way to humans as the so-called *peripersonal space*. By explicitly using and annotating this space, a robot can adopt behavioral patterns from humans and (re-)act accordingly. A record of interaction histories can help the robot in disambiguation tasks such as which object to grasp next, or what part of a building it has to reference during dialog.

Within this work, we investigate the expectations of humans in these scenarios with the help of a video questionnaire. Participants attributed the robot behavioral patterns such as a preference for the right hand and distance dependent choice which is influenced by the presence of others. As the peripersonal space of humans is believed to have similar properties, we developed a simple representation for a robot
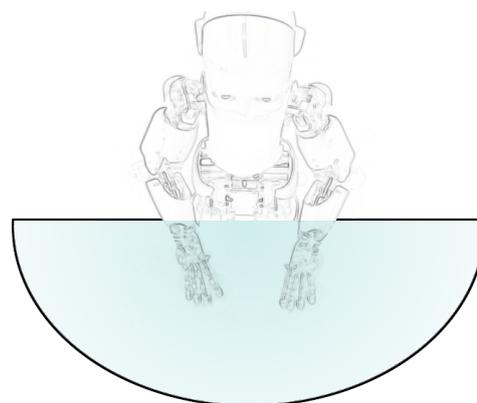

Fig. 1: The peripersonal space of a robot. A black ellipse approximates the outer limits of its reachable area.

which we call the active peripersonal space. We believe that our a system can increase a robot's context awareness and therefore its usefulness in general by aligning to the interlocutors behavior.

II. Related Work

A. Peripersonal Space

As outlined by [3], primates as well as humans have dedicated brain areas for a multi-model representation of the area nearby the body, which is called *peripersonal space*. It is believed that this representation is fundamental for efficient self-monitoring and object manipulation. It also plays a crucial role in interaction with others as these spaces influence each other [4]. If two peripersonal spaces overlap, they form a common area which we call the *interaction space* of these agents. See Fig. 1 for a schematic depiction of an agent's peripersonal space. While such representations may be necessary for human communication, there have also been various attempts to make use of these findings in computer science. For example [5] and [6] use the properties of a virtual and a robotic agent's peripersonal space for learning their respective body structures by exploration and self-touching.

B. Alignment

One way of humans being communicative while still having a relatively low cognitive load is explained via the concept of alignment [7]. The key idea is that interaction partners align

each other to certain syntactic, semantic, and other behavioral structures automatically. This way, high cognitive costs for explicit negotiation of concepts can be avoided. [8] investigate the alignment phenomenon in the context of human-computer interaction and state that effects there might be even more important than in human-human interaction. To also make robots use these concepts, at least a common understanding of the scene is required and thus, human behavior should be modeled and exhibited.

C. Effective Human-Robot Interaction

Being expressive and attentive is crucial for effective HRI [9]. While some theoretical concepts (like peripersonal space and alignment) can be adopted from studies about the interaction between humans, it is important to note that humans have different expectations when interacting with a robotic rather than a human interlocutor [1]. On the one hand, they attribute the robot skills or properties which these don't necessarily have and on the other hand do not use available functionalities because they don't know that they exist. It is therefore especially important to model every behavioral feature and design carefully [10]. Consequently, each artificial feature has to be verified by actual users of the system.

III. SPATIAL INTERACTION STRATEGIES

In order to evaluate what kinds of behavior humans expect during the interaction with a robotic agent on a table, we conducted an on-line survey. Our main hypothesis for the study has been that humans expect human-like behavior of a robot by projecting their own spatial representation to the robot to some degree. To support this theory, three questions have been addressed explicitly:

1) Humans attribute the robot handedness.
2) Humans have distance-dependent expectations on the robot's behavior in peripersonal space.
3) Even the mere presence of a human hand influences the human's expectations on a robot's behavior in terms of territoriality and focus.

A. Study Setup

To investigate these hypotheses, we conducted an on-line study containing images and videos of an interaction situation between the iCub and a human. Firstly, personal data such as age and occupation have been asked.

