# How Can I Help?

## Spatial Attention Strategies for a Receptionist Robot

**Patrick Holthaus · Karola Pitsch · Sven Wachsmuth**

**Abstract** Social interaction between humans takes place in the spatial environment on a daily basis. We occupy space for ourselves and respect the dynamics of spaces that are occupied by others. In human-robot interaction, spatial models are commonly used for structuring relatively far-away interactions or passing-by scenarios. This work instead, focuses on the transition between distant and close communication for an interaction opening. We applied a spatial model to a humanoid robot and implemented an attention system that is connected to it. The resulting behaviors have been verified in an online 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 earlier and to a higher degree than other strategies.

**Keywords** Human-robot interaction · Attention · Interaction opening · Experimental evaluation

## 1 Introduction

The acceptance of a robot as an communicative partner in domestic or public environments fundamentally depends on social factors in that people feel comfortable and confident during an interaction [1]. 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 [2], the initiation period is 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.

To combine these approaches, we provide a robot with a 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.

Of the many scenarios in which such an ability is relevant, in this paper, we have chosen a receptionist setting, as it is prototypical for many real-world cases. 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 signaling the human interest in an increasing manner as she comes closer towards the robot.

Patrick Holthaus · Karola Pitsch · Sven Wachsmuth
Applied Informatics
Faculty of Technology
Bielefeld University, Germany

In the following, we examine two questions: (i) whether the dynamic adaption of attention is accepted by the users, and (ii) if it lets them understand better how the robot can be used. To achieve a broad user-base, we have conducted the expreiment online through the video-study paradigm. In it, a fully autonomous system (cf. Sec. 3) has been parametrized with different strategies, to record a number of videos of the system interaction in different styles. These videos have then been shown to the participants in a random sequence, without any information on their difference, and user's provided information on the point in time when they reached their conclusions, as further explained in Sec. 4. We provide results showing that interpretability was significantly different between the different behaviors (cf. Sec. 5 and 6).

## 2 Related Work

To render human-robot interaction intuitive for the untrained human user, research in HRI has begun to explore which social cues - often derived from HHI - might be beneficial for a robot system [3].

If robot systems are supposed to work in a receptionist setting, a central task consists in entering in contact with a human user and to initiate an interaction with her: The robot needs to signal to the human that it is available, to engage her into an interaction and to achieve a joint focus of attention. To do so and to make the system easily accessible and understandable for untrained users, a set of social cues could be used which are inspired by authentic human interactional practices.

Studies on HHI show that participants negotiate their mutual engagement as a fine-grained stepwise process [4] and - in co-present face-to-face interaction - make use of a set of different multimodal cues (gaze, talk, body orientation). Kendon reveals - analysing the arrival of guests at a garden party - that participants tend to first gaze at each other at a greater distance (here: six to ten meters), then redirect their gaze and will only make eye-contact again at about two meters and then proceed to a greeting exchange ("hello", hand-shaking) [5].

However, in the fields of social robotics and HRI, the moment of this initial contact has received only little attention: Most experimental studies only start when the human is already placed in the appropriate starting position in front of the robot (e.g. [2]) or an operator remote-controls the start of the interaction for an otherwise autonomous system (Shiomi et al. [6]).

If autonomous systems do include a module for 'opening an interaction', these comprise generally of a single step: a greeting, such as "hello", on the verbal level or a hand waving action (e.g. Shiomi et al. [7]). While these systems acknowledge the need to explicitly mark the beginning of an interaction, they do not take into consideration its processural character.

Opposed to these studies, Pitsch et al. consider - for the example of a museum guide robot - the opening of a focused encounter as a dynamic process [8], in which the robot closely monitors the visitor and attempts to adjusts its own conduct accordingly: Once the system detects a person approaching, it turns its head towards her and monitors her head orientation (classified as gazing / non-gazing at the robot). In case of loosing the visitor's face, it pauses and restarts parts of its opening utterance ("may i // offer you some // information about this painting?") to adjust the progress of its talk to the visitor's interest in receiving some explanation. While this "pause & restart"-procedure (Goodwin [9], Kuzuoka et al. [10]) is only a very simple mechanisms using only one observational cue (visitor's gaze), it nevertheless enables the robot system to engage in a stepwise contingent opening with an approaching visitor for about 50 per cent of the cases. Investigation of these cases reveal that - in comparison to a non-contingent opening - visitors stay longer in the interaction with the robot and produce more social behavior towards it (responses, bidding farewell). While this model provides for the visitors' different speeds when approaching the exhibit / robot and is able to draw them closer to the robot, it does not consider the 'spatial dimension' in its own right.

In contrast, the dimensions of 'space' and 'proxemic behavior' have been well explored over the last years in HRI, in particular referring to the works of Hall [11] and Kendon [5] on HHI. Hall has introduced a generalistic concept of proxemic conduct, in which he distinguishes between certain physical distance classes that are used during conversations. These classes include a personal, social, and a public distance, representing the social distance (i.e. the level of comfort, familiarity, etc.) of the interaction partners.

