The Design of Gaze Behavior for Social Robots

Ph.D. Dissertation Proposal

25 June 2008

Bilge Mutlu

Human Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
bilge@cs.cmu.edu

Thesis committee:

Jodi Forlizzi (Co-chair), Human-Computer Interaction Institute
Jessica Hodgins (Co-chair), Robotics Institute and Computer Science Department
Sara Kiesler (Member), Human-Computer Interaction Institute
Justine Cassell (External Member), Northwestern University

Abstract

Humanlike robots are designed to communicate using human verbal and nonverbal language and engage in conversations with people where social gaze cues play an important role. While research in human-robot interaction has shown that people understand these cues and interpret them as valid signals for human communication, whether social gaze cues can serve as an effective communicative mechanism that leads to significant social outcomes remains unknown. Furthermore, work in this area lacks theoretical and empirical foundations for the design of communicative mechanisms to work with the human communicative system. The main research questions in this problem space are as follows: What are the design variables for social gaze behavior? How do we design gaze behavior for social robots? Can designed behaviors lead to positive, significant social outcomes in human-robot interaction such as better learning, stronger affiliation, and active participation in conversations?

This dissertation seeks to find answers to these questions through (1) exploring the design space to identify design variables for social gaze, (2) adapting an approach based on modeling human behavior to designing robot behaviors, and (3) evaluating the social outcome of designed behaviors in three studies that focus on different functions of social gaze behavior. Study I focused on designing gaze behavior for communication of attention and found strong learning effects induced by a simple manipulation of how much a robot looks at one person more than the other. Study II looked at how different patterns of a robot's gaze behavior lead to different levels of participation in a conversation with the robot and found strong effects of gaze cues in behavioral and subjective measures of participation, attentiveness, liking, and feelings of groupness. These studies will be followed by a proposed third study that will look at how cues from a robot's gaze might be used in establishing joint attention and evaluate how changes in the robot's gaze at objects affect foreign language learning.

This research will contribute to (1) the design of robotic systems with a set of design variables for social gaze and a theoretically and empirically grounded methodology for the design of communicative mechanisms for social technology and (2) human communication research with a deeper understanding of how human communicative mechanisms respond to artificially created social stimuli.

1. Introduction

In the future, robots might serve a variety of tasks as informational agents at public spaces, caregivers or companions for the elderly, educational peers for children, and so on. These robots would have to communicate using human verbal and nonverbal language and carry out conversations with people where aspects of gaze will play an important role. For example, suppose an educational robot's task is to tell stories at a primary school and make sure that everyone in the class is following the story. What would the robot do if it realized that one of the students is not attending to its story? What would human teachers do? Let's read the following excerpt (Woolfolk and Brooks, 1985):

Professor: *How do you know when your teacher really means what she says?*

Third Grader: *Well, her eyes get big and round and she looks right at us. She doesn't move and her voice is a little louder, but she talks kinda slowly. Sometimes she stands over us and looks down at us.*

Professor: *What happens then?*

Third Grader: *The class does what she wants!*

As seen in the excerpt above, human teachers change aspects of their verbal and nonverbal language to communicate to the student that he should be attending to the story. In fact, research has shown that simply looking at that student more improves his learning (Otteson and Otteson, 1980; Sherwood, 1987). What should the robot in our scenario do? The obvious hypothesis is that the robot should look at the student more. But, do human communicative mechanisms work with robots?

Researchers have been developing robotic systems that support human communicative mechanisms for the last decade (Breazeal, 1998; Nourbakhsh, 1999; Brooks et al., 1999; Scassellati, 2001; Kanda et al., 2002; Dautenhahn et al., 2002; Pineau et al., 2003; Minato et al., 2004). A number of studies have shown the importance of the use of nonverbal behaviors in human-robot communication (Kanda et al., 2003; Breazeal et al., 2005; Sidner et al., 2005, 2006) and a few of these studies focused on aspects of gaze (Imai et al., 2002; Sidner et al., 2004; Yoshikawa et al., 2006; Yamazaki et al., 2008). Imai and his colleagues (2002) showed that people can accurately perceive a robot's orientation of attention using cues from its gaze. When robot's gaze behavior was contingent with that of participants, people's perception of the robot's orientation of attention was stronger (Yoshikawa et al., 2006). In a study Sidner and her colleagues (2004) conducted, the robot's use of gaze cues and gestures significantly increased people's engagement as well as their use of gaze cues to communicate with the robot. Yamazaki and her colleagues (2008) further showed that when a robot followed simple rules of conversational turn-taking to coordinate its gaze behavior and verbal utterances, people were more likely to display nonverbal behaviors at turn boundaries.

While these studies provide strong evidence that robot gaze can serve as an effective communicative mechanism, a systematic study of how robot gaze could lead to significant social outcomes in different situations is still lacking. For instance, can robot gaze affect learning? Can a robot use gaze cues to regulate turn-taking and conversational participation? Can robot gaze help people in making associations between two pieces of information? Furthermore, how social gaze behavior should be designed to work with human communicative mechanisms needs further exploration.

This dissertation seeks to find answers to these questions through developing a theoretical understanding of how robot gaze might serve as a communicative mechanism, a set of design variables that designers of social robots can manipulate to achieve a desired social outcome, and a methodology for extracting knowledge from human communication in order to design social behaviors for robots. This knowledge will contribute to (1) the design of robotic systems with a set of design variables for social gaze and a theoretically and empirically grounded methodology for the design of communicative mechanisms for social technology and (2) human communication research with a deeper understanding of how human communicative mechanisms respond to artificially-created social stimuli.

The next section describes the research context with scenarios that motivate the research questions and the approach taken for addressing these questions. Section 3 provides a review of related work on social gaze from literatures on human communication research, human-computer interaction, and robotics. Section 4 provides details on the design of and results from completed empirical studies and an outline of the proposed study. Section 5 provides a general discussion of the results, limitations of the work, and future research. The last section outlines the schedule for the proposed work.

2. Research Context and Approach

Because few robots exist in our everyday lives and most research in robotics builds on a future vision for everyday use of humanoid companions and assistants, the research questions posed in this dissertation are motivated by a set of future scenarios. Below are three scenarios that provide context for the three empirical studies that look at how robot gaze might serve as a communicative mechanism and methodological inquiry into the design of humanlike behavior.

