

A Spatial Model of Engagement for a Social Robot

Marek P. Michalowski
Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
Email: michalowski@cmu.edu

Selma Sabanovic
Science and Technology Studies
Rensselaer Polytechnic Institute
Troy, New York 12180
Email: sabans@rpi.edu

Reid Simmons
Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
Email: reids@cs.cmu.edu

Abstract—Even if a socially interactive robot has perfect information about the location, pose, and movement of humans in the environment, it is unclear how this information should be used to enable the initiation, maintenance, and termination of social interactions. We review models that have been developed to describe social engagement based on spatial relationships and describe a system developed for use on a robotic receptionist. The system uses spatial information from a laser tracker and head pose information from a camera to classify people in a categorical model of engagement. The robot's behaviors are determined by the presence of people in these different levels. We evaluate the system using observational behavioral analysis of recorded interactions between the robot and humans. This analysis suggests improvements to the current system: namely, to put a stronger emphasis on movement in the estimation of social engagement and to vary the timing of interactive behaviors.

I. INTRODUCTION

The ability to detect, track, and recognize the location, identity, and behavior of human beings is an important component of human-robot social interaction. The use of cameras, lasers, microphones, and other simple sensors in human-oriented perception has improved to the point that we should expect social robots to have a reasonable understanding of where people in the environment are and, to some extent, what they are doing. However, even if a robot could obtain perfect perceptual information about the location (distance and direction), pose, and movement of human interactors, it is still unclear how this information should be used by the robot to enable the natural initiation, maintenance, and termination of social interactions with humans.

One of the primary functions of nonverbal behavior is the regulation of face-to-face social interaction, for which space and distance are of fundamental importance [1]. Rauterberg et al. [2] introduce the concept of “shared social space,” characterized by visibility (expression, gesture), audibility (voice, intonation), and social nearness (physical distance). Our attributions of mental states to interaction partners are strongly influenced by the spatial and temporal qualities of physical behaviors in the shared space [3]. Having a model of people's attention, intention, and affect according to their physical stance relative to us is important for selecting behaviors and thereby steering interactions in an appropriate direction.

A number of models have been proposed to characterize the meaning of the spatial relationships between individuals in social interactions. We review these frameworks and

consider their relevance and application to the development of social robotics. We then describe a system that was developed for and implemented on a robotic receptionist. We conducted an observational analysis of this system from recorded interactions. The analysis revealed limitations of our engagement model as well as of the robot's behaviors toward people at different levels of engagement. We evaluate the system and discuss suggested modifications to our model.

II. MODELS OF SOCIAL SPACE

E.T. Hall [4], [5] developed a conceptual framework known as “proxemics” that is concerned with human perception and use of space. He proposed a basic classification of distances between individuals:

- Intimate distance (0-18in): unmistakable involvement with another body (lover or close friend).
- Personal distance (18in-4ft): comfortable separation, interaction with friends.
- Social distance (4-10ft): reduced involvement, interaction with non-friends.
- Public distance (>10ft): outside circle of meaningful involvement, public speaking.

Increasing distance naturally results in degraded thermal, olfactory, visual, and aural sensations between interactors. Voice volume increases with distance between individuals, while intimacy of conversational content decreases to a public nature. Hall notes, however, that these distances were deduced from observation of American and European subjects. Specific distance between interactors actually varies by culture, gender, status, age, familiarity, relationship, pose, etc. [6]. Walters et al. [7], [8] have found that these distances are generally applicable to human-robot interaction as well.

Schefflen [9] proposes micro-territories called “spots,” “cubits,” “k-spaces,” “locations,” and “modules” to characterize units of space that generally determine common distances in face-to-face interaction, dimensions of furniture, seating configurations, and room layouts. The distances involved here are highly dependent on the dimensions of the human body, suggesting that the size of a socially interactive robot should be taken into account when anticipating the spatial factors that will affect its interactions.