Kendon has shown that groups of people tend to form a set of systematic spatial arrangements (f-formations) in co-present face-to-face interaction, which allow all participants equal and direct access (vis-a-vis, L-shaped, side-by-side). In this, the lower part of the body (in particular, the feet) is dominant in forming the spatial arrangement while the upper body part can be engaged in actions of shorter duration. These observations on human proxemics and spatial conduct are at the basis of a range of models and experiments in HRI, which consider how a robot should control its position with regard to a human:

(i) Making use of Hall's proxemics, some systems are designed to respect peoples' usual spatial habits, e.g. Tasaki et al. suggest a spatial mapping of a system's friendliness [12] and Nakauchi & Simmons developed a robot that stand in line with humans by using a model of personal space [13]. Pacchierotti et al. use Hall's distances to make a robot pass people in hallways in appropriate distances [14], and Kirby et al. have investigated a system's following behavior [15]. Takayama et al. even find for HRI settings that proxemics is influenced by eye contact [16] which suggests a tight coupling of different communicative cues.

(ii) Another set of studies explores the spatial dimension of a robot approaching a human. Dautenhahn et al. reveal that seated persons dislike frontal approach of the robot and instead prefer it to come closer from the right or left side [17]. Koay et al. show also a relationship between the system's approach direction and the robot's appearance [18].

(iii) Some consideration has been given to the formation of spatial arrangements in HRI following Kendon's suggestions. Huettenrauch find that people tend to form f-formations with robots [19], and Yamaoka et al. have proposed a model for proximity control, in which the robot establishes certain spatial configurations with the human and in relation to an object [20].

(iv) Spatial behavior has also been shown to be consequential for engaging humans: Shiomi et al. report that a robot in a science museum can attract visitors and engage them in interaction by moving around [21].

A receptionist robot may not necessarily be able to move in space, but is as well likely to be positioned in one particular place, i.e. at the receptionist's desk. Thus, it will be relevant to realize proxemic conduct in terms of observing humans and displaying certain conduct with regard to their distance.

To do so, it can use a range of other social cues which show its orientation to spatially located events: Pitsch et al. found for a museum guide robot, which is placed in a fix position, that visitors find it more positive and systematic/reactive if the robot - to initiate an interaction - moves its head slowly to random positions as if pro-actively looking for a visitor as opposed to wait in a fixed position [22].

Yamazaki et al. suggest - based on human conduct - that a care robot should display its availability by rotating its head to look at people and, if someone has signaled need for help, the system should display its recipiency of this reaction [23].

Based on Schegloff's study on human "body torque" [24] (i.e. different orientation of parts of the body, in particular lower and upper body parts) Kuzuoka et al. investigate the effect of the robot's body rotation on reconfiguring an f-formation [25]. They explore whether the lower segments of a robot's body have a greater effect on changing the spatial arrangement than the upper body segments or the head. Their preliminary results suggest that both the mere rotation of the robot's head and a "body torque"-configuration has not much effect on the users for repositioning themselves while a change in the orientation of the whole body had a strong effect on people repositioning themselves.

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 [2, 8, 26]. 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 [27] 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.

From this background, the following implications arise for the design of our robot system:

(i) As a constraint, our receptionist system is fixed in its position and therefore cannot approach a human. Nevertheless, it can use its various degrees of freedom to engage in a conversation and express attention.

(ii) In order to act socially based on physical distances, it is required for the robot to detect possible human interaction partners and their position over time.

(iii) As a final requirement, the robot has to dynamically adjust its behaviors to reflect the current interaction situation, i.e. interpret its interaction partner and act accordingly.

## 3 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 communication with a human it can use gesture and speech. While the basic interaction with the robot has already been shown in [28], we now present nonverbal means for establishing interaction spaces *before* and maintaining them *during* the actual interaction at the desk.
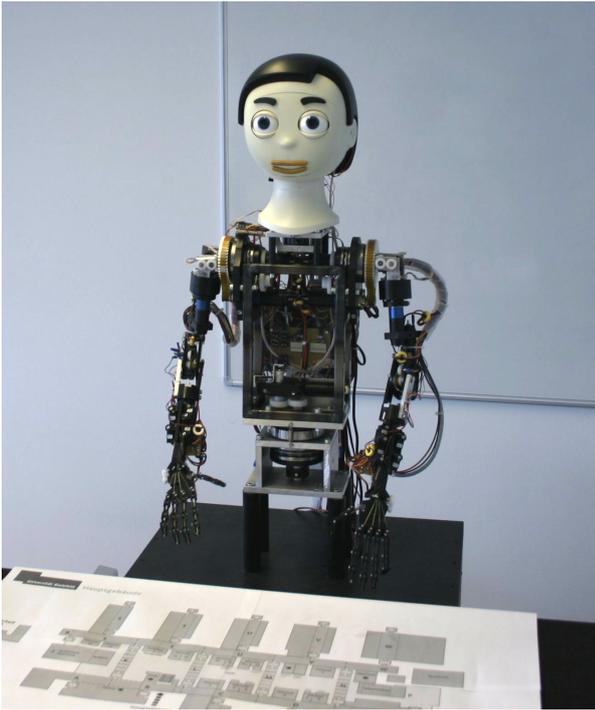
**Fig. 1** Picture of the hardware setup. The robot torso BARTHOC with the Flobi head has been placed behind a desk to act as a receptionist.