Then, the actual scene with two equal objects on a table between a human and the robot has been presented from the participants' perspective (Scene I). The introductory text told that during the interaction, human and robot should grasp and inspect objects on the table after each other. Participants were told that currently it is the robot's turn and below the picture they had to choose which object the robot should pick and inspect in their opinion.
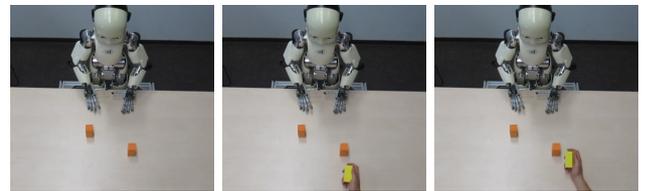
On the next page, they saw a video of the same situation, in which a human hand enters the scene and moves towards one of the objects (Scene II). The hand holds a third object and that is turned around inspected by the human. Again,
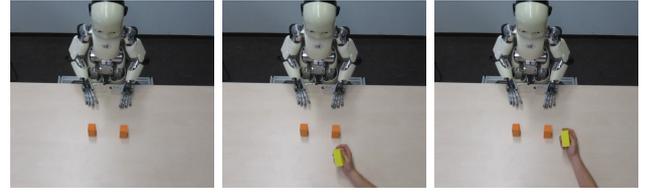
(a) Snapshots from the **Middle** condition.
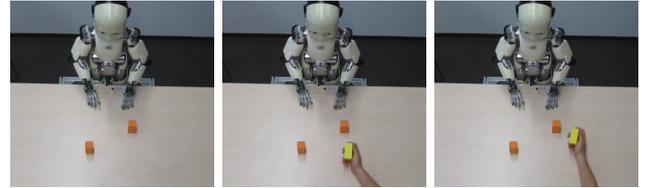
(b) Snapshots from the **Robot** condition.

(c) Snapshots from the **Human** condition.

Fig. 2: Object and hand positions presented in the video study (mirrored scenes left out). From left to right, you can see the initial image without a human hand, then the human hand present, and on the rightmost image the very close human hand condition.

users had to choose which object the robot should pick in its next move.

To avoid effects of handedness, exactly the same situation was presented a third time only flipped horizontally (Scene III). Mirrored scenes have been shown in random order, i.e. in some cases scene one and two, i.e. picture and first video, were mirrored and in others only the second video.

On the later pages, people also answered whether they think that the robot is capable of grasping and recognizing the object, and whether the object's distance or the human's gesture actually influence the robot's choice.

B. Variations in the Condition

In order to clarify which object users want the robot to pick depending on the distance, we varied the positioning of the objects in the following three conditions:

- Both objects lie in the **Middle** of the table with the same distance to the robot.
- The right object lies closer to the **Robot** than the left one.
- The right object is closer to the **Human** than the left one.

Depending on a second condition, the manipulation takes place either in the **close** surroundings between the human and the object, or in **very close** distance almost besides the right object. A third condition constitutes the ending of the video:

Either the hand **leaves** the scene or it **stays** in the same spot after investigating the object.

Please refer to Fig. 2 for an overview of the different placements of the objects and the human's hand. Mirrored conditions and different endings are omitted from the graphics.

C. Participants

A total of 154 people participated in the questionnaire, of which 14 were excluded because they did not accept the privacy statement plus another three who did not answer all questions. 51% are females and the participants' age ranges from 23 to 61 with a mean of 30.5 years. The questionnaire was available in German and English, with each person fluent in at least one of the languages. The average self-assessed German knowledge is 3.89 on a five-point Likert scale (0-4).

The participants' occupation varies greatly although many of them are either students (35%) or scientific staff (23%). Most participants have a common technical understanding with an above average computer experience of 2.99. The average experience with robots is lower (1.46), so most of them are naive to the actual experiment.

IV. HUMANS' EXPECTATIONS

All questions have been evaluated with an χ^2 goodness of fit test [11] against equal probability to distinguish preference for an answer from chance. Also, a χ^2 test of independence [11] has been conducted between the conditions to clarify whether the differences influence the participants' choice.