Much of our sense of engagement with interactors also relates to our perception of their attention. Langton et al. [10] review various psychological and neurological models for how humans perceive attentional cues (body, head, eyes, gestures, verbalizations) and how these cues might mediate



Fig. 1. The Roboceptionist in its booth.

each other in the brain. The distance between interactors determines the relative salience of these visual cues, shaping our perception of attention and therefore of engagement.

Other researchers in social robotics have developed generative models of engagement for the purposes of guiding the behavior of embodied interactive systems. The robot Kismet [11] implicitly used an interactor's spatial configuration in a set of reactive behaviors. These behaviors included seeking, avoiding, calling, and greeting people based on the distance, speed, and sound of interactors. Sidner et al. [12] have developed a social robot that uses head pose, gaze, and deictic gestures to imitate and take turns with a conversational partner. Their understanding of engagement does not necessarily take into account relative spatial positioning and movement.

Literature on the role of physical space in the structuring of social interactions has come from a number of fields, such as psychology, cognitive science, human-computer interaction, and robotics. Some models come from the study of a single factor such as distance in observed interactions, while others are crafted for the purpose of designing an interactive system using available technologies. There is a need to develop a more comprehensive model of social engagement that

- accounts for many different cues (distance, head pose, gaze, facial expression, etc.),
- has a basis in psychological or cognitive theories of the perception of social attention,
- holds explanatory power in describing real interactions between people, and
- suggests how to generate naturalistic behaviors for socially interactive systems.

We have designed a model of engagement for a robotic receptionist. The goal of the robot is to engage visitors so that they begin interacting with the robot and, once an interaction has begun, to maintain their interest. We consider the role of a receptionist to be suitable for studying principles of initiation, maintenance, and termination of interactions that are applicable to more general social situations.

III. THE ROBOCEPTIONIST'S MODEL OF SOCIAL ENGAGEMENT

The Roboceptionist [13] (Fig. 1) is a robotic receptionist situated in a booth near the entrance of the Robotics Institute at Carnegie Mellon University. It consists of an animated face on a flat-panel monitor that turns on a pan-tilt unit, which is mounted on a mobile base. The robot uses a keyboard interface for input. It can answer questions about the weather and office locations, display building maps, and hold conversations about its own fictional changing story line.

A number of capabilities depend on good sensing and modeling of people in the environment, such as polite greeting, appropriate verbalizing, attentive listening, determining how to be of service, and evaluating effectiveness. Three design criteria, then, motivate our tracking and classification of people: (1) the robot should correctly attend to people (by turning its monitor and face in the direction of a person's face) as they approach and stand around it, (2) the robot should recognize when people arrive and depart so as to maintain context for ongoing interactions, and (3) the robot should verbalize (e.g. greet visitors) in an appropriate manner. For example, if the robot's goals are only to maximize the number and length of interactions, it might be reasonable for it to address every passerby it perceives. However, the role of a receptionist entails adhering to particular social norms. Therefore, it is our approach that the robot should selectively initiate interactions only with people who appear interested in interacting.

The first of these design criteria calls for location information for people in the environment, the second calls for tracking people's continuity of identity, and the third calls for inference about a more abstract property of the participants – that is, their level of engagement or their direction of attention. For our purposes, we have designed a categorical model of engagement:

- *Present*: people who are standing far from the robot.
- *Attending*: people who are idling closer to the robot, perhaps observing it and likely aware of its presence.
- *Engaged*: people who are next to the booth and passively observing the robot's behavior.
- *Interacting*: people who are actively participating in an exchange with the robot.

While these categories and their associated distances are loosely analogous to Hall's social distances [4], it should be noted that this model is neither motivated by, nor reflective of, the psychology of human attention. Rather, it is a generative model designed for the purposes of specifying different behaviors of the robot toward people at different levels of engagement. For example, the robot is to acknowledge the presence of Attending people by turning to their general direction. It should turn to and verbally greet Engaged people, followed by occasional nods in their direction. It should verbally prompt Interacting people for input if they are not typing. See [13] for a full description of the robot's behaviors.