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

## 3.1 The Robot System

The proposed system is implemented on the humanoid robot torso BARTHOC [29]. Due to huge improvements in the technical construction and design, the original head has been replaced by a newer version called Flobi [30]. It has been explicitly designed to produce social behaviors and human-like feedback [31] as well as integrating sensor functionality.

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. 1.

## 3.2 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 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.

### 3.2.1 Person Localization

An interaction partner for the robot is detected with a standard face detection algorithm [32] providing a 2D rectangle at image coordinates from the in-eye camera. Assuming an average size of the detected rectangle on a real face ($\approx 15cm \times 15cm$), two estimations for the distance can be calculated by triangulation. One considering the horizontal camera resolution and opening angle, a second one based on the corresponding vertical values. The distance of a person is then defined as the mean of the horizontally and vertically estimated distances.

According to Hall [11], we can now classify whether the person stands either in an intimate ($\leq 45cm$), personal ($\leq 120cm$), social ($\leq 360cm$), or public ($\geq 360cm$) distance to the robot. Persons with their face turned away from the robot are disregarded. Thereby, we can assure that only those are attended to which show signals of attention themselves. In Fig. 2 you can see a human in a close social distance to the robot, ready to enter the personal distance.

### 3.2.2 Compensation Angles

The robot is prompting its attentence by gazing at the person using its cameras. For the horizontal pan and vertical tilt individual compensation angles $\Phi_{pan}$ and $\Phi_{tilt}$ are computed. The robot constantly turns by these angles in order to keep the person's face in the image center which reflects the current gaze direction of the robot. Final angles are determined by the width and height normed vertical ($d_y$) or horizontal ($d_x$) deviation from the image center multiplied with a basic angle $\phi$. For the intimate distance a factor of $\phi = 2°$ is used, $\phi = 1.5°$ for personal, $\phi = 1°$ for social, and $\phi = 0.5°$ for public distance. If the compensation angle is below a threshold $\epsilon$ no movement is performed:

$$\Phi_{pan} = \begin{cases} -\phi & d_x > \epsilon \\ \phi & d_x < -\epsilon \\ 0 & \text{otherwise} \end{cases} \qquad \Phi_{tilt} = \begin{cases} -\phi & d_y > \epsilon \\ \phi & d_y < -\epsilon \\ 0 & \text{otherwise} \end{cases}$$

Because the resulting angle compensation for the 2D deviation in the image is distant specific, this already leads to a stronger engagement of the robot when the person comes nearer.
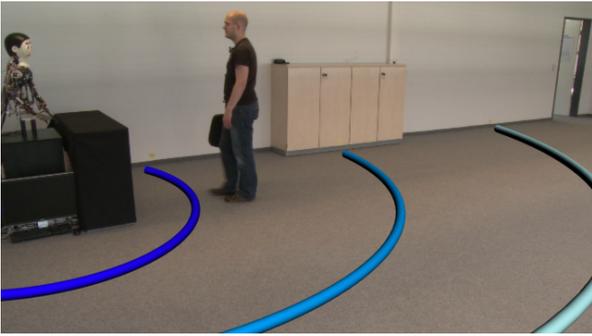
**Fig. 2** A person in social distance to the receptionist. The augmented circuits surrounding the robot mark the different distance classes from proxemics theory: dark blue surrounds the personal, lighter blue marks the social, and the outer circle limits the close public distance.



**Fig. 3** Distribution of basic angle $\phi$ used for compensation and inertia values with regard to distance class. Dark blue marks the values used in personal distance, values in the social distance are highlighted with a lighter blue, and values with the lightest blue are used in public distance.

### 3.2.3 Decomposition of Compensation Angles Into Robot Postures

These relative pan and tilt angles are distributed among the robot's joints specific for the distance class. 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.

Here, a second method for adapting the attention of the robot to the current interaction situation is applied. Depending on Halls distance classes [11], the usage of certain joints is restricted. A so-called inertia value (in the sense of stiffness) 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 50% 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. Please refer to Fig. 3 in which the principle of compensation angles in conjunction with an inertia value is depicted exemplary for a single joint.
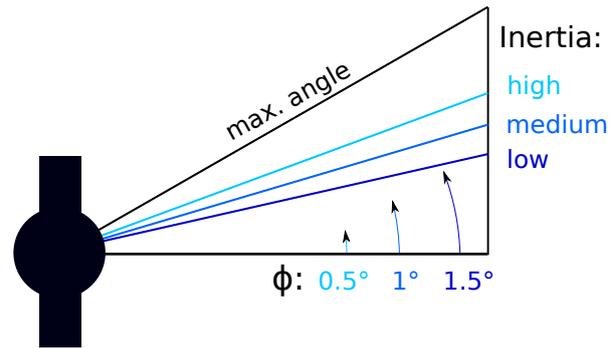
### 3.2.4 Attention Distractors

Since humans do not stare consistently at each other during a conversation [33], we also suggest the implementation of distracting random gazes. These shift the robot's 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 view angle is shifted relatively from the current gaze location and 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.