2.1. Scenario 1

> *Melanie works at the Liberty Elementary School in Pittsburgh, Pennsylvania as an English instructor. ASIMO (Sakagami et al., 2002) also works at this school as an aid to English and history instructors. Melanie teaches English to third graders and has three classes a week - on Mondays, Wednesdays, and Fridays. On Mondays, ASIMO tells the class stories of Melanie's choice. On Wednesdays, Melanie discusses the story with the class and asks the class to write a one-page review of the story and bring it to class on Friday.*
>
> *Recently, Melanie realizes that Justin, one the students in her third grade class, has not been participating in the discussions and his essays are very brief. She talks to Justin and has a phone conversation with his mother to see if there is anything bothering him or any trouble at home. Nothing seems to stand out. She talks to the history and math teachers about the recent change in Justin's attention but both instructors don't seem to notice a change.*
>
> *Melanie decides that Justin might be distracted, or he might even be losing interest in English. She tells ASIMO to pay particular attention to Justin during storytelling. She hopes to monitor Justin's behavior and direct his attention to class.*

Melanie's problem is not uncommon. In fact, research on classroom behavior has shown that classroom inattentiveness has negative effects on literacy (Rowe and Rowe, 1999). However, teachers can positively affect student attentiveness using aspects of nonverbal language such as interpersonal space, gestures, gaze, and tone of voice (Woolfolk and Brooks, 1985). Gaze behavior, in particular, is shown to improve learning in primary schoolers (Otteson and Otteson, 1980) and college students (Sherwood, 1987). Should ASIMO simply look at Justin more to direct his attention to class?

Researchers have developed pedagogical virtual agents that direct students' attention using gaze cues and gestures (Rickel and Johnson, 1999, Ryokai et al., 2003). The use of gaze cues by robots is also shown to have a positive effect on engagement (Bruce et al., 2002; Sidner et al., 2004). However, whether cues from the gaze of a robot can direct attention in a way that it leads to better learning is unknown.

Can robot gaze communicate attention and lead to better learning? How can we design robot gaze behavior to attract attention and improve learning? What are the design variables? The first study in this dissertation sought to find answers to these questions.

2.2. Scenario 2

> *Aiko is a shopper at the Namba Parks shopping mall in Osaka where Robovie (Ishiguro et al., 2001) serves as an information booth attendant. Aiko is trying to find the closest Muji store and wants to know if the store also sells furniture. She approaches Robovie to inquire about the shop.*

The conversational situation that Robovie will have to manage in this scenario is a two-party conversation where Robovie and Aiko will take turns to play the roles of speaker and addressee (Clark, 1996).

> *As Aiko and Robovie talk about how to get to the Muji store, another shopper, Yukio, approaches Robovie's booth. Yukio wants to get a program of this month's shows at the amphitheater. When Yukio approaches the information booth, Robovie acknowledges Yukio's presence with a short glance but turns back to Aiko signaling to him that he has to wait until the conversation with Aiko is over.*

What is different in this scenario is the addition of a non-participant (Clark, 1996) into the social situation who is playing the role of a bystander (Goffman, 1979).

> *After Robovie's conversation with Yukio is over, a couple, Katsu and Mari, approach the booth inquiring about the Korean restaurants in the mall. Robovie asks Katsu and Mari a few questions on their food preferences and understanding that they don't like spicy food, he leads the couple to Shijan located on the sixth floor of the mall.*

This last situation portrays a three-party conversation where Robovie plays the role of the speaker and Katsu and Mari are addressees for most of the conversation. While Robovie needs to carry on conversations in all of these situations, the differences in levels of participation require him to also provide the appropriate social signals to regulate each person's conversational role. When Yukio approaches the booth, Robovie has to make sure that Aiko's status as addressee doesn't change, but that he also signals to Yukio that his presence is acknowledged and approved. In talking to Katsu and Mari, he has to make sure that both feel equally respected as addressees.

Considerable evidence suggests that people use gaze cues to perform this social-regulative behavior (Bales et al., 1951; Schegloff, 1968; Sacks et al., 1974; Goodwin, 1981). Research has shown that these cues are also effective in regulating conversational participation when they are used by virtual agents (Bailenson et al., 2001, 2005; Rehm and Andre, 2005). While robot gaze is shown to be effective in performing conversational functions such as supporting turn-taking behavior (Kuno et al., 2007; Yamazaki et al., 2008) and showing appropriate listening behavior (Trafton et al., 2008), how these cues might shape different forms of participation remains unexplored. Furthermore, whether the cues used by humans can be carried over to robots to create the desired social outcome is unknown.

Can simple cues from a robot's gaze lead to different forms of conversational participation? How can we design gaze behavior that leads to such outcome? And what are the design variables? The second study addressed these questions.

2.3. Scenario 3

Mai is a recent college graduate who just accepted a job offer at the Osaka International Airport as a customer service representative. Her new job requires Mai to attend to a training program including sharpening her English skills before she can start working.

To study English, Mai goes to a second language center near the Osaka University campus where Sakura works as an instructor. Mai attends Sakura's class twice a week to improve her grammar. Mai also studies vocabulary for an hour with Geminoid (Nishio et al., 2007), an android robot that works at the language center. Geminoid's job is to help students with learning and practicing new vocabulary. When he works with a student, he and the student sit across a table with a computer screen next to them. As pictures of objects and places appear on the screen, Geminoid uses different techniques to assess whether the student knows the English word for what is shown on the screen. If the student recalls the word, he helps the student practice the word in an example. If not, he helps the student with learning the word with its correct pronunciation.

After working with Mai for a couple of weeks, Geminoid realizes that Mai has trouble recalling the names of particular objects. For instance, the word "cabbage" is very difficult for Mai to remember.

What can Geminoid do to help Mai recall these difficult words? Lazaraton (2004) conducted an microanalysis of an English as a second language teacher and found that nonverbal behaviors played an important role in teaching new vocabulary. Research on language development showed that the ability to follow a teacher's gaze and establish joint attention towards objects significantly improves vocabulary learning through pairing the observed object and its vocalized name (Baldwin, 1995; Mundy and Gomes, 1998).

Could Mai learn these difficult words at home using audio or video recordings of Sakura or Geminoid uttering the words? Kuhl and her colleagues (2003) compared foreign language learning between infants exposed to the language through interaction with a physically-present speaker and those exposed to it through audio and audio/visual material. Their results showed that no learning took place when infants were exposed the language through media, while more learning took place when they interacted with the native speaker in comparison to the control group suggesting that social interaction is a necessary condition for foreign language learning.

Researchers have built robots with the ability to establish joint attention (Scassellati, 1999; Nagai et al., 2003). Imai and his colleagues (2003) showed that a robot can effectively build joint attention with people using gaze cues, pointing gestures, and deictic references to objects in the environment and experimentally showed the importance of gaze cues in establishing joint attention. However, whether gaze cues from a robot can serve as an effective communicative mechanism that leads to better learning is unknown.