Almost every participant (96.7%), thinks that the robot is capable of grasping the objects on the table, whereas only 57.6% believe that it can recognize the human's gesture. Because it heavily influences answers to the other questions whether people believe in the functionality of the robot or dislike it (cf. [12], [13]), people with a negative vote on either of these questions have been excluded from the rest of the analysis. In total, 79 participants remain to be evaluated.

A. Side and Object Choice

Fig. 3 shows an overview of which object participants want the robot to grasp. In Fig. 3a, results are grouped by the position of the object summed over all conditions. Participants answers significantly depend on the scene ($\chi^2 = 6.8, df = 2, p < 0.034$). Also, a highly significant preference for the left object can be observed in the first scene ($\chi^2 = 9.2, df = 1, p < 0.003$). Please notice that in this scene (Scene I), there's no hand visible at all, as it is a still image. In the later videos (Scenes II, III), this bias cannot be reproduced anymore, each side is equally chosen. Answers in the third scene significantly differ from those of the fist one ($p < 0.025$).

Part 3b describes the results from another perspective. Answers there are grouped by the place, in which the human hand is active, independent from the object position. The hand could appear either on the right side (default) or on the left side (mirrored setup). Again, in the first scene, no hand is visible. Nevertheless, one of the objects lies at the position where the hand appears in the later videos. No significant
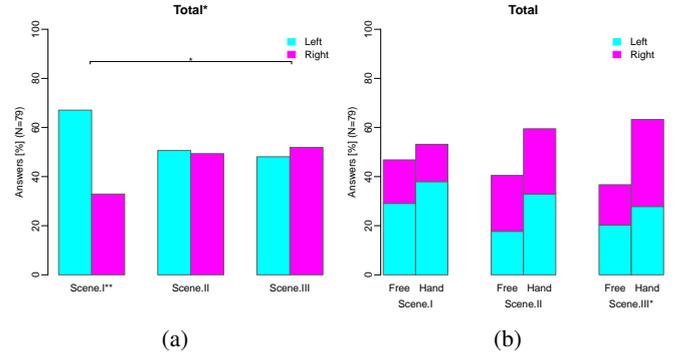

(a)   (b)

Fig. 3: Overview of the participants' suggestion for the robot in the three scenes. In (a), it is depicted whether the right or the left object has been chosen. (b) contains choices grouped by the object, i.e. whether a hand will be presented nearby the object or not. Significance levels ($* := p < .05$, $** := p < .01$) of the goodness of fit against equal probability are given below the columns. Levels resulting from the independence test are given as bars between the columns and at the title for the complete dataset.

differences are observable in the first and second scene. A non-significant tendency in the second, and a significant one in favor of the object with a hand in its surroundings is visible in the third scene ($\chi^2 = 5.6, df = 1, p < 0.018$).

B. Positioning of the Object

The distance between the object and the robot has significant influence on the object choice of the participants, as depicted in the first columns of Fig. 4 (a), (b), (c). For the first scene, a (highly) significant preference for the object that lies closer to the robot can be observed in the **Robot** (Fig. 4a, $\chi^2 = 17.6, df = 1, p < 0.001$) and **Human** (Fig. 4c, $\chi^2 = 6.5, df = 1, p < 0.011$) conditions. Consequently, if both objects are equally far away from the robot (**Middle** condition, Fig. 4b), the answers cannot be distinguished from chance in the still image.

A comparison of the conditions in Fig. 4 (d) shows that the answer for the first scene is highly dependent on the condition ($\chi^2 = 24.1, df = 2, p < 0.001$) and that all conditions differ significantly from each other except the middle condition from all results. The presence of a hand influences the choice in scenes two and three, cf. Fig. 4 (e) and (f). Conditions cannot be distinguished from each other because all (except the Human condition) show a significant favoring of the object closer to the human's hand in both cases.

