Can I Help You?
A Spatial Attention System for a Receptionist Robot*

Patrick Holthaus[1], Ingo Lütkebohle[1], Marc Hanheide[2], and Sven Wachsmuth[1]

[1] Applied Informatics
Faculty of Technology
Bielefeld University, Germany
[2] School of Computer Science
University of Birmingham, England

Abstract. Social interaction between humans takes place in the spatial dimension on a daily basis. We occupy space for ourselves and respect the dynamics of spaces that are occupied by others. In human-robot interaction, the focus has been on other topics so far. Therefore, this work applies a spatial model to a humanoid robot and implements an attention system that is connected to it. The resulting behaviors have been verified in an on-line video study. The questionnaire revealed that these behaviors are applicable and result in a robot that has been perceived as more interested in the human and shows its attention and intentions to a higher degree.

1 Introduction

To let robots work and cooperate in domestic or public human environments, it is necessary for humans to interact with them without the need for special training or external instruction [**?**]. At the same time, the acceptance of a robot fundamentally depends on social factors in that people feel comfortable and confident during an interaction [**?**]. Therefore, a general goal in human-robot interaction (HRI) is to understand and mimic communicative cues observed in human-human interaction (HHI). Recent work in social robotics has explored these aspects in distant interactive situations (in terms of proxemics) as well as close-up situations (in terms of joint attention). In this paper we are looking at the intersection or transition between close and distant HRI, in particular, at the distance-based modification of attention behaviors while a person is approaching the robot. As also reported in [**?**], the initiation period is the most critical for a successful human-robot interaction. In most close-up experimental scenarios the human partner is externally briefed about the setup and task, while in most distant experimental setups the robot does not show any reactive or initiative behavior apart from approaching. Such studies typically stop just before the actual communication is established. Within this work, we provide a robot with a

* This work has been supported by the German Research Society (DFG) within the Collaborative Research Center 673, Alignment in Communication.

system that allows it to respond to proxemic features in an interactive situation. Particularly, the robot is able to use the distance to a human as an input that triggers a behavioral output that is based on proxemic cues. The resulting robot's attention is made transparent by the body posture, facing direction, and gaze so that, in turn, the human is aware of the intentions of the robot.

This becomes relevant in receptionist scenarios, for example. To deploy a robot into a hotel lobby or a museum, one should consider which impact a robot's presence could have on the human. E.g., people far away may be less interested in an interaction with the robot than people coming closer towards it. With the presented system, the robot is able to respect the dynamics that humans use by adapting its attention accordingly. An interaction can actively be established by signalling the human interest in an increasing manner as she comes closer towards the robot.

In the following, we conducted a video study to reveal whether the dynamic adaption of attention is accepted by the users and if it lets them understand better how the robot can be used.

1.1 Related Work

Social cues in HRI have been extensively explored in recent years. A first part is dedicated to proxemics as introduced by Hall [**?**], i.e. respecting people's personal spaces. Comparing to HHI, they report similar factors influencing proxemic behavior in HRI [**?**,**?**,**?**]. Kirby et al. [**?**] and Pacchierotti et al. [**?**] study this for person following or passing behaviors. Takayama et al. [**?**] even find for HRI settings that proxemics is influenced by eye contact which suggests a tight coupling of different communicative cues.

While studies on proxemics typically focus on distant human-robot interaction, another line of work looks at maintaining user engagement in close human-robot scenarios [**?**,**?**,**?**]. Here one of the key ideas is to convey intentionality either by appropriate feedback or mixed-initiative strategies that guide the partner through the interaction. An interesting result by Muhl & Nagai [**?**] suggests that – once a mutual interaction between the partners has been established – short distractions of the robot leads to a higher engagement of the human partner.

2 Scenario

Our receptionist scenario consists of a multi-modal interaction system that is implemented on a humanoid robot. It is designed to help users find their ways to offices of colleagues or other university buildings. For the interaction with a human it can use gesture and speech. While the basic interaction with the robot has already been shown in [**?**], we now present nonverbal means for establishing interaction spaces *before* and maintain them *during* the actual interaction at the desk.