Can cues from a robot's gaze help in establishing joint attention in a way that it leads to better language learning? How can we design such cues? And, again, what are the design variables? The proposed third study will seek answers to these questions.

2.4. Research Approach

The studies that are motivated by these scenarios follow an interdisciplinary approach combining knowledge and methods from a number of disciplines including design, psychology, linguistics, and computer science and a three-stage process of design, implementation, and evaluation. The first stage involves extracting and gathering theoretical and empirical knowledge on human behavior and using this knowledge to design humanlike behavior for a set of social contexts. This stage is informed by knowledge from communication, social and developmental psychology, and sociolinguistics, and employs methods from interaction design, social psychology, computer graphics, and machine learning. In the second stage, the design specifications are converted into computer algorithms implemented on robotic platforms with the appropriate capabilities to produce the designed behavior. This stage is informed by knowledge and methods from computer science, particularly artificial intelligence and robotics. Finally, the implemented behaviors are evaluated for communicative efficacy in controlled laboratory experiments where human subjects are placed in a social situation that allows for testing hypotheses of human communication. This stage is informed by knowledge and methods from communication and social psychology. Below are details of each stage in this process.

2.4.1. Theoretically and Empirically Grounded Design

The goal of this stage is to ground the design of robotic gaze behavior in theory on human communication and empirical data collected in the social context of interest. Grounding design decisions in empirical data is particularly important from a design standpoint because human communication theory looks at human behavior from the standpoint of understanding its significance for social interaction; therefore, theory is not fully informative in recreating human behavior. For instance, human communication theory tells us that people occasionally look away from their partners during conversation, but does not tell us where they look, which is an important variable in the design of gaze behavior.

Theoretical grounding is done through using knowledge from human communication research in the design of gaze behavior. For instance, Wang and his colleagues (2005) designed a pedagogical agent using theory on politeness that was developed by Brown and Levinson (1987). Similarly, robots developed by Matsusaka and his colleagues (2001) and Kuno and his colleagues (2007) used a theory on conversational turn-taking suggested by Sacks and his colleagues (1974).

Empirical grounding is done through building models of human behavior using various forms of data such as a distribution of how long a person looks at a particular target. Examples of empirically-grounded design is less frequent: Cassell and her colleagues (2006) collected data from 28 individuals to build a model of how people use gestures to make spatial references in giving direction and used their model to design gestures for a conversational agent called NUMACK. Similarly, Kanda and his colleagues (2007) used data from 25 dyads to model the delay between one's deictic gesture and and the other's changing orientation towards the direction of the reference and designed orienting behaviors of a robot based on this model.

The theoretical knowledge that informed the design process in this dissertation was created through a broad survey of literature on gaze in sociology, social and cognitive psychology, communication, linguistics, sociolinguistics, and neurophysiology. The process of empirical grounding was informed mainly by methods of behavioral and discourse analysis. Data collection techniques involved audio and video recordings, gaze target information obtained from eye-trackers, and head and body orientation information obtained from motion-capture systems.

2.4.2. Algorithmic Implementation

The goal of the implementation stage is to create computer algorithms that can automatically generate gaze behaviors based on designed models. The approach to creating these algorithms is to integrate rules from theory (e.g. that a speaker looks at an addressee before a turn as suggested by Sacks and his colleagues, 1974) and findings from empirical data (e.g. that the distribution of the length of a speaker's gaze at an addressee follows a two-parameter continuous distribution with values θ and k) into a hybrid rule-based/stochastic algorithm.

2.4.3. Experimental Evaluation

The evaluation of designed behaviors are done through controlled laboratory studies where human participants are asked to perform an experimental task with robots. These tasks are designed to allow participants to immerse themselves in an experience with the robot for a relatively long time. Gill and his colleagues (1998) suggest that people have poor mental representations of those with whom they have little experience; therefore, they cannot confidently make judgements about them. For example, "a hiring committee may be intrigued by a job candidate on skimming her vita but will hire her only after boosting its confidence through conversations with the candidate herself" (Gill et al., 1998). In judging robots, people might rely on similarly poor mental representations because they have very little experience with robots suggesting that the more experience people have with a robot, the richer their mental representation of the robot and the more confident and consistent their judgments of the robot would be. Furthermore, Lee and her colleagues (2005) showed that people's beliefs about a robot's capabilities and knowledge are developed mostly through extrapolating from their own knowledge. In this process, they rely on simple cues from the robot such as the robot's origin and language. Therefore, the more cues they get from the robot, the more consistent their judgments of the robot would be.

The experiences with robots are designed to follow social "interaction rituals" (Goffman, 1971) or "social episodes" (Forgas, 1979), particularly rituals of greeting and leave-taking. Goffman (1955) describes greetings as serving "to clarify and fix the roles that participants will take during the occasion of the talk and to commit participants to these roles," while leave-taking rituals as providing "a way of unambiguously terminating the encounter." Following these rituals is particularly important in the context of this research because gaze cues are shown to play a significant role in producing these behaviors (Kendon and Ferber, 1973). Therefore, in each sequence of interaction with human participants, robots followed greeting and leave-taking rituals.

The technical implementation of the experiments combined automatic algorithms (e.g. gaze generation algorithms) and Wizard-of-Oz techniques (e.g. fixing people's location to avoid the cost of implementing face tracking). Implementation details are provided in the study descriptions.

2.5. Expected Contributions

This work impacts the design of robotic systems and research on human communication including the fields of human-robot interaction (HRI), human-computer interaction (HCI), computer-supported collaborative work (CSCW), and computer-mediated communication (CMC). For the design of robotic systems, it provides designers with a set of design variables that they can use to create social gaze behavior for a variety of social situations. It also contributes with a theoretically and empirically grounded methodology for modeling human behavior in a particular social situation and recreating it for humanlike platforms to act in similar situations. This work will provide human communication research with a deeper understanding of how human communicative mechanisms respond to artificially created social stimuli and the social outcome of this response. Furthermore, the process of designing humanlike behavior through careful analyses of human behavior might lead to new knowledge. Specifically, this dissertation provides:

Methodological

- A methodology for grounding design decisions in theory and empirical data in designing humanlike behavior for robots.
- An integrated process for the design, implementation, and evaluation of humanlike behavior for social robots.

Practical

- A set of design variables for creating social gaze behavior for robots.
- An algorithm for synchronizing a robot's gaze behavior with its speech to address an audience of two in an oratory situation (e.g. storytelling).
- An algorithm for synchronizing a robot's gaze behavior with its speech to achieve floor management activities such as turn-yielding and turn-taking, and regulate participation in different conversational configurations.