In the **Human** condition, a significant preference in the first scene disappears in the videos with a hand. People choose an object more often if there is a human activity in the surroundings than in the first observation. The opposite holds for the **Robot** condition where people select the free object more often, at least in the first scene with a hand. In the second video, people again prefer the object close to the robot. In the **Middle** condition, a hand nearby the object produces a significant difference from chance in favor of the
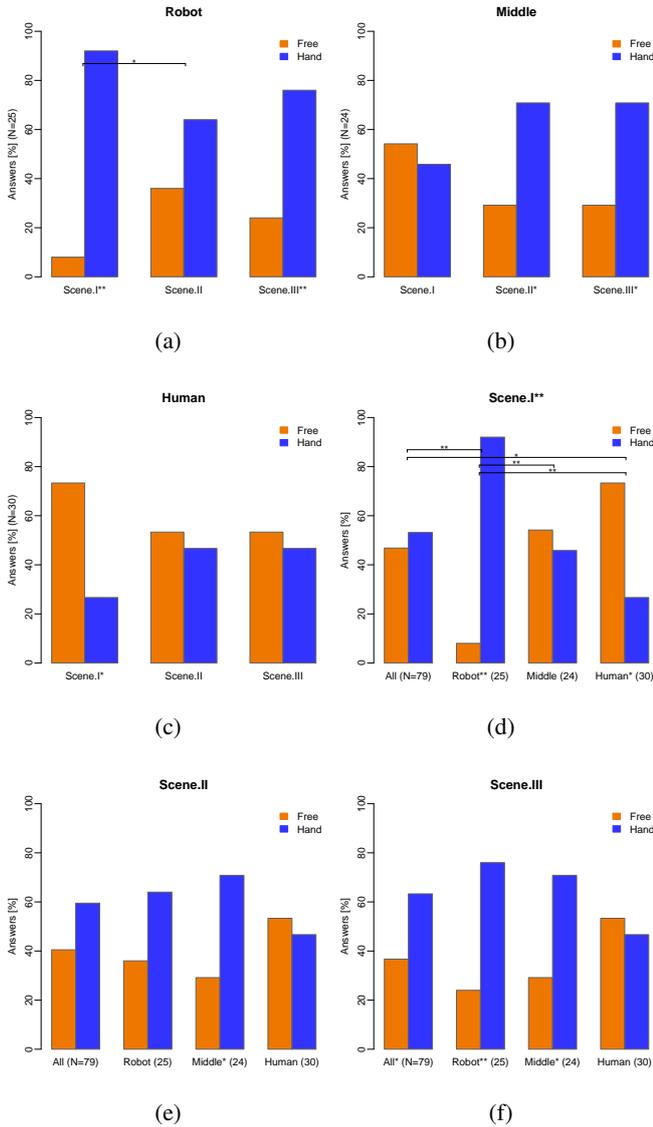
(a)

(b)

Fig. 5: Participants' answers to the influence on the object distance (a) and the human's gesture (b) on the robot's choice grouped by the **Robot**, **Middle**, and **Human** condition. Significance levels ($* := p < .05$, $** := p < .01$) are given along the columns, as bars between the columns, and at the title.

to the object, no significances can be found. On the contrary, if the hand is in the **close** surroundings of but behind the object from the robot's perspective, people choose the object significantly more often in scene three ($\chi^2 = 4.8, df = 1, p < 0.028$). A tendency is observable in scene two.

D. Decision Questions on Influences

In the questions after all scenes, people vote significantly for the **distance** to have an influence on the robot's choice equally among all conditions ($\chi^2 = 17.3, df = 1, p < .001$). One exception from that is the middle condition, where both objects have the same distance from the robot. Answers in this case differ significantly from the other conditions and could not be distinguished from chance, cf. Fig. 5 (a).

The **gesture** instead is not believed to have a strong influence on the robot's choice in all conditions, see Fig. 5 (b). While the overall influence is approved by the sum of all votes ($\chi^2 = 12.2, df = 1, p < 0.001$) and especially by the votes from the middle condition, answers from people in the human and robot condition showed a trend but did not produce a significant result.

V. IMPLICATIONS FOR A ROBOT

The results of the study suggest that a robot's representation of its surroundings should on the one hand model objects in a distance dependent manner during face-to-face interactions on a table. On the other hand, also the presence and history of hands has to be remembered as participants expect human presence to influence the robot's attentional focus. In general, we could show that humans tend to project their own spatial representation onto the robot.