Therefore, we have enhanced our robot with an attention system and a method to calculate the distance to a person in the same room.

2.1 The Robot System

The proposed system is implemented on the immobile humanoid robot BARTHOC [?]. Due to huge improvements in the technical construction and design, the original head has been replaced by a newer version called Flobi [?]. It has been excplicitly designed to produce social behaviors and human-like feedback [?] as well as integrating sensor functionality.

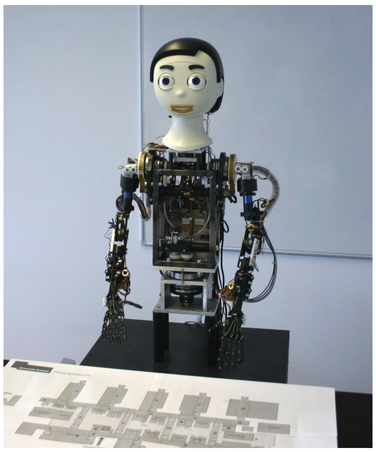

Fig. 1. Picture of the hardware setup. The robot torso BARTHOC with the Flobi head has been placed behind a desk to act as a receptionist.

Of the 45 degrees of freedom (DOF), only the hip, head, and eyes are being used in this scenario (6 DoF). The head is equipped with two fire-wire cameras in the eyes and microphones in the ears. Since the cameras are attached to the eye-balls, their image always reflects the current view direction of the robot. For an image of the hardware setup please see Fig. **??**.

2.2 The Proximity-Based Person Attention System

The person attention system is based on a simple sensor-actor loop that follows the face of a human using of the in-eye cameras of the robotic head. First the distance and deviation of the human face from the camera center is computed. Then the compensation pan-tilt angles are decomposed differently between the hip, head_turn, and eye_turn of the robot depending on the intimate, personal, social, or public distance class.

Face localization is done with a standard face detection algorithm [?] providing a 2D rectangle at image coordinates. Then, the distance is caculated assuming an average size of the detected rectangle on a real face (height $\approx 15cm$). It is estimated considering the horizontal camera resolution and the opening angle of the camera. The distance of a person is defined as the mean of the horizontally and vertically estimated face distance. According to Hall [?], we can now classify whether the person stands either in an intimate, personal, social, or public distance to the robot. In Fig. **??** you can see a human in a close social distance to the robot, ready to enter the personal distance.

Compensation Angles are computed for the horizontal pan and vertical tilt in order to keep the face in the image center which reflects the current gaze direction of the robot. Because the angle compensation (ϕ_{pan}) for the 2D deviation (d_x) in the image is distant specific, this already leads to a stronger engagement of the robot when the person comes nearer. For the intimate distance a factor of $s = 2°$ is used, $s = 1.5°$ for personal, $s = 1°$ for social, and $s = 0.5°$ for public distance (with $\phi_{pan} = -sd_x$). If the compensation angle is below a threshold no movement is performed.

Decomposition of Compensation Angles Into Robot Postures is done specific for the distance class.

These relative turn and pitch angles are transformed to robot postures by the motor control component. The turn is distributed among the hip, head_turn, and eye_turn joints. The head_pitch and eye_pitch joints combine to the overall pitch angle.

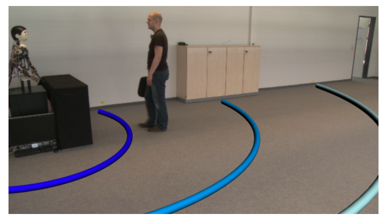

Fig. 2. A person in social distance to the receptionist. The augmented circuits surrounding the robot mark the different distance classes from proxemics theory: dark blue surrounds the personal, lighter blue marks the social, and the outer circle limits the close public distance.