## 4 Experimental setup

The proposed attention system has been evaluated with the help of an online questionnaire. Participants had to answer questions referring to videos that show a human approaching the receptionist. Further they had to mark the time of the robot's first interaction attempt in the videos. Two main questions have been addressed in this survey:

(i) To what extent does the dynamic modification of the attention behavior alter people's perception of the robot?
(ii) Which influence does the addition of random gazes have on the perception of the robot?

### 4.1 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. Comparability within and between participant groups could only be guaranteed because the interacting person's behavior, especially his path towards the robot, stays the same in all videos.

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".

In Sec. 3.2 we proposed a distance-adapted attention model in conjunction with in-between random gazes. To evaluate this model, we compare the dynamic movements to two static behavior styles. If the robot behaves dynamically, inertia value $i$ as well as compensation angles $\Phi_{pan}$ and $\Phi_{tilt}$ are adapted to the actual distance of a person as in the introduced model. Contrarily, static behavior means that during close and far interaction styles, these parameters are fixed. The robot behaves as if an interlocutor would be located in either a personal (close) or public (far) distance to the robot. Furthermore, we differed among normal movment styles to normal plus additional random movement.

As a consequence, eight videos of the same situation but with different interaction styles have been recorded:

- **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. 4 you can
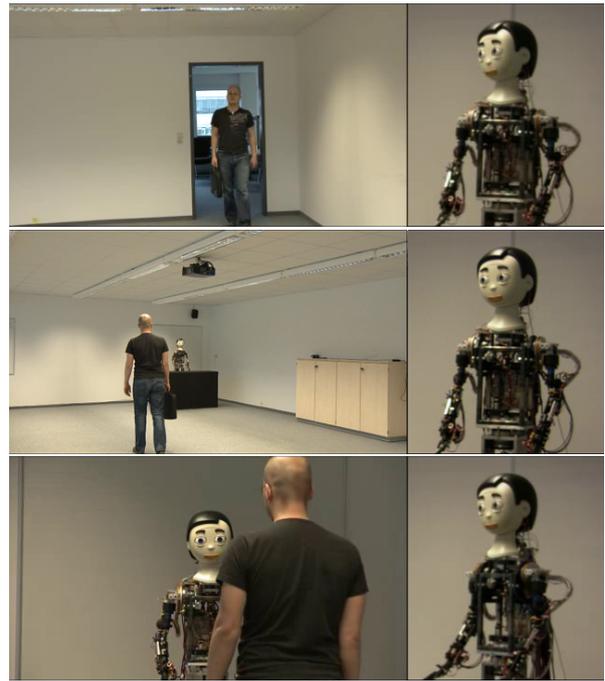


**Fig. 4** Video screen-shots from the study. The left camera image follows the person as he comes closer to the robot. In the right image a close-up of the robot is shown to let people identify the robot's motions reliably.

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.

### 4.2 Questionnaire Design

The participants had to fill out an online questionnaire where they were shown three different 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 have been put in one of the following five experimental conditions:

- **NR** Videos differ 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)

For each of the videos, participants had to determine the *Timestamps* when they thought the robot had realized that the human wanted to interact with it. They had to do so by stopping the video at exactly this time. The video could not be watched any further beyond that point.

After identifying the timestamps in all three conditions, the videos have been presented a second time. Here, participants had the possibility to watch the video as a whole and as many times as they wanted. Beneath the video, they 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 movement'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.

### 4.3 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 highly varied 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.

## 5 Results

Pausing time of the video and answers to the questionnaire have been evaluated for significant deviations of their mean value. As a method for the comparison, a paired-samples Wilcoxon signed-rank test with a significance level $\alpha = 5\%$ has been used.

### 5.1 Goal Directed Movements

Almost all of the questions asked produced significant differences between the Z video (zero movement) and every other video that has been shown. Participants rated all of the robot's attributes higher for videos that showed a moving than for a still robot ($p < .037$). Also, participants thought the robot realizes its human interaction partner faster if it was moving. Times in the stopping task were significantly shorter compared to the no-movement condition ($p < .009$).
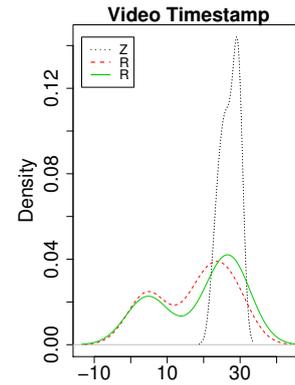
The RR group with 12 participants is an exception to the others: Fig. 5 shows in detail that videos containing pure random movements only produced significant changes in the participants' ratings for the robot's *Human-Likeness* and *Attention*. Instead, *Interest*, *Appropriateness*, *Naturalness*, *Autonomy*, and *Intention* could not be distinguished from videos without any robot movement. Only the first of both random videos has been stopped significantly earlier than the video without movement ($p < .024$). Pausing times of the second random video are higher again and hence no significant differences could be found. In Fig. 6 exemplary the densities of the video timestamps for the RR group are shown. There is an obvious difference between the densities for the zero condition in comparision to both random videos. Additionally, a shift to the right for the second random video is noticable.