A. Sensor fusion

The robot uses a video camera and a laser range scanner to detect people. The robot is situated behind a desk in a

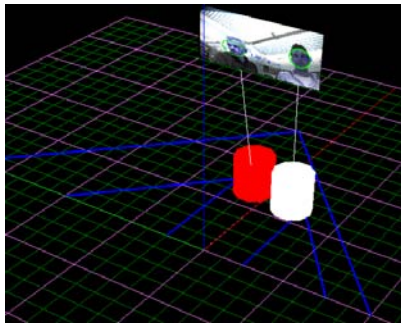


Fig. 2. Visualization of sensor fusion between laser and camera in tracking people.

booth, the laser is mounted in the desk behind a slit at human knee-level, and the camera is installed on the front and top of the robot near human chest-height. Adjacent laser range values that deviate from a learned background model are clustered and tracked using a Kalman filter (see [14]). The vision software returns a list of tracked faces with coarse pose information (whether the face is frontal or not). Frontal faces are detected using the method in [15] and are used to initialize (and re-initialize) a mean shift color tracker described in [16].

The information from these two sensors is fused on two levels [17]. First, the system needs to build a model of people's spatial location and pose (where they are standing on the floor plane, where their heads are located in space, and their direction of head orientation) so that the robot can visibly and appropriately direct its attention to people around its desk. In order for output from the sensors to be correlated, the location features of detected objects must be transformed into a common reference frame. 2D polar coordinates from the laser and 2D image-space coordinates from the camera are transformed into 3D Cartesian coordinates, translated, rotated to compensate for the sensors' positions and orientations relative to the robot, and matched based on nearness (Fig. 2). The robot is therefore able to attend appropriately to a person's direction (from the laser scanner) and to the height of the person's face (from the camera). Tracking the identity of interactors is also improved by allowing one sensor to correct for discontinuities from failed tracking in the other. Discontinuities result in a loss of context in an interaction, as the robot would think that the current interactor has left and that a new one has arrived.

Next, these multiple sources of information must be integrated to estimate people's level of engagement. Each of the two sensory modalities is able to contribute an independent estimate: for example, the classification can be done by laser scanner alone, based on physical location, or by camera alone, based on head pose estimation. This symbolic fusion is done in a hierarchical manner inspired by work in the cognitive study of social attention. Perrett et al. (see [10]) suggest that if only a body is visible, its position and pose are used. If the head is clearly visible, its direction is used to adjust the estimated attentional focus. If the eyes are clearly visible, gaze direction is incorporated into the estimate.

In our system the robot uses spatial location as an initial estimate of attention (Fig. 3). If a frontal face has been associated with that location, the classification is adjusted

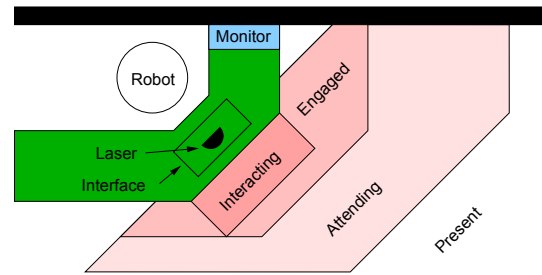


Fig. 3. Spatial regions used as initial estimate of engagement.

to a higher level of engagement. If the face is no longer frontal, or has been lost by the tracker, the classification returns to a lower level of engagement. We are currently developing a head pose tracker that can track the 3D rotation of interactors' heads. This directionality will be incorporated into the estimate of social engagement as well as used to guide the verbal content of interactions (if the person is looking at an object in the booth) and the social cues exhibited by the robot (following the interactor's gaze).

B. System performance

The robot operated for almost two years using the laser scanner alone. The addition of the camera and sensor fusion system improved the robot's attentiveness to nearby people, as it is able to detect, focus on, and track the location of people's faces. It also reduced the occurrence of lost context in interactions: in a week of operation, the sensor fusion system achieved a 34% reduction in the number of interactions that would have been confusingly cut short by single-modality tracking failures.