Here, a second method for adapting the attention of the robot to the current interaction situation is applied. Depending on Halls distance classes [**?**], the usage of certain joints is restricted. A so-called inertia value determines to what extend the complete range of a joint is being exhausted. A virtual boundary limits the theoretically possible angle that a joint can be maximally moved.

With a high inertia value the individual joints are limited least, i.e. they can be moved to half of their real maximum. Because of that, most of the movement is accomplished using the eyes only. The head is used for changes in gaze directions that cannot be reached by the eyes alone. The hip remains practically unused. When the inertia is set to medium, the joints are virtually limited to use only 40% of their range. In this setup, the head is used much more frequently for changing the posture. A low inertia value limits the joints to 30%. Therefore, also the hip joint contributes very often to the actual turn value.

The limitation above does not introduce a hard boundary, but a soft one instead. If the angle cannot be distributed the aforementioned way, then the remaining part will be added to joints that have not already reached their real maximum.

Attention Distractors Since humans do not stare consistently at each other during a conversation [**?**], we also suggest the implementation of distracting random gazes. These shift the robots focus from a human to another location for a short time of approximately one second. The robot's attention seemingly gets caught by some other entity in the room.

The resulting view angle is decomposed exactly the same as in the case of a detected face. The only difference is in the usage of joints. The inertia value is even higher than if a human is detected. Thus, the joints are only limited to 70% of their range. This way, one can assure that the robot does not turn its body away from a human in a face-to-face situation.

3 Experimental setup

The proposed attention system has been evaluated with the help of an on-line questionnaire. Participants had to answer questions referring to videos that show an interaction situation. Two main questions have been addressed in this survey:

1. To what extent does the dynamic modification of the attention behavior alter people's perception of the robot?
2. Which influence does the addition of random gazes have on the perception of the robot?

Videos of the Different Conditions We videotaped an interaction between a human and our robot. This way, we could ensure that each participant group rates exactly the same robot behaviors. Furthermore, the experimental results could not be influenced by the various ways people would try to interact with the robot.

The robot has been placed behind a desk in the corner of a room: A human enters this room, walks through it, and eventually stands in front of the desk. When the human arrives and enters the robot's personal distance, it says: "Hello, my name is Flobi. How may I serve you?". The human answers: "Tell me the way to Patrick's office".

The nonverbal behavior of the robot differed between trials and was categorized into eight different conditions:

- **Z** The robot does not move at all (**Z**ero movement).
- **R** The robot's gaze is shifted only **R**andomly.
- **CN** The robot tries to focus its counterpart but acts as if he were permanently in a personal (**C**lose) distance, **N**o random movements added.
- **DN** Again, the human is focused. This time, the movement is **D**istance dependent.
- **FN** The gaze is shifted as if the person were in a public (**F**ar) distance.
- **CR** Same as CN, but **R**andom movements are added in between.
- **DR** Distance dependent as DN, but with **R**andom movements.
- **FR** Like FN, with **R**andom movements added.

The interaction has been recorded from two perspectives. One camera has been following the human all the time and another one shot a close-up of the robot. Both of the videos have been combined to a single one that shows the perspectives side by side. In Fig. **??** you can see three screen shots of the resulting video that has been shown to the participants.

All of the videos have been synchronized to the frame one could spot the robot in the left video for the first time. They fade to black while the human answers the robot to suggest an ongoing interaction between the two agents.

Questionnaire Design The participants had to fill out an on-line questionnaire where they were shown three videos. The first video always showed the Z condition, in the second and third video, the participants could see two videos from different conditions. To prevent side effects of sequence, these videos were shown in random order. Altogether participants can be categorized into the following five groups:

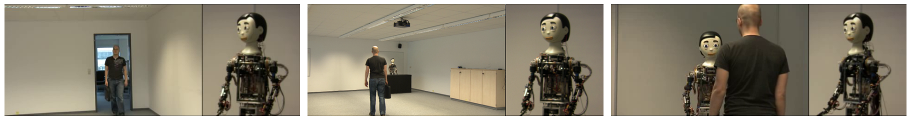

Fig. 3. Video screenshots from the study. The left camera image follows the person as he comes closer to the robot. In the right image a close-up of the robot is shown to let people identify the robot's motions reliably.