When people only see an image with two equally distant objects on a table, they expected the robot to grasp the object on the left, which is the robot's right side. As almost all participants are right-handed, their intuitive response seems to attribute their own preference for the right hand to the robot. If other factors such as human hands are involved, this first

---

(a)

(b)

(c)

(d)

(e)

(f)

Fig. 4: Participants' choices in the three scenes for each object position grouped by the hand placement. In (a), (b), and (c) answers from the **Robot**, **Middle**, and **Human** condition are plotted. (d), (e), and (f) show a comparison of these conditions in each scene. Significance levels ($* := p < .05$, $** := p < .01$) are given along the columns, as bars between the columns, and at the title.

object to be chosen in scene two and three ($\chi^2 = 4.17, df = 1, p < 0.042$).

C. Positioning and Presence of the Hand

Our results show that participants in scene three prefer an object significantly more if the hand **stays** nearby the object than if it **leaves** the scene ($\chi^2 = 4.6, df = 1, p < 0.033$), although there are also tendencies in the latter case. Trends are visible in scene two as well, but no significances can be observed.

Similar results can be found when comparing the position of the hand in relation to the object. If the hand is **very close**
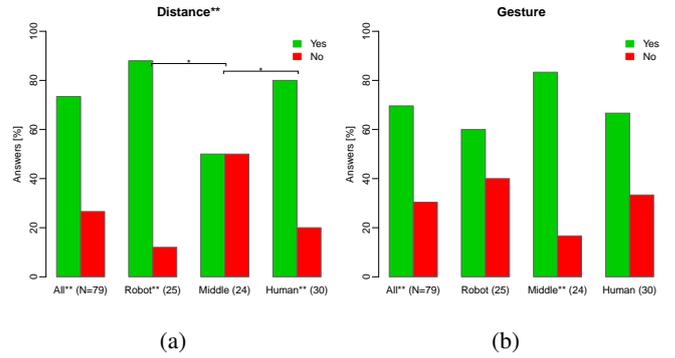
assumption might still be valid, but apparently superseded by these factors.

If one object lies closer to the robot, people initially prefer this one, because it is easier for the robot to reach than the one farther away. This choice again indicates an assignment of a spatial model which is similar to the human's. Such distance dependent behavior can be modeled effectively in a spherical representation of the peripersonal space, because decisions can be based on only one single parameter.

Distance alone cannot be used as a deciding factor for the robot in an interaction scenario. Answers from the video study clearly showed that also the presence of the interaction partner has to be considered. A human hand on the one hand can raise the decision frequency if the object is at the outer limits of the peripersonal space. On the other hand, it can inhibit a decision for an object that is clearly in the robot's personal zone. The results indicate that the human can function as an attention getter in the first case but also occupy areas for personal use in the latter case.

A human hand does not necessarily have to touch, refer or point to the object directly. In the experiment, the hand holds a third object the whole time which makes a direct reference impossible. Instead, the object is marked indirectly through the positioning of a hand. The decision for the object does also not only depend on the distance between hand and object but additionally on the direction from where the hand approaches the object. If the hand happens to be very close besides the object, the effect is weaker. Possibly, this again indicates some form of occupancy.

The ending of the scene also influences peoples choice of the object. Only if the hand stays in the scene, a significant effect is observable. The hand might have been visible for a too short amount of time to be useful as an attention getter. Nonetheless a trend towards the object close to the hand is noticeable in the data.

The above findings are supported by the participants' answers to the decision questions. People know and expect the distance as well as the presence of a hand to have an influence on the robot's choice. To make interaction between naive users and robots smoother, these expectations should be respected in the robot's behavior.

VI. Spatial Representation of a Robot

Based on the survey results, we have developed a preliminary model for a robot's representation of peripersonal space in the context of a receptionist scenario. In such face-to-face interactions which involve a desk, the robot needs to have an understanding of where it is allowed to grasp and point in order to facilitate interaction. Our representation is designed to natively support decisions that are distance dependent and influenced by human presence.