The addition of vision also improved the classification of social engagement from use of the laser alone. The act of typing at the keyboard interface may be considered "ground truth" that a person is interacting. If fused sensor data can more accurately predict when someone is about to interact with the robot, then verbal greetings will be more appropriate and less annoying (for people who approach the booth but do not intend to interact). In two weeks of operation, the sensor fusion system decreased by 69% the number of inappropriately greeted people from the use of the laser scanner alone. However, the system increased the number of people who were correctly predicted to begin typing to the robot from 19% to only 25%.

There is clearly much room for improvement. Unnatural or uncomfortable situations can arise from perceptual failures or misclassifications of interactors' engagement: the robot may address someone who is not open to interaction, or it may ignore someone else who is. We continue to improve the use of our sensors, but it is difficult to automatically identify failures and determine how best to address them. More immediately, it is necessary to evaluate the validity of our engagement model in order to maximize the utility of already available sensor data. To this end, we have performed an analysis of recorded interactions with the robot to evaluate the current model and to identify possible improvements. These changes are motivated by the general characteristics of human social behavior while at the same time taking into account the constraints imposed by currently available technology. It is still difficult to obtain reliable information on



Fig. 4. Frame from coded video.

humans using machine vision, especially at longer distances. The laser scanner, however, is quite accurate even at long distances. We have found that the high degree of movement through the space surrounding the Roboceptionist requires earlier anticipation of human behaviors, in order that the robot may begin to behave appropriately before it is too late. Therefore, a model of engagement based on stationary positions may be insufficient for increasing the frequency and quality of interactions.

IV. OBSERVATIONAL ANALYSIS

Two video cameras in the Roboceptionist's booth (Fig. 4) are positioned to view all people in the vicinity of the booth (including the space in Fig. 3). Over a period of two days, 12.51 hours of operation were recorded. The video was coded using Noldus Observer software [18], aligned and merged with logged data relating the robot's internal state and behaviors, and analyzed statistically and temporally.

A manual coding schema was created to capture the interactive behaviors (movement, position, gesture, pose, gaze, etc.) of humans and the robot, as well as the robot's internal state. The schema can be summarized as follows:

- Interaction
 - Look to/away from ... (person only)
 - Speak to ...
 - Type (person only)
- Movement
 - Pass (person only)
 - Approach/walk away from ... (person only)
 - Stand in front of/next to/behind ... (person only)
 - Turn to/away from ...
- Internal state (robot only)
 - Idle/chat/greet/interact state
 - Person enters/leaves Engaged/Interacting zones

Coding of the video was done on a frame-by-frame basis with manual entry of time-stamped codes. Each code represents either a discrete event (e.g. the start of typing or speaking, stepping, seeing a person enter a particular zone) or a transition between states in each of the three behavioral classes (e.g. approaching, standing, walking away).

Our fine-grained behavioral analysis focused on the initiation, maintenance, and termination of social interactions. A statistical analysis was used to describe the cumulative occurrence of events, while a lag-sequential analysis was

used to determine the co-occurrence of different events within a small time window (e.g. 2-5 seconds).

There was at least one person interacting with the robot (starting by approaching, looking at, or standing next to it; interacting by typing or attending to someone who was typing; and ending by walking away) for 4.13 hours out of 12.51 coded hours of video. The robot, meanwhile, was in the interacting or greeting state for 4.08 hours, of which 3.15 hours overlap with the actual 4.13 hours of interaction. Therefore, 22.8% of the robot's interactive behavior was directed toward people who were not interacting with the robot, and the robot did not exhibit interactive behavior for 23.7% of the time that people were interacting with it. We would like to reduce both of these percentages for more targeted interactive behavior.

We first look at this difference between the robot's perception of the social state of the world and our description of actual physical interactions; our goal is to determine the limits of our current model of engagement and to consider how it might be improved. Next, we look at the temporal relationship between robot behavior and ensuing human behavior; here, our goal is to determine the most appropriate behaviors for the robot to perform with people at different levels of engagement. Finally, we present a number of observations regarding human interest in the robot and discuss how these might be inferred automatically, in order to determine whom the robot might be able to attract into interacting.