Theoretical

- An understanding of fundamental aspects of human gaze behavior (i.e. where people look at, how long and how frequently they gaze at each location, distributions of how long each location is looked at, etc.), specifically in the context of storytelling and casual conversations with different levels of participation.
- An understanding of gaze signals that regulate conversational participation.
- An understanding of the relationship between patterns of gaze shifts and thematic structure of casual conversations.

- Evidence that robot gaze can lead to significant social outcomes such as better recall and participation in a conversation.
- Evidence that how much a robot looks at an individual affects that individual's performance in recalling the information presented by the robot.
- Evidence of strong gender effects on the perception of gaze behavior, particularly on positive evaluations of the robot.
- Evidence that turn-taking signals in a robot's gaze are correctly interpreted by people and lead to fluid, natural sequences of turn-taking in human-robot conversations.
- Evidence that gaze cues alone (i.e. whether or not a robot looks at a person at interactionally significant points in a conversation) can lead to varying levels of task attentiveness, liking of the robot, and feelings of groupness.

This section described the scenarios that outline the research context and motive research questions. An overview of the approach taken to addressing these questions and a summary of expected contributions are also provided. The next section provides a review of related work on gaze from research on human communication, embodied conversational agents, and humanlike robots focusing specifically on the social contexts outlined by the scenarios presented in this section.

3. Background

Research in Human-Computer Interaction has shown that people respond to computers in fundamentally social ways (Nass et al., 1993; Reeves and Nass, 1996; Sproull et al., 1996). Nass and his colleagues (1993) proposed a framework called "Computers as Social Actors" (CASA) and showed in a series of studies that people make attributions of gender stereotypes (Nass et al., 1997) and personality to (Moon and Nass, 1996), respond to flattery (Fogg and Nass, 1997) and humor from (Morkes et al., 1998), give credit to (Moon and Nass, 1998), and show politeness towards (Nass et al., 1999) computers. They argued that these responses are mindless and automatic (Nass and Moon, 2000) following the proposition that people automatically respond to relevant social stimuli (i.e. when the stimulus follows common norms and patterns of interaction) (Langer et al., 1978; Bargh et al., 1996).

While Nass and his colleagues (1995) argue that a minimal set of cues provides sufficient social stimuli for people to respond to computers in the same way they would respond to people, Cassell (2001) argues that humanlike cues are a necessity because people have a propensity to seek an embodiment for intelligence and a social locus of attention. In support of the latter argument, Sproull and her colleagues (1996) showed that explicit humanlike cues such as a humanlike face presented on a computer screen as opposed to a text-based computer led people to make stronger attributions of personality and present themselves more positively to the computer, and feel more aroused by the computer. These results suggest that humanlike cues provide a sense of presence and disambiguate what communicative channels are open to people (e.g. speech, gaze, facial expressions, gesture, etc.) making communication more fluent and allowing people to have a more accurate mental representation of the computer (Kiesler, 2005).

This dissertation builds on these works, but focuses particularly on understanding how human communication mechanisms respond to humanlike cues, specifically those from a robot's gaze. It is informed by literature on social gaze behavior in human communication research, embodied conversational agents, and robotics. The literature survey below encompasses a review of related work from all three literatures, with a focus on the social contexts provided by the scenarios presented in the previous section.

3.1. Gaze Cues in Social Interaction

During social interaction, people look at others for an average of 61% of the time—longer than they speak (Argyle and Ingham, 1972). Through gazing at others, people study others' behavior and appearance and look particularly in the region of their eyes (Cook, 1977). The eyes are such an important source of social information that even infants aged four weeks are able to locate the eyes of an observer (Wolff, 1963). Newborns prefer faces with visible eyes (Batki et al., 2000) and moving pupils (Farroni et al., 2002). For any social interaction to be initiated and maintained, parties need to establish eye-contact. Goffman (1963) argued that, through establishing eye-contact, people form "an ecological eye-to-eye huddle" through which they signal each other that they agree to engage in social interaction. Simmel (1921, as quoted in Argyle and Cook, 1976) describes this mutual behavior as "a wholly new and unique union between two people [that] represents the most perfect reciprocity in the entire field of human relationship."

Research has shown that people are extremely sensitive to being looked at (Gibson and Pick, 1963). This sensitivity may have evolved as a survival mechanism in order to detect whether a predator is attending (Emery, 2000). Neurophysiological evidence shows that this mechanism might be supported by a dedicated 'eye direction detector' in the brain (Baron-Cohen, 1995). Pictures of eyes (Bateson et al., 2006)—even simulated 'eyespots' on a computer screen (Haley and Fessler, 2005)—are found to influence people's decision making behavior. On the road, drivers and pedestrians move off more rapidly from stop lights when they are stared at (Ellsworth et al., 1972). Drivers are more likely to stop for hitchhikers who establish eye-contact with them (Sydner et al., 1974).

Research on gaze, which started in early 1960s, has shown that gaze behavior is tightly intertwined with many other aspects of social interaction. Gender (Exline, 1963; Argyle and Ingham, 1972; Bayliss et al., 2005), personality differences (Strongman & Champness, 1968; Kleck & Nuessle, 1968; Mobbs, 1968), conversational role (Exline & Winters, 1965; Kendon, 1967), the topic of conversation (Exline, 1963; Exline et al., 1965; Abele, 1986), whether interaction takes place in public (Goffman, 1963; Kendon, 1973; Kendon and Ferber, 1973), the familiarity of the parties (Exline, 1963; Noller, 1984), and many other factors are found to affect gaze behavior. The tight coupling between gaze behavior and many other aspects of social interaction has made the study of gaze behavior central to social psychology. Argyle and Cook (1976) argue that "any account of social behavior which fails to deal with the phenomena of gaze is quite inadequate."

3.2. What is Gaze?

Most human communication research literature on gaze is concerned with the direction of eyes. While eyes are the primary source of information on the direction of attention, social gaze involves a complex coordination of the eyes, the head, and body orientation that is sensitive to the social context (Emery, 2000; Frischen et al., 2007). For instance, when cues from eyes and head are congruent, people can interpret direction of attention faster than they can when cues are incongruent (Langton and Bruce, 1999).