NR Differs in containing **R**andom movements or **N**ot.(DN and DR, or FN and FR, or CN and CR)
FD The robot acts as if the human is either **F**ar away or dynamically adjusts its movement to the **D**istance.(FN and DN, or FR and DR)
CD The robot treats the human either as **C**lose to the robot or dynamically adjusts to the **D**istance.(CN and DN, or CR and DR)
CF The robot acts as if the human is either **C**lose or **F**ar away.(CN and FN, or CR and FR)
RR The robot only shows **R**andom movements in both videos.(Control group)

Participants were presented each video. They had the possibility to watch the video as a whole and as many times as they wanted. Beneath the video, the participants were asked to rate certain aspects of the robot's behavior on a five-point Likert scale (0-4):

- The robot's *Interest* in the human
- The *Appropriateness* of the robot's behaviors
- The robot's *Human-Likeness*
- The *Naturalness* of the robot's movements
- How much *Attention* the robot payed to the human.
- The robot's *Autonomy*
- How much of its *Intention* the robot revealed.

Participants Altogether 111 users participated in the study, of which 39.6% were female and 60.4% were male. Their age varied between 16 and 70 years with an average of 30.5. Almost half of them were affiliated with the university, either as students (31.8%) or as scientific staff (18.2%). The vast majority of 88.3% were native German speakers. The rest stated a high understanding of English or the German language. The questionnaire was available in English and German languages, so the questions could be well understood and answered by every participant.

The robot experience varied greatly between subjects. A very large part (84.7%) did not rate their robot experience higher than average on a five-point Likert scale (0-4). The mean value for the participant's robot experience has been at 1.04. In contrast, most of them rated their computer experience either 3 or 4 (67.9%). With an average of 2.94, the computer knowledge seems to be fairly high among the participants. In general, one can say that although the majority of participants are naive to the subject, they have a common technical understanding.

4 Results

Answers to the questionnaire have been evaluated for significant deviations of their mean value. As a method for the comparison, a paired-samples T-Test with a significance level $\alpha = 5\%$ has been used.

4.1 Goal Directed Movements

Almost all of the questions asked produced significant differences between the Z video (zero movement) and every other video that was shown to the participants. Participants rated all of the robots attributes higher for videos that showed a moving robot than for a non moving robot ($\alpha = 5\%, p < .027$).

The RR group with 12 participants is an exception to the others: Videos that showed pure random movements only produced significant changes in the participants ratings for the robot's *Human-Likeness* and *Attention*. Instead, *Interest*, *Appropriateness*, *Naturalness*, *Autonomy*, and *Intention* could not be distinguished from videos without any robot movement. Table **??** shows detailed results of the RR group.

Table 1. Mean ratings ∅ from the RR group with 12 participants, sorted by the video type. The two-tailed significance pR1 of the differences between Z and R1 as well as pR2 between Z and R2 are also depicted if $p < \alpha$.

	∅Z	**∅R1**	**∅R2**	***p*R1**	***p*R2**
Human-L.	1.00	2.00	1.82	.010	.025
Attention	.92	2.25	2.17	.001	.004
Interest	1.25	1.58	1.92	-	-
Intention	.92	1.33	1.42	-	-
Appropr.	1.67	1.75	2.00	-	-
Natural.	1.09	1.50	1.50	-	-
Autonomy	1.25	1.42	1.58	-	-

4.2 Distance Dependent Modification of Behaviors

Only one of the FD, CD, and FC groups showed significant deviations in the ratings of the robot's behaviors. Groups CD (21 users) and FC (24) did not show any differences between the two videos that were presented to them. Responses in the FD condition (26) instead could be distinguished. Participants rated the robot's *Interest*, *Attention*, and its *Intention* higher in the distance dependent video than in the far away condition. The result of this comparison is shown in Table **??**.