A. The robot's social view of the world

By coding people's behaviors as well as the robot's internal state, we were able to compare the robot's perception of people's social engagement with our observed description of what was really happening.

The Roboceptionist saw 1500 people enter the Engaged zone, and of those, 772 went on to enter the Interacting zone. In our observational analysis, on the other hand, we saw people stand next to or close to the robot ("engaged") 174 times, and they stood in front of the robot ("interacting") 195 times. These numbers fall in line with what the Engaged and Interacting zones were designed for; that is, interacting requires standing in front of the robot, and standing in the Engaged zone implies deciding whether to interact or observing an already occurring interaction. The large difference in the robot's perception is a result of people moving through these regions. 485 people passed directly next to the booth, possibly entering the Interacting zone, and 2458 more people passed close to the booth, some of whom may have passed through the Engaged zone.

This misperception has an effect on the robot's interactive behaviors, which consist of turning and speaking. Both behaviors may be used when a person enters the Engaged zone and the robot is otherwise idle. Their purpose is to begin an interaction or to attract someone who may be considering interacting with the robot. The behaviors are generally appropriately directed toward a person (moving or stationary) but are occasionally directed elsewhere or are performed too late. In the case of turning, this may be because of sensor noise, software lag, or a fast-moving person. 66% of the robot's 816 turns toward a person were

directed at people who passed the robot, and 51% of these occurred 3 or more seconds after the person passed the robot and would have been unable to see the action. Overactive turning is perhaps acceptable, as humans often turn to look at salient events in their environment without the expectation that this will result in some sort of interaction. In the case of the Roboceptionist, however, the turning is often accompanied by speaking.

It is reasonable for the robot to speak to people that it considers “engaged” or “interacting;” however, we see that the robot applies these labels quite liberally. Unlike head-turning, we usually do expect speech actions to be reciprocated, yet 54% of the robot’s directed greetings are not followed by human interactive behavior. We have learned, anecdotally, that people are somewhat disturbed when the robot addresses them as they pass the booth with no intention to interact; returning the greeting is burdensome, and they may feel “bad” about not taking the time to do it. This verbal overactivity, then, is a shortcoming of both the engagement model and of our selection of behaviors that accompany it. Motion through the regions must be considered differently from people who are stationary in the regions, in order to reduce the number of people classified as Engaged. Specifically, a person’s *direction* of motion should be used in addition to their static location. Using a short history of a person’s location to estimate a motion vector, it may be possible to infer that a person is approaching the booth from learned (or designed) rules. This would serve as a more discriminating estimate of initial engagement with the robot.

B. Effects of robot behavior on human behavior

As discussed, the robot makes turning motions that may or may not be properly directed toward people in the environment. In the case of speaking, the robot additionally makes periodic “phone calls” to appear active. We observed significant differences between people’s responses to properly directed interactive behaviors and their responses to undirected behaviors.

Out of the 816 times that the robot turned to a person, 26% were followed within 5 seconds by an interactive action by a person (e.g. approaching, turning to, or typing to the robot). On the other hand, when the robot turned elsewhere while there was a person present, they performed an interactive behavior only 10% of the time.

Similarly, when the robot spoke to people, they were more likely to interact with it than if it had been speaking on the phone. Out of 741 times that the robot addressed a person, 48% were followed by an interactive behavior. When the robot spoke on the phone and there was a person present, they only interacted 15% of the time.

Currently, turning and speaking are both used as responses to human arrival. People are more responsive to speaking than to turning (48% of directed verbalizations being followed by human interactive behavior, versus 26% of directed turns), but excessive speaking is more costly (in terms of human comfort, as discussed) than excessive turning. We therefore suggest separating the robot’s turning and speaking behaviors in such a way that the former is used liberally and the latter is used more conservatively, and that they

are used at different times in the interaction. Motion, in the form of turning, is appropriate for *anticipation* of interaction, as it is unobtrusive to passersby and attractive to curious potential interactors. It may be used more often as long as it is used earlier (to avoid the delay described above), and it should use the modified engagement model that considers a person’s direction of movement in order to capture as many potential interactors as possible. Audible speaking, on the other hand, is appropriate as a *response* when the robot is rather confident that a human is initiating an interaction.