### 5.2 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 participants, FR vs. DR or FN vs. DN) instead could be distinguished. The result of this comparison is shown in Fig. 7. The robot's initative has been spotted earlier and participants rated the robot's *Interest*, *Attention*, and its *Intention* higher in the video showing the distance dependent behavior than in the far away condition.
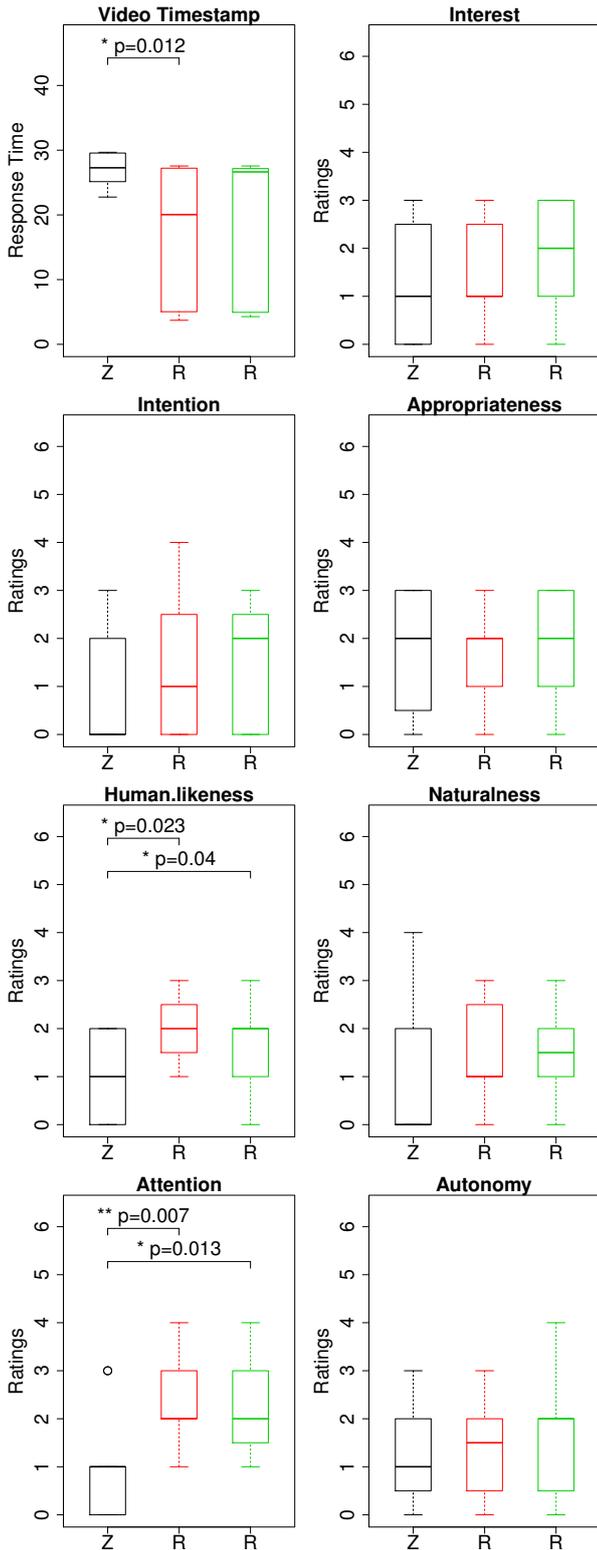
**Fig. 5** Boxplot of video response time (continiously) and ratings (discrete) of the RR group with 12 participants. The zero movement condition (Z) is compared to two different random only movement types (R). Median values are marked with a bold line, the box contains central 50% of given answers. The rounded (3 dig.) two-tailed significance $p$ of the statistical test is depicted if the differences of means are either significant ($*$) or highly significant ($**$).



**Fig. 6** Density of the video timestamps in seconds from the RR group with 12 participants. Densities for the zero movement condition (Z) as well as the first and second random only movement types (R) are shown.

### 5.3 The Influence of Random Movements

The participants' answers of the NR group (27) differed significantly in five categories. Please refer to Fig. 8 for detailed results. 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). Also, the robot's intention to communicate has been perceived earlier if in-between random movements occur. Other attributes did not show significant differences in the users' ratings.

## 6 Discussion

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.