A. Articulated Scene Model

The articulated scene model [14] provides a way to annotate a complex three-dimensional scene with interaction relevant information such as a background and foreground model as well as interaction objects. Although it has its original use
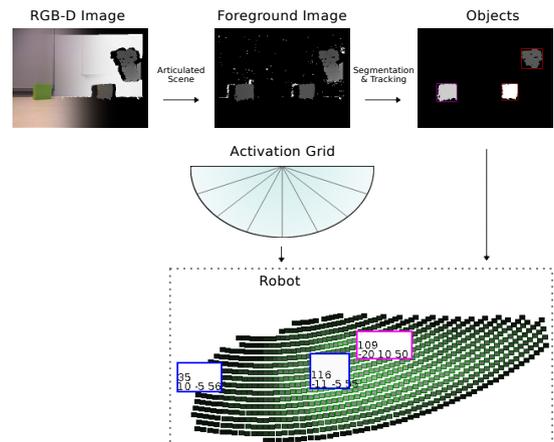

Fig. 6: Depiction of the processing graph resulting in a combined representation of the robot's active peripersonal space. Two objects with the id 35 and 116 are marked in blue, a human hand is drawn in violet (109). Smaller black and green rectangles mark a grid point with the brightness of the color indicating the number of activations.

in the detection and tracking of persons, it can be used as a prior for a distinction between hands and objects [15]. In this work, we use their object segmentation as a preprocessing for our spatial model.

B. Properties

The proposed system is composed of a dual representation of space. On the one side, there is the active peripersonal space. It consists of a spherical coordinate system centered around the robot and limited to the robot's reachable distance. The space is modeled by an activation grid in which each point stores its own history of interactions. Every time a human hand is registered in one of the grid points, its presence is remembered. After for a short time interval, similar to the human's short term memory [16], the activation is forgotten about.

With this model, the robot is able to limit and focus its attention towards the peripersonal space and at the same time represent which parts of the space overlap with the partner's space and form an interaction space. Also, the interaction records allow for regions to be treated specially, for example by supporting the robot in grasping decisions.

On the other side, entities such as objects and people's hands are represented and tracked individually from the grid points. This way, it is ensured to have precise knowledge about the position and extent of objects and especially of people to enable grasping and collision avoidance. See Fig. 6 for a visualization of the processing graph and dual representation of our spatial model.

C. Implementation

While the model itself has been implemented platform-independently, it has been incorporated on the humanoid robot iCub [17] in the context of our receptionist scenario: The child-size robot with 16 degrees of freedom (DoF) in

each arm has been explicitly designed for the study of human-robot interaction [18]. Its arms are controlled via a minimum jerk inverse kinematics framework [19].

A scene is established with the help of the Microsoft Kinect RGB-D sensor[1], accessed through the OpenNI framework[2]. Based on the articulated scene model [15], background, movable objects, and self-moving objects can be identified and tracked.

Human generated movement is stored in the activation grid. The grid itself has an adjustable decay and resolution which in practice can be configured relatively high as the reachable area is very limited (a radius of approx. 37cm in case of the iCub robot). For our current setup, we use a decay of 30sec and a resolution of $\alpha = \beta = 1°; d = 1$cm.

VII. Conclusion and Further Research

Within this work we have shown that humans expect robot's grasping decisions in its peripersonal space to be side and distance dependent and influenced by the presence of an interlocutor. In order to align the robot's behavior with the human (and therefore facilitate interaction), a robot's model should support behaviors that meet these expectations.

We have implemented a preliminary spatial model for humanoid robots that inherits important properties such as handedness, distance-awareness, and a fine-grained sense of where interlocutors act within the interaction space. Effects of occupancy and attention can be considered with the help of our implementation. This becomes important in interaction scenarios where the peripersonal spaces of robot and human overlap.

As our model lays the foundations of further decision making for the robot, we would like to investigate several interaction strategies in future studies within the receptionist scenario. Also, it would be interesting whether the effects that have come up during the video study will be the same or even stronger when people are interacting with a real robot instead of observing a video scene.