C. Human interest in robot

In order to attract those people who may be considering interacting, and to account for the delay in the robot’s behaviors as described above, it is necessary to anticipate people’s level of interest earlier than we currently do. Looking for people who are approaching in the direction of the booth is one method, but we looked at other factors as well.

We classified passersby according to approximate distance from the booth. *Next to* is within 2ft of the booth, *close* is 2-4ft from the booth, and *far* is > 4ft from the booth. We found that passersby were slightly but significantly more likely to look at the Roboceptionist if they were close to the booth (33%) than if they were next to or far from it (both 29%) ($\chi^2(2, N = 4196) = 7.367, p = 0.0251$). This suggests that there may be a comfortable distance at which to establish eye contact with the robot; closer is too intimate, and further results in lower visual detail. Given the difficulty of estimating gaze for people who are far away (and moving), the possibility that people at a certain distance are more likely to be paying attention to the robot is a useful piece of information.

We expected that a person’s speed might indicate their likelihood of interacting; that is, people who are moving slower would be more likely to interact. Having coded approaches as slower than, close to, or faster than average walking speed, we found that, indeed, almost all people who interacted with the robot approached at average or below average speed. As the speed of passersby was not coded, these results are still inconclusive. However, it is possible that a simple rule incorporating speed, perhaps learned from logged sensor data, might provide reliable information regarding a person’s future engagement with the robot.

We expected that passersby would be more interested in the robot (and therefore look at it more) when there was already someone interacting with the robot, but we found no significant difference. We did, however, find a significant portion of interactions (46%) to consist of groups of more than one interactor. It is often the case that a person is drawn into interacting with the robot not by the robot itself, but by another person. This suggests an opportunity, if multiple people are present, for the robot to initiate interactions with other members of a group.

V. DISCUSSION & FUTURE WORK

People who encounter the Roboceptionist may be roughly divided into three classes: people who will approach the robot to interact with it regardless of its initial behavior, people who will not interact (passersby), and people who are undecided. For people who will interact, there is currently a delay in the robot’s behaviors. During their interactions,

especially those with groups, we also see a great deal of movement for which a stationary model of engagement is insufficient. For people who will not interact with the robot, the overactive verbalizing can be jarring. For people who are undecided, on the other hand, the robot is not active enough: there is a missed opportunity to draw in people who are ripe for engagement (e.g. by turning to them earlier).

The robot was designed mostly with one-on-one interactions in mind. It is apparent from our observations that interaction with groups makes up a significant portion of the Roboceptionist's interaction time. It is necessary to more carefully examine the patterns of human behavior with the robot and with each other when part of a group. The spatial movement of group members is different from that of individual actors, and this needs to be taken into account when estimating social engagement. Furthermore, the behaviors of the robot should be suitable for a number of interactors, within our constraints that only one can be typing to the robot at any time. It is unclear what types of behaviors to exhibit toward a group of people, and how to account for the fact that they often take turns typing to the robot.

To the extent that gaze and even gestures may be used to obtain information about engagement, we continue to apply computer vision methods to sensing cues beyond location and movement. However, even if we had accurate head pose estimation, it is unclear how this information should feed an estimate of engagement, as people frequently look around. Unfortunately, long-distance perception of human head pose and gaze is still impossible. Selectively attracting interested passersby through directed interaction is therefore unlikely.

However, in the process of coding the videos, we were often able to "guess" from a person's early motions whether they would eventually interact with the robot. The fact that humans are very good at anticipating social engagement from low-resolution video suggests to us that movement through the social space may be as useful, if not more useful, than finer cues such as head pose or gaze. We expect that a more intelligent use of the laser tracker data, in which movement and directionality over a short period of time are considered before a person becomes stationary, can result in earlier anticipation of social interaction (for temporally appropriate generation of initial cues), more selective generation of verbal cues, and more effective maintenance of ongoing interactions.