### 6.1 Random Only Movements

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. Another possibility would be that participants attribute the robot some kind of attention because it can shift its gaze to places somewhere in the room. On the one hand,
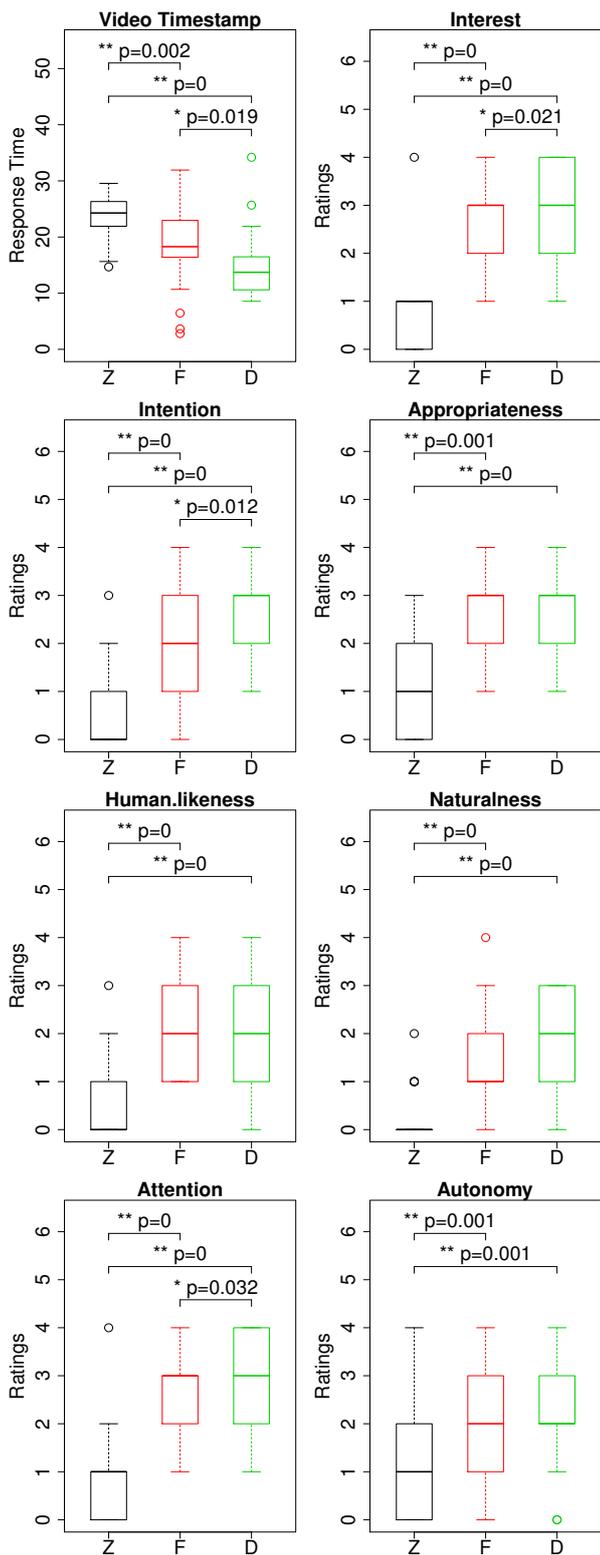
**Fig. 7** Boxplot of video stop time in seconds (cont.) and ratings (disc.) by video type of the FD group (26 participants). Z marks zero movement videos, F consists of videos from the far away condition (FN, FR), and D contains videos with dynamic movement adaption (DN, DR). Median values are marked with a bold line, the box contains central 50% of given answers. The rounded (3 dig.) two-tailed significance $p$ of the statistical test is depicted if the differences in the ratings are either significant ($*$) or highly significant ($**$).
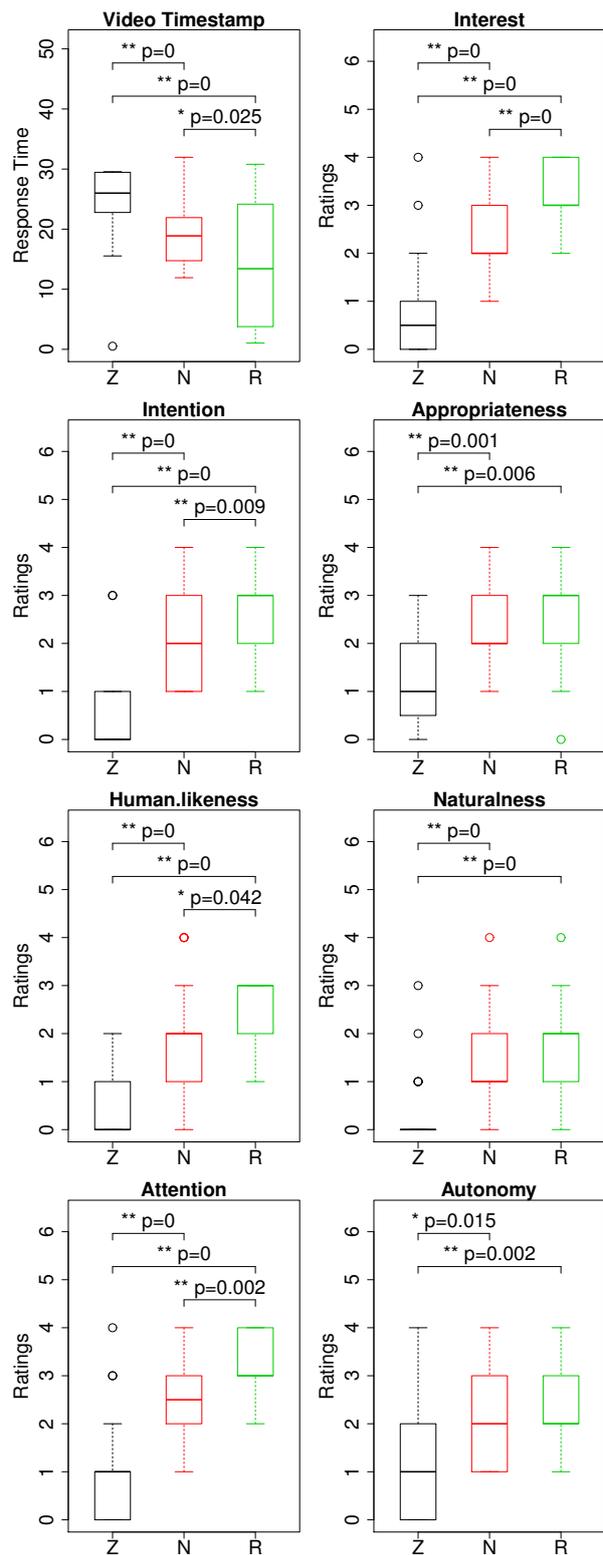
**Fig. 8** Boxplot of video stop time in seconds (cont.) and ratings (disc.) by video type of the NR group (27 participants). Z marks zero movement videos, N consists of videos with straight person-directed gaze (FN, DN, CN), R contains videos with additional random gazes (FR, DR, CR). Median values are marked with a bold line, the box contains 50% of given answers. The rounded (3 dig.) two-tailed significance $p$ of the statistical test is depicted if the differences in the ratings are either significant ($*$) or highly significant ($**$).