In the absence of information from the eyes, cues such as head orientation, body posture, and pointing gestures might also indicate direction of attention (Langton et al., 2000). Research on neurophysiological aspects of gaze has shown that signals created in part of the brain that is responsible for processing social information by observing eye direction are stronger than those evoked by observing head direction, which suggests that directional information from gaze, head, and body cues might be combined hierarchically in a mechanism dedicated to detect another's direction of attention (Perrett et al., 1992). The existence of this hierarchy is supported by behavioral evidence. When people are at greater distances, head orientation becomes a stronger cue than information from the eyes in determining direction of attention (Von Cranach and Ellgring, 1973).

In addition to this hierarchical relationship, eye and head orientation might convey different social information. For instance, Gibson and Pick (1963) showed that people misjudge gaze direction of a target when the head is not facing them. Hietanen (2002) argues that this misjudgment is because "facing away" might be interpreted as "socially disinterested" delaying the processing of eye direction, supporting the hypothesis that head orientation and eye direction convey different messages.

3.3. Definitions of Social Gaze Behavior

A review of literature on gaze from different perspectives reveals a number of terms, concepts, and social situations where gaze plays a significant role. Below, definitions are provided for concepts that are relevant to this thesis.

- *Gaze, one-sided gaze, eye-gaze, looking at, visual orientation towards* - A looks at B in or between the eyes, or, more generally, in the upper half of the face (Cook, 1977).
- *Mutual gaze, eye-contact* - Both A and B look into each other's face, or eye region, thus acting simultaneously as sender and recipient (Von Cranach and Ellgring, 1973).
- *Averted gaze, gaze avoidance, gaze aversion, cut-off* - A avoids looking at B especially if being looked at, and/or moves the gaze away from B (Von Cranach and Ellgring, 1973; Emery, 2000).
- *Gaze following* - A detects B's direction of gaze and follows the line of sight of B to a point in space (Emery, 2000).
- *Joint attention, visual co-attention, deictic gaze* - A follows B's direction of attention to look at a fixed point in space (such as an object) (Butterworth, 1991).
- *Shared attention* - Both A and B look at a fixed point in space and are aware of each other's direction of attention (Baron-Cohen, 1995; Emery, 2000).

The remainder of this review focuses on three social functions of gaze behavior following the three scenarios presented earlier; communication of attention, regulating conversational participation, and establishing joint attention. In covering literature on each function, related work from research on human communication, humanlike virtual agents, and humanlike robots is reviewed.

3.4. Gaze Cues in Communication of Attention

An important aspect of human cognitive system is the ability to orient attention to information in the environment that is relevant to one's behavioral goals (Posner, 1980; Frischen et al., 2007). One of the most salient cues of this orientation is gaze direction (Baron-Cohen, 1995; Emery, 2000). People direct their gaze at each other to signal that their attention is directed at the other (Goffman, 1963). Being looked at by another produces an immediate heightening of arousal (Nichols & Champness, 1971; Patterson, 1976; Kleinke, 1986). This response is induced by an "eye-direction detector" (Baron-Cohen, 1995) or a "direction of attention director" (Perrett and Emery, 1994) located in the Superior Temporal Sulcus (STS), the part of human brain involved in deriving social meaning (Perrett et al., 1992; Baron-Cohen, 1995). The STS responds to being looked at more than it does to other gaze stimuli, as it is engaged in the processing of the social information (Pelphrey et al., 2004). This processing funnels the attention to the looker and might delay any response to other stimuli (Senju and Hasegawa, 2005).

3.4.1. Attention and Learning

The establishment of mutual orientation of attention can lead to significant social outcomes such as increased intimacy (Argyle and Dean, 1965; Patterson, 1976), attraction (Exline and Winters, 1966; Mason et al., 2005; Frischen et al., 2007), and attention (Langton et al., 2000), all of which contribute to increased attention on and better recall of verbal communication (Exline and Eldridge, 1967; Fry and Smith, 1975; Otteson and Otteson, 1980, Sherwood, 1987). For instance, primary schoolers who were looked at by a storyteller had better recall of the story than those who were not looked at (Otteson and Otteson, 1980). Sherwood (1987) replicated these results with college students. Fry and Smith (1975) showed that participants performed better in a digit-encoding task when the instructors made as much eye contact as possible while reading the instructions than when they made as little eye contact as possible.

In the classroom, significant differences were observed between experienced and/or effective teachers and inexperienced and/or ineffective teachers in the frequency of direct eye contact with students during the first week of class (Brooks, 1985). Woolfolk and Brooks (1985) suggest that teachers consciously and explicitly use gaze cues to attract the attention of their students. In fact, eye contact is found to be one of the main factors to increase the efficacy of verbal reprimands in the classroom (Van Houten et al., 1982).

3.4.2. Person Perception

Gaze cues are also used in evaluations of personality (Goffman, 1963; Kleck and Nuessle, 1968; Kendon and Cook, 1969; Cook and Smith, 1975). In general, positive evaluations of a partner increase consistently with the amount of gaze from zero to normal but decrease with too much gaze (Argyle et al., 1974). People who look at others 80% of the time are rated as more friendly, self-confident, natural, mature, and sincere, while those who look at others 15% of the time are perceived as cold, pessimistic, cautious, nervous, defensive, immature, evasive, submissive, indifferent, sensitive, and lacking confidence (Kleck and Nuessle, 1968; Cook and Smith, 1975).

People are sensitive to not only the amount of gaze they receive but also the patterns of gaze. We expect others to look in certain ways and are disturbed when we encounter unusual gaze patterns (Goffman, 1963). People who look in long, infrequent gazes are preferred over those who look in short, frequent ones (Kendon and Cook, 1969). People who are observed to move their eyes to establish eye-contact are evaluated more likable than those who are observed to break eye-contact (Mason et al., 2005).

Feelings of being looked at is also shown to affect economic decisions (Bateson et al., 2006; Haley and Fessler, 2005; Burnham and Hare, 2006). For instance, Bateson and others (2006) placed images of either a pair of eyes or flowers on an “honesty box” that was used at a school cafeteria to collect money for drinks and found that people paid nearly three times as much for their drinks when they saw images of eyes than when they saw images of flowers. Similar results were found in a study where participants saw images of MIT’s Kismet robot on their computer screen (Burnham and Hare, 2007) and in another study where their computer backdrop contained schematic eyes (Haley and Fessler, 2005).

3.4.3. Gaze Cues and Communication of Attention in Humanlike Virtual Agents

A number of studies looked at communication of attention in the design of humanlike virtual agents (Khullar and Badler, 2001; Peters et al., 2001; Peters, 2005; Peters and O'Sullivan, 2003). However, these studies focused mostly on the automatic production of gaze cues to communicate an agent's direction of attention based on the detection of salient stimuli in the environment or the attention level of a conversational partner, and did not consider how different aspects of the agent's gaze affected the partner's attention level or perceptions of and interactions with the agent. For instance, Khullar and Badler (2001) developed a model for automatically animating an agent's gaze direction based on the outputs of different cognitive mechanisms such as visual search and tracking. Peters and his colleagues (2005) developed a virtual agent with the ability to interpret its partner's level of interest and use gaze cues to provide feedback on its attention in order to maintain conversational flow.