Table 2. Mean ratings ∅ from the FD and NR groups, sorted by the type of video. The two-tailed significance p of the differences is depicted in the last column if it is below α.

(a) FD-group (26 participants)

	∅ **F**	∅ **D**	p
Interest	2.58	2.92	.036
Attention	2.58	3.04	.043
Intention	2.12	2.60	.020
Human-L.	1.88	2.12	-
Appropr.	2.50	2.58	-
Natural.	1.50	1.73	-
Autonomy	2.00	2.23	-

(b) NR-group (27 participants)

	∅ **N**	∅ **R**	p
Interest	2.30	3.22	.001
Attention	2.42	3.23	.002
Intention	2.15	2.74	.008
Human-L.	1.85	2.37	.037
Appropr.	2.26	2.33	-
Natural.	1.59	1.78	-
Autonomy	2.19	2.33	-

4.3 The Influence of Random Movements

The participants' answers of the NR group (27) differed significantly in four categories. The robot's *Interest*, *Human-Likeness*, *Attention*, and *Intention* have been rated better in videos with random movements (CR, DR, FR) than in videos without random movements (CN, DN, FN). Other attributes did not show significant differences in the users' ratings. See Table **??** for more detailed results.

5 Interpretation

The above results show that the presented system can serve as an entry point for a human-robot interaction. Each of the presented movement types is more appealing to a human user than no movement at all. Even totally random movements (RR group) suggest a certain human-likeness of the robot. The significance in the ratings of the attention in the random-only case might be caused by the fact that the robot accidentally looked straight into the human's eye as it began to speak. If this had not been the case, the attention ratings of the random behavior would possibly also not be distinguishable from the no-movement case.

Random gazes in conjunction with person-directed gaze can lead to a better user experience than person-directed gaze alone (NR group). Participants believed that the robot had more interest in the human, was more human-like, paid more attention to the human, and expressed its intentions to a greater degree when the robot exhibited random gazes.

At a first glance it might be confusing that especially the attention is rated higher when the robot looks away from time to time. We believe that these distracting looks actually help to communicate an attention to the human because the robot re-focuses on the human every time it had looked away. Therefore, the robot shows that its attention is caught again by the human. While the random gazes help to assign a certain personality to the robot, they do not have an influence on the appropriateness, naturalness, of the behaviors and the autonomy of the robot. The robot apparently does not lose any of its functionality by the addition of distracting gazes.

No differences could be found between the groups that saw the two distance independent behaviors of the robot (FC group). The difference in these conditions obviously did not lead to a higher valuation in one of them. While all cases in this group differed significantly from the zero movement video, participants did not prefer one solution over the other.

Also the distance-dependent condition is not distinguishable from the condition in which the robot acts as if the person stands directly in front of it (CD group). We believe that this could be caused by the similarity of the videos for these cases. Participants could not really tell the difference between the two conditions. That might be a problem of the video itself but could also be a consequence of the experimental setup. Since people were not in the same room with the robot but saw a video instead, their comfortable feeling could not be violated by a robot that doesn't respect personal distances. Therefore, the ratings for the robot are almost identical in the case of direct response as in the dynamic case.

Between the far-away and the distance-dependent condition, significant differences could be found in the user's ratings of the robot's interest, attention and intention. Apparently, the robot was experienced as more responsive and expressive in general, if it uses more of its capabilities and turns its body earlier and more frequently to the interaction partner. Therefore, the distance-dependent behaviors should be preferred over the artificially restricted ones.

6 Conclusion

In this work, we have presented a robot equipped with a spatial model of its surrounding. Also, an attention system has been developed that controls the robot's movements. Both of these components have been combined in an integrated system that allows the robot to exhibit distance dependent social behaviors. We have shown that this system can serve as an entry point for a face-to-face interaction in a receptionist scenario and should be preferred over a non-moving or randomly moving robot.