participants identified an initiative by the robot earlier in the first random video compared to no movement. On the other hand, in the second random video, the timestamp could instead not be distinguished from the zero condition. Participants apparently mis-interpreted the random movements as a sign of interaction in the first place, but realized the movements were intentionless while watching the second video.

## 6.2 Additional Random Gazes

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. Also, they noticed a robot-triggered interaction earlier in this case.

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 gazes 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. At the same time the robot communicates that it is interactive in an effective way that can easily and almost immediately be detected by a human interaction partner. 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.

## 6.3 Distance Dependent Modification

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 (FD group), significant differences could be found in the user's ratings of the robot's interest, attention, intention, and video timestamp. 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. As these movements are perceived sooner and rated higher, the distance-dependent behaviors should be preferred over the artificially restricted ones.

## 7 Conclusion

In this work, we investigated an attention model for a receptionist robot that reflects the distance between the robot and its interaction partner. The system allows the robot to exhibit distance dependent social behaviors which connect close and distant HRI. We have shown that the proposed dynamic approach can serve as an entry point for a face-to-face interaction in a receptionist scenario and should be preferred over other strategies such as 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 in terms of user experienced robot intention, attention, interest and human-likeness. Involvement of the robot should be shown in a distance dependent manner to increase the perceived intention, attention and interest. 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. T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. Robotics and Autonomous Systems, 42(3):143–166, March 2003.
2. 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.
3. C. Breazeal, A. Takanishi, and T. Kobayashi. Social robots that interact with people. Springer Handbook of Robotics. Springer, 2008. 1349-1369.
4. Schegloff, E. A. (2002). Opening sequencing. In J. E. Katz & M. Aakhus (Eds.), Perpetual contact: Mobile communication, private talk, public performance. (pp. 326-85). Cambridge: Cambridge University Press.
5. Kendon, A. Conducting interaction: Patterns of social behavior in focused encounters. New York: Cambridge University Press 1990.
6. Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. A larger audience, please!: Encouraging people to listen to a guide robot. In HRI '10: Proceeding of the 5th ACM/IEEE international conference on human-robot interaction.
7. Shiomi, M., Sakamoto, D., Kanda, T., Ishi, C. T., Ishiguro, H., and Hagita, N. A semi-autonomous communication robot: A field trial at a train station. In HRI '08: Proceedings of the 3rd ACM/IEEE international conference on human robot interaction.
8. K. Pitsch, H. Kuzuoka, Y. Suzuki, P. Luff, C. Heath, K. Yamazaki, A. Yamazaki, and Y. Kuno. "The first five 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.
9. Goodwin, Charles Conversational organization: Interaction between speakers and hearers. Academic Press (New York), 1981.
10. Kuzuoka, H., Pitsch, K., Suzuki, Y., Kawaguchi, I., Yamazaki, K., Kuno, Y., Yamazaki, A., Luff, P. and Ch. Heath (2008). Effects of restarts and pauses on achieving a state of mutual gaze between a human and a robot. In CSCW 2008.
11. Edward T. Hall. Proxemics. Current Anthropology, 9(2/3):83, 1968.
12. T. Tasaki, K. Komatani, T. Ogata, and H. Okuno. Spatially Mapping of Friendliness for Human-Robot Interaction. Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 52–526, Edmonton, August 2005.
13. Y. Nakauchi, and R Simmons. A social robot that stands in line. Proc. of the IEEE/RSJ Intern. Conference on Intelligent Robots and Systems, pages 357–364, 2000.
14. E. Pacchierotti, H. I. Christensen, and P. Jensfelt. Evaluation of passing distance for social robots. In IEEE Workshop on Robot and Human Interactive Communication (RO-MAN), Hartfordshire, 2006.