In one study, however, Bailenson and his colleagues (2005) looked at how gaze cues of a speaker could be "augmented" in an immersive virtual environment to direct the attention of the speaker towards two listeners simultaneously. They compared participants' evaluations of the speaker across augmented and normal gaze conditions and found that women agreed with speaker's message more in the augmented gaze condition than in the normal gaze condition.

3.4.4. Gaze Cues and Communication of Attention in Humanlike Robots

Two studies in human-robot interaction looked at communication of attention. Bruce and her colleagues (2002) designed a social robot called Vikia with the ability to detect and orient its gaze at passersby in a hallway. They evaluated whether the ability to orient attention towards a person would increase people's likelihood to interact with the robot and found that passersby were more likely to stop when the robot oriented its attention at people than when it did not. Imai and his colleagues (2002) conducted an experiment where they seated eight participants in a circle and placed a humanlike robot at the center of the circle. As the robot oriented its gaze direction at different people, participants were asked whom they thought the robot was attending to. Their results showed that people had a high sensitivity to being looked at and could identify when the robot looked at them with 91% accuracy and when it looked at the person sitting next to them with 80% accuracy.

While these two studies provide evidence that gaze cues can communicate a robot's direction of attention and people respond to them as valid stimuli, whether manipulations on these cues such as increased gaze might lead to significant social outcomes such as better recall of information or more favorable evaluations of the robot is unknown.

In the next subsection, an overview of literature on the conversational use of gaze cues, particularly in regulating participation, intimacy, and role exchanges, and on the design of gaze cues for embodied conversational agents and humanlike robots is provided.

3.5. Gaze Cues in Regulating Conversational Participation

In using language, people work together as participants (Clark, 1996). The roles of the participants and how these roles might shift during social interaction, a phenomenon described by Goffman (1979) as "footing," are particularly important in understanding spoken discourse (Hymes, 1972; Hanks, 1996). At the core of these roles are those of the speaker and the addressee (Clark, 1996). While these roles might be fixed in some social settings (e.g. lectures), most conversational settings allow for shifting of roles. At any "moment" (Goffman, 1979) in a two-party conversation, one of the participants plays the role of the speaker and the other is the addressee. Conversations with more than two participants also involve side participants who are the "unaddressed recipients" of the speech (Goffman, 1979; Wilkes-Gibbs and Clark, 1992; Clark, 1996).

In addition to these "ratified participants" (Goffman, 1979), conversations might involve "non-participants" (Clark, 1996). For instance, there might be bystanders whose presence is acknowledged by the participants and who are observers of the conversation without being participants in it (Goffman, 1979; Clark and Carlson, 1982; Clark, 1996). There might also be hearers whose presence is not acknowledged by participants, but who are following the conversation closely, such as overhearers who are unintentionally listening to the conversation and eavesdroppers who have engineered the situation to purposefully listen to the conversation (Goffman, 1979). Figure 1 provides an abstract illustration of different levels of participation.

The direction of gaze plays an important role in initiating and maintaining conversational participation. In conversations, the gaze of a speaker towards another participant can signal that the speaker is addressing that participant (Sacks et al., 1974; Goodwin, 1981). In this situation, the speaker indicates a "communication target" (Bales et al., 1951). When there is no intended target (i.e. when a speaker is addressing a group), gazing at a participant long enough might create the belief that the speaker is addressing that participant (Bales, 1970). On the other hand, when there is an intended target and the speaker does not signal by means of gaze which target is being addressed, breakdowns might occur in the organization of conversational roles (Schegloff, 1968).

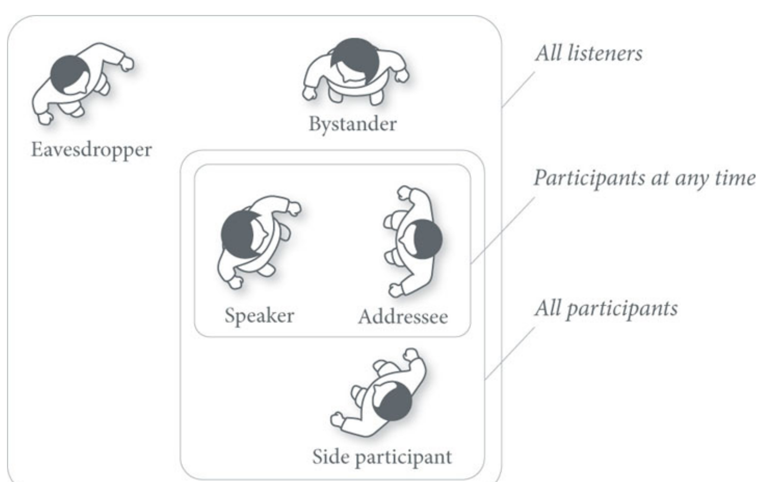


Figure 1. Levels of participation (adapted from Clark, 1996).

Listener gaze can also signal participation (Argyle and Cook, 1976; Goodwin, 1981). By looking at or away from others, people signal their social accessibility (Goffman, 1964). For instance, in therapy groups, four patterns of gaze behavior that signal accessibility were observed: In the "alert" state, the participant gazes at others and is responsive and open to interaction; The "huddled" state involves avoiding visual contact and displaying withdrawal from the group; In the "closed" state, the participant gazes away from the focus of the group, does not participate voluntarily, but is responsive when addressed; and the "away" state involves gazing into space signaling unresponsiveness to interaction (Callen et al., 1973).

Gaze direction also serves as an important cue in shifting roles in conversations through taking turns (Nielsen, 1962; Duncan, 1972; Sacks et al., 1974; Goodwin, 1980, 1981) and overlapping talk (Schegloff, 2000). For instance, speakers might look away from their addressees to indicate that they are in the process of constructing their speech and do not want to be interrupted while looking at their addressees might signal the end of a remark and the passing of the floor to another participant (Nielsen, 1962). In this context, the addressee at whom a speaker looks at the end of a remark would be more likely to take the role of the speaker next (Weisbrod, 1965 as described in Kendon, 1967). Shifting of roles might be delayed when remarks do not end with gazing at another participant (Kendon, 1967; Vertegaal et al., 2000). When gaze levels are particularly low, such as in a conversation between strangers, gaze plays a particularly important role in cueing role exchanges (Beattie, 1981).