While random movements alone are not suitable as an entry for the interaction, the overall behavior can benefit from the addition of random directions to the person-directed gaze. Involvement of the robot should be shown in a distance dependent manner. Restricting the robot's hip movement in face-to-face situations leads to a lower overall rating of the robot's responsiveness. The opposite case of immediate response remains a question that should probably be addressed again, since we have not found any significant differences but doubt that an immediate response would be appropriate under real-world conditions.

References

1. C. Breazeal and B. Scassellati. How to build robots that make friends and influence people. In Intelligent Robot Systems (IROS), pages 858–863, Kyonjiu, Korea, 1999.
2. T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. Robotics and Autonomous Systems, 42(3):143–166, March 2003.
3. I. Lütkebohle, J. Peltason, L. Schillingmann, C. Elbrechter, B. Wrede, S. Wachsmuth, and R. Haschke. The curious robot - structuring interactive robot learning. In International Conference on Robotics and Automation, Kobe, Japan, 2009. IEEE.
4. Edward T. Hall. Proxemics. Current Anthropology, 9(2/3):83, 1968.
5. Leila Takayama and Caroline Pantofaru. Inuences on proxemic behaviors in human-robot interaction. In Intelligent Robots and Systems (IROS), St. Louis, MO, 2009.
6. D. S. Syrdal, K. Dautenhahn, M. L. Walters, and K. L Koay. Sharing spaces with robots in a home scenario anthropomorphic attributions and their eect on proxemic expectations and evaluations in a live HRI trial. In Proc. AAAI Fall 2008 Symposium AI in Eldercare: New Solutions to Old Problems, Washington, DC, USA, 2008.
7. van T. Oosterhout and A. Visser. A visual method for robot proxemics measurements. In Proceedings of Metrics for Human-Robot Interaction: A workshop at the Third ACM/IEEE International Conference on Human-Robot Interaction (HRI08), pages 61–68. University of Hertfordshire, 2008.
8. Rachel Kirby, Reid Simmons, and Jodi Forlizzi. Companion: A constraint optimizing method for person-acceptable navigation. In IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pages 607–612, September 2009.
9. E. Pacchierotti, H. I. Christensen, and P. Jensfelt. Evaluation of passing distance for social robots. In IEEE Workshop on Robot and Human Interactive Communication (ROMAN), Hartfordshire, 2006.
10. K. Pitsch, H. Kuzuoka, Y. Suzuki, P. Lu, C. Heath, K. Yamazaki, A. Yamazaki, and Y. Kuno. The rst ve seconds: Contigent step-wise entry as a means to secure sustained engagement in human-robot-interaction. In International Symposium on Robot and Human Interactive Communication, Toyama, Japan, September 2009.
11. C. Muhl and Y. Nagai. Does disturbance discourage people from communicating with a robot? In The 16th IEEE International Symposium on Robot and Human Interactive Communication, Jeju, Korea, 2007.
12. N. Beuter, T. Spexard, I. Lütkebohle, J. Peltason, and F. Kummert. Where is this? - gesture based multimodal interaction with an anthropomorphic robot. In International Conference on Humanoid Robots, Daejeon, Korea, 2008. IEEE-RAS.
13. M. Hackel, M. Schwope, J. Fritsch, B. Wrede, and G. Sagerer. Designing a sociable humanoid robot for interdisciplinary research. Advanced Robotics, 20(11):1219–1235, 2006.
14. I. Lütkebohle, F. Hegel, S. Schulz, M. Hackel, B. Wrede, S. Wachsmuth, and G. Sagerer. The bielefeld anthropomorphic robot head "flobi". In IEEE International Conference on Robotics and Automation, Anchorage, Alaska, 2010. IEEE.
15. F. Hegel. Gestalterisch konstruktiver Entwurf eines sozialen Roboters. PhD thesis, Bielefeld University, 2010.
16. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition (CVPR), volume 1, pages 511–518, 2001.
17. A. Kendon. Some functions of gaze-direction in social interaction. Acta Psychologica, 26:22 – 63, 1967.