15. 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.
16. Leila Takayama and Caroline Pantofaru. Influences on proxemic behaviors in human-robot interaction. In Intelligent Robots and Systems (IROS), St. Louis, MO, 2009.
17. Dautenhahn, K., Walters, M., Woods, S., Koay, K. L., Nehaniv, C. L., Sisbot, A., Simeon, T. (2006). How may I serve you?: A robot companion approaching a seated person in a helping context. In Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction.
18. K. L. Koay, D. S. Syrdal, M. L. Walters and K. Dautenhahn. Living with Robots: Investigating the Habituation Effect in Participants: Preferences During a Longitudinal Human-Robot Interaction Study. 16th IEEE International Confer-ence on Robot & Human Interactive Communication, 2007.
19. H. Huettenrauch, K. S. Eklundh, A. Green, and E. A. Topp. Investigating Spatial Relationships in Human-Robot Inter-action. Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 5052–5059, 2006.
20. Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. How close? A model of proximity control for information-presenting robots. In HRI 2008.
21. Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. (2006). Interactive humanoid robots for a science museum. In HRI'06. Salt Lake City, Utah, USA.
22. Pitsch, K., Wrede, S., Seele, J. -C. h., and Süssenbach, L. Attitude of german museum visitors towards an interactive art guide robot. In HRI2011.
23. Yamazaki, K., Kawashima, M., Kuno, Y., Akiya, N., Burdelski, M., Yamazaki, A., and Kuzuoka, H. Prior-To-Request and request behaviors within elderly day care: Implications for developing service robots for use in multiparty settings. In ECSCW 2007.
24. Schegloff, E. A. (1998). Body torque. Social Research, 65(3), 535-596.
25. Kuzuoka, H., Suzuki, Y., Yamashita, J., and Yamazaki, K. Reconfiguring spatial formation arrangement by robot body orientation. In HRI '10: Proceeding of the 5th ACM/IEEE international conference on human-robot interaction.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. F. Hegel. Gestalterisch konstruktiver Entwurf eines sozialen Roboters. PhD thesis, Bielefeld University, 2010.
32. 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.
33. A. Kendon. Some functions of gaze-direction in social interaction. Acta Psychologica, 26:22 – 63, 1967.

**Patrick Holthaus** is a PhD student and member of the Applied Informatics group and CRC 673 "Alignment in Communication" at Bielefeld University, Germany. In 2009 he received his M.Sc. in Intelligent Systems also at Bielefeld University with the top result of the year. His current research interest is focussed to Human Robot Interaction and particularly social communication signs in the spatial dimension.

**Karola Pitsch** is a research fellow in the Applied Informatics Group and the Research Institute for Cognition and Robotics (CoR-Lab) at Bielefeld University, where she currently works on the EU project "iTalk" and co-heads the project "Alignment in Augmented Reality based Cooperation" in the CRC 673 "Alignment in Communication". In 2006, she received her PhD in Linguistics from Bielefeld University. From 2005 to 2008, she has been a research fellow in the Work, Interaction and Technology Research Group at King's College London working on the EU-project "PaperWorks" before joining the Applied Informatics Group at Bielefeld University in 2008. She has undertaken extended research stays at EHESS (France), UCLA (USA), Universidad de Buenos Aires (Argentina) and Saitama University (Japan). Her research focuses on multimodal human interaction in authentic and technically mediated settings, the integration of qualitative and quantitative research methods and the design and evaluation of human-robot-interaction.

**Sven Wachsmuth** holds a faculty staff position in the Applied Informatics and, since 2008, is heading the Central Lab Facilities of the Center of Excellence Cognitive Interaction Technology (CITEC). He received his Diploma and PhD in Computer Science from Bielefeld University in 1997 and 2001, respectively. In 2002, he spent a sabbatical year supported by the DFG at the AI group, University of Toronto. He was technical coordinator of the FP5 project VAMPIRE and is PI in the DFG CRC673 "Alignment in Communication" and CoR-Lab. He is currently working in the fields of visual scene analysis and cognitive robotics.