3.5.1. Gaze and Discourse Structure

Research on conversational functions of gaze showed that gaze behavior is closely linked with speech (Argyle and Cook, 1976). Kendon (1967) identified patterns in speakers' and addressees' gaze during role exchanges. For instance, he found that speakers mostly look away from their addressees at the beginning of an utterance, but look at their addressees at the end of an utterance. As parties switch roles at the beginning of the next utterance, the new speaker looks away from the new addressees (Kendon, 1967). In this situation, looking away at the beginning of an utterance and during hesitant speech indicates holding the floor (Nielsen, 1962; Kendon, 1967) and serves to avoid information overload in the planning of the utterance (Goodwin, 1981). Looking at the addressee at the end of an utterance, on the other hand, communicates that the speaker is ready to pass the floor to the addressee (Nielsen, 1962).

Information structure of speakers' utterances are also found to account for gaze shifts (Cassell et al., 1999). When utterances are looked at as theme-rheme progressions (Halliday, 1967), at the beginning of each theme, speakers look away from the addressees 70% of the time, and at the theme-rheme junction, they look at their addressees 73% of the time (Cassell et al., 1999).

3.5.2. Gaze Behavior as an Intimacy-Regulation Mechanism

In conversations, gaze direction can be used to regulate interpersonal intimacy (Argyle and Dean, 1965). The "Intimacy Equilibrium Theory" suggests that participants in a conversation develop an equilibrium for interpersonal intimacy, which is a function of gaze, physical proximity, intimacy of the conversational content, amount of smiling, etc. (Argyle and Dean, 1965). For any set of participants, the level of intimacy is at a certain degree and participants try to keep this degree constant over the course of the interaction. When one of the components in the model changes (e.g. an increase in physical proximity), people tend to maintain the equilibrium by shifting one or more of the other components in the reverse direction.

Argyle and Dean (1965) experimentally demonstrated that gaze at partners were reduced at closer distances. Kendon (1967) found that smiling and the amount of gaze were inversely correlated, confirming Argyle and Dean's theory. Exline and others (1965) showed that intimacy of the conversational content affected the amount of gaze. Their results were partly consistent with the Argyle and Dean's theory; participants looked at their partners more when they were speaking, but not while listening. One explanation of this effect is that speaking about intimate topics evokes more embarrassment than listening to others talk about intimate topics (Argyle and Cook, 1976). On the other hand, Abele (1986) manipulated the amount of intimacy in a conversation and found that this did not affect how much participants looked at their partner. Although, she found that when partners talked about intimate topics, the level of gaze increased over the course of the conversation, while it decreased when participants talked about non-intimate topics. These results suggest that people do regulate their intimacy with their partners when they talk about intimate topics. However, people might also adopt to the level of intimacy in the course of a conversation.

3.5.3. Gaze Cues and Conversations with Embodied Virtual Agents

The use of nonverbal cues for conversational participation has been extensively studied in the design of embodied conversational agents (Cassell et al., 1994, 1999a, 1999b, 2002; Garau et al., 2002; Thorisson, 2002; Lee et al., 2002; Pelachaud et al., 2002; Rehm and Andre, 2005; Heylen et al., 2005). The largest contribution to this literature is from Cassell and her colleagues who developed a number of systems that use verbal and nonverbal behaviors to regulate conversational participation. One of these systems simulated conversations among multiple human-like agents that showed appropriate and synchronized speech, intonation, and non-verbal behavior (Cassell et al., 1994). In a text-based chat system called BodyChat, Vilhjalmsson and Cassell (1998) allowed users to control the nonverbal behaviors of an on-screen avatar to initiate and maintain conversations. Another example of these systems generated appropriate verbal and nonverbal signals to manage conversational mechanisms such as turn-taking, feedback, and repair in two-party conversations (Cassell et al, 1999a).

While these systems combined nonverbal cues such as gaze, facial expressions, hand gestures, and postural shifts in the design of the agent (Cassell et al., 1999a, 2001), gaze cues were considered as the most salient signal to establish conversational roles and regulate turn-taking (Cassell et al., 1999b; Vertegaal et al., 2001). Furthermore, signals that are designed to resemble human gaze behavior (as opposed to randomly generated signals) lead to more efficient conversations, better task performance, and more positive evaluations of the agent (Garau et al., 2001; Heylen et al., 2005). Similarly, Colburn and others (2000) asked participants to sit across from a computer screen and talk to a confederate through an avatar displaying human-like gaze patterns, the same avatar with fixed gaze, and the confederate's voice without the avatar's face on the screen. They found that participants gazed significantly more at the screen when the avatar was present and marginally more when the avatar displayed human-like gaze behavior. Similarly, Garau and others (2001) found that participants paid more attention to the avatar with humanlike gaze than they did to the avatar with random gaze. However, in contrast to Colburn and his colleagues' results, they also found that simply having an avatar face did not increase people's evaluations of their partners over talking over audio. Lee and others (2002) developed an animated face with human-like eye movements and asked participants to evaluate the face with no eye movement, randomly generated movement, and movement generated using data collected from human speaker and listeners. They found that people rated faces with human-like eye movements to be more interested, engaged, lively and friendly than the other two kinds of gaze behavior.

Another body of work on the use of gaze cues in conversational agents focuses on understanding how these cues affect people's interactions with virtual agents (Bailenson et al., 2001; Rehm and Andre, 2005). Bailenson and his colleagues (2001) asked participants to interact with a virtual character in an immersive virtual environment to test the intimacy equilibrium theory and found that the theory held for female participants: Females maintained larger interpersonal distance with agents when the agents engaged in eye-contact with them. Rehm and Andre (2005) asked two participants to play a game with Greta, a virtual character developed by Pelachaud and her colleagues (2002), where each player played the roles of the speaker and the addressee and evaluated people's involvement in the conversation. They found that when participants played the role of the speaker, they showed gaze behavior similar to human conversations. On the other hand, when the agent played the role of the speaker, people looked at the agent significantly longer than they looked at human speakers. Their results might imply that people do not regulate their intimacy with agents. However, that the task used in their experiment required participants to read deceptive cues in the agent's verbal and nonverbal behaviors might have caused longer gaze at the agent.