In short, we envision a modified engagement model along these lines:

- Present: not of interest for interaction; moving quickly or in a direction tangential to the robot's booth.
- Interested: moving toward the robot, or moving slowly in the vicinity. The robot should turn its head toward Interested people.
- Engaged: mostly stationary near the booth, perhaps with a detected face. The robot may speak to Engaged people.
- Interacting: probably responsible for typing that has occurred. Multiple people may be considered Interacting.

VI. CONCLUSION

We have reviewed a number of models of social engagement that take into account the relative spatial configuration

of interaction partners. We presented the implementation of a comparable model on a social robot. Distance, direction, and head visibility and pose are taken into account for selecting appropriate interactive behaviors. Observational behavioral analysis of this system in practice has demonstrated that a static location-based model of social engagement, such as those described in much of the literature, is insufficient for naturalistic regulation of social interactions. We suggest that direction and speed of motion are more appropriate measures of engagement than location alone, and that movement and speech should have different roles in the regulation of interaction.

Acknowledgements

This project was partially supported by National Science Foundation ITR projects #IIS-0329014 and #SES-0522630 and a GRFP.

REFERENCES

- [1] M. L. Patterson, *Nonverbal Behavior: A Functional Perspective*. New York: Springer-Verlag, 1983.
- [2] M. Rauterberg, M. Dtwyler, and M. Sperisen, "From competition to collaboration through a shared social space," in *Proceedings of the East-West International Conference on Human-Computer Interaction (EWHCI95)*, B. Blumental, J. Gornostae, and C. Unger, Eds., vol. II, 1995, pp. 94–101.
- [3] U. Frith and C. Frith, "The biological basis of social interaction," *Current Directions in Psychological Science*, vol. 10, pp. 151–155, October 2001.
- [4] E. T. Hall, *The hidden dimension: man's use of space in public and private*. New York: Doubleday, 1966.
- [5] E. T. Hall, "Proxemics," *Current Anthropology*, vol. 9, pp. 83–108, 1968.
- [6] M. L. Patterson, "Spatial factors in social interactions," *Human Relations*, vol. 21, pp. 351–361, 1968.
- [7] M. L. Walters, K. Dautenhahn, R. te Boekhorst, K. L. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry, "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," in *Proceedings of IEEE Ro-man*, 2005, pp. 347–352.
- [8] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. te Boekhorst, C. Nehaniv, I. Werry, and D. Lee, "Close encounters: Spatial distances between people and a robot of mechanistic appearance," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, Tsukuba, Japan, 2005, pp. 450–455.
- [9] A. E. Schefflen, *Micro-Territories in Human Interaction*. Mouton, 1975, pp. 159–173.
- [10] S. R. Langton, R. J. Watt, and V. Bruce, "Do the eyes have it? Cues to the direction of social attention," *Trends in Cognitive Sciences*, vol. 2, no. 2, February 2000.
- [11] C. Breazeal, *Designing Sociable Robots*. MIT Press, 2002.
- [12] C. L. Sidner, C. Lee, C. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, May 2005.
- [13] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, and J. Wang, "Designing robots for long-term social interaction," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS '05)*, August 2005.
- [14] R. Simmons et al., "GRACE: An autonomous robot for the AAAI Robot Challenge," *AAAI Magazine*, vol. 24, no. 2, pp. 51–72, 2003.
- [15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition*, 2001.
- [16] G. R. Bradski, "Computer vision face tracking as a component of a perceptual user interface," in *Applications of Computer Vision*, 1998.
- [17] M. P. Michalowski and R. Simmons, "Multimodal person tracking and attention classification," in *Proceedings of the Conference on Human-Robot Interaction (HRI 2006)*, Salt Lake City, Utah, March 2006.
- [18] "Noldus Information Technology: The Observer," URL: <http://www.noldus.com/products/observer>.