VIII. Acknowledgments

References

[1] M. Lohse, *New Frontiers in Human-Robot Interaction*. John Benjamins Publishing Company, 2011, ch. The role of expectations and situations in human-robot interaction, pp. 35–56.

[2] C. Breazeal and B. Scassellati, "How to build robots that make friends and influence people," in *In Intelligent Robot Systems (IROS)*, 1999, pp. 858–863.

[3] N. P. Holmes and C. Spence, "The body schema and the multisensory representation(s) of peripersonal space." *Cognitive Processing*, vol. 5, no. 2, pp. 94–105, 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16467906

[4] D. M. Lloyd, "The space between us: a neurophilosophical framework for the investigation of human interpersonal space." *Neuroscience & Biobehavioral Reviews*, vol. 33, no. 3, pp. 297–304, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18926850

[5] N. Nguyen and I. Wachsmuth, "From body space to interaction space: modeling spatial cooperation for virtual humans," in *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*, ser. AAMAS '11. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 1047–1054. [Online]. Available: http://dl.acm.org/citation.cfm?id=2034396.2034418

[6] M. Hersch, "Adaptive sensorimotor peripersonal space representation and motor learning for a humanoid robot," Ph.D. dissertation, Lausanne, 2009. [Online]. Available: http://library.epfl.ch/theses/?nr=4289

[7] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue." *The Behavioral and brain sciences*, vol. 27, no. 2, pp. 169–90; discussion 190–226, Apr. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15595235

[8] H. Braningan and J. Pearson, "Alignment in human-computer interaction," in *Proceedings of the Workshop on How People Talk to Computers, Robots, and Other Artificial Communication Partners*, K. Fischer, Ed., 2006.

[9] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 4, 2002, pp. 4138 – 4142 vol.4.

[10] F. Hegel, S. Gieselmann, A. Peters, P. Holthaus, and B. Wrede, "Towards a typology of meaningful signals and cues in social robotics," in *ROMAN 2011 The 20th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2011, pp. 72–78. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6005246

[11] K. Pearson, "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine Series 5*, vol. 50, no. 302, pp. 157–175, 1900. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/14786440009463897

[12] L. D. Riek, T.-C. Rabinowitch, P. Bremner, A. G. Pipe, M. Fraser, and P. Robinson, "Cooperative gestures: Effective signaling for humanoid robots," *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 61–68, mar 2010.

[13] L. Takayama and C. Pantofaru, "Influences on Proxemic Behaviors in Human-Robot Interaction," in *Intelligent Robots and Systems (IROS)*, St. Louis, MO, 2009.

[14] A. Swadzba, N. Beuter, S. Wachsmuth, and F. Kummert, "Dynamic 3d scene analysis for acquiring articulated scene models," in *International Conference on Robotics and Automation*, IEEE. Anchorage, AK, USA: IEEE, 2010.

[15] ——, "The articulated scene model: Model-less priors for robot object learning?" Vasteras, Sweden, 2011. [Online]. Available: http://ias.cs.tum.edu/events/rgbd2011

[16] A. Baddeley, S. D. Sala, T. W. Robbins, and A. Baddeley, "Working memory and executive control," *Philosophical Transactions: Biological Sciences*, vol. 351, no. 1346, pp. pp. 1397–1404, 1996. [Online]. Available: http://www.jstor.org/stable/3069185

[17] N. Tsagarakis, G. Metta, G. Sandini, D. Vernon, R. Beira, F. Becchi, L. Righetti, J. Santos-Victor, A. Ijspeert, M. Carrozza, and D. Caldwell, "icub: the design and realization of an open humanoid platform for cognitive and neuroscience research," *Advanced Robotics*, vol. 21, no. 10, pp. 1151–1175, 2007.

[18] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. Von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, A. Bernardino, and L. Montesano, "The iCub humanoid robot: an open-systems platform for research in cognitive development." *Neural Networks*, vol. 23, no. 8-9, pp. 1125–1134, 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20864311

[19] U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots," *Dynamical Systems*, pp. 1668–1674, 2010. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5650851

[1] http://www.xbox.com/en-US/kinect/

[2] http://openni.org/