3.5.4. Gaze Cues and Conversations with Humanlike Robots

In human-robot interaction, a more recent but growing body of literature looks at social gaze behavior (Imai et al., 2002; Kanda et al., 2003; Sidner et al., 2004; Mutlu et al., 2006; Yoshikawa et al., 2006; Kuno et al., 2007; Staudte and Crocker, 2008; Trafton et al., 2008; Yamazaki et al., 2008). Among these, a few promising studies looked at the conversational effectiveness of robot gaze, particularly in regulating turn-taking in two-party (Kuno et al., 2007; Yamazaki et al., 2008) and multi-party conversations (Matsusaka et al., 2001; Bennewitz et al., 2005; Trafton et al., 2008).

Kuno and others (2007) developed a museum guide robot that looked at its addressee at turn-relevant places (Sacks et al., 1974) to regulate turn-taking. Yamazaki and others (2008) showed in an experiment that looking at the participants at turn-relevant places evoked more backchannel responses than looking at random places. A robot developed by Matsusaka and his colleagues (2001) could participate in multi-party conversations following the turn-taking model suggested by Sacks and others (1974) for human conversations. Bennewitz and others (2005) developed a similar robot that used a turn-taking model developed by Thorisson (2002) for virtual agents to take part in multi-party conversations. Trafton and others (2008) developed a robot that can listen to two-party conversations as a bystander and experimentally showed that people found the robot's gaze behavior more natural when the robot looked at the speaker only at turns as opposed to at turns and backchannel responses.

While these studies provide strong evidence that gaze cues from a robot support conversational functions such as turn-taking and showing appropriate listening behavior, whether these cues might shape different forms of conversational participation and affect people's perceptions of and interactions with the robot remains unknown.

In the next subsection, literature on the use of gaze cues in establishing joint attention is reviewed from social and developmental perspectives. A review of related work on the design of embodied virtual agents and humanlike robots is also provided.

3.6. Gaze Cues in Establishing Joint Attention

Another important aspect of social gaze behavior is the propensity to orient attention to the same source of information that other people are looking at (Frischen et al., 2007). This behavior is called "joint attention," also called "deictic gaze" or "visual co-orientation" (Butterworth, 1991). Joint attention has been studied from two perspectives. The first one is concerned with the perception of another person's gaze direction as it relates to social interaction while the second one looks at the development of the ability to follow another person's gaze direction and its significance to development of knowledge. A brief review of research from each area is provided below.

3.6.1. Joint Attention in Social Interaction

People tend to follow others' gaze direction because "the direction of a person's gaze usually indicates what object he is interested in or what person he is responding to in the sphere of the environment" (Gibson and Pick, 1963). Milgram and his colleagues (1969) showed the social nature of this seemingly automatic response in a field experiment where they asked confederates to gaze at the sixth floor window of an apartment in Manhattan and counted the number of passersby who followed the gaze direction of the confederates. They found that 20% of the people looked up when they saw one confederate looking up while 80% of the passersby looked up when they saw five confederates gazing at the window.

Gaze behavior in joint attention situations shows significant differences from conversational gaze. For instance, an observational study of sales staff and customers at a department store found that people looked at each other very little (Argyle, 1972). While this behavior would be considered to be asocial in a conversation, it is acceptable in joint attention situations as parties are expected to direct their attention at the object of interest. Argyle and Graham (1976) looked at the factors that affect how much people look at objects vs. a partner and found that the amount of gaze at objects increased and gaze at the partner decreased as the complexity of the object of attention and its relevance to the interaction increased.

3.6.2. Joint Attention as a Developmental Paradigm

What makes joint attention interesting from a developmental perspective is that the ability to follow another person's gaze direction develops before language (Scaife and Bruner, 1975; Mundy and Newell, 2005) and plays an important role in social development (Moore and Dunham, 1995), particularly in language acquisition (Baldwin, 1995; Morales et al., 2000; Brooks and Meltzoff, 2005) and in social cognition (Tomasello, 1995; Carpenter et al., 2000; Charman et al., 2001). Infants as young as 3 months can follow gaze shifts presented on a computer screen (Hood et al., 1998). Morales and his colleagues (2000) found that infants' ability to follow a caretaker's gaze shifts at six months is correlated with their vocabulary size at 18 months. Baldwin (1995) argues that the ability to take a caregiver's perspective might improve the learning of nouns through making associations between an observed object and its name. Charman and his colleagues (2001) found that joint-attention ability at 20 months was a predictor of theory-of-mind ability at 44 months showing the importance of joint attention in the development of social cognition.

3.6.3. Gaze Cues and Joint Attention with Humanlike Virtual Agents

Johnson and Rickel (1997) developed a pedagogical agent named Steve that used gaze and pointing to direct students' attention to important information in a virtual environment. Another pedagogical agent, Cosmo, developed by Lester and his colleagues (1999), used gaze cues to make deictic references to representations of computers in a network in a virtual world. Cassell and her colleagues (2002) built an embodied conversational agent called MACK with the ability to build joint attention with people over a physical map located between people and the screen on which the agent was projected. Their system used an empirical model of nonverbal grounding based on gaze cues to establish joint attention over the map. They evaluated the communicative effectiveness of their model in an experiment where they manipulated whether or not MACK used gaze cues to establish joint attention and found that when the agent used gaze cues, people's interactions with the agent showed gaze routines that were similar to those observed in interactions between people.

3.6.4. Gaze Cues and Joint Attention with Humanlike Robots

Several researchers have built robotic systems with the ability to establish joint attention with people (Brooks et al., 1999; Scassellati, 1999; Nagai et al., 2003; Kozima et al., 2004; Calinon and Billiard, 2006). Brooks and his colleagues (1999) developed a robot called Cog with joint attention abilities. These abilities were developed in different modules as suggested by Baron-Cohen (1995) including eye contact, gaze following, and pointing with gaze (Scassellati, 1999). Kozima and his colleagues (2004) developed two robots, Keepon and Infanoid, with the ability to establish joint attention with children and conducted longitudinal studies of how children interacted with the robots. They showed that two-year-olds and over not only recognized the direction of attention of the robot but also initiated joint attention through pointing at or holding objects (Kozima et al., 2003). Nagai and her colleagues (2002) developed a model for Infanoid that allowed the robot to learn joint attention abilities from a human caregiver. Imai and his colleagues (2003) developed a joint attention mechanism for a humanoid robot that used verbal deictic references in combination with gaze and pointing gestures to direct people's attention towards objects in the environment. They also conducted an experiment where they assessed the use of eye contact in establishing joint attention and found that the robot could only direct people's attention towards objects when it established eye contact with them. As a follow-up to this work, Sugiyama and his colleagues (2006) developed a model of joint attention for a humanoid robot that used gaze cues and verbal and gestural deictic references to direct people's attention at objects in the environment and detected when people used these cues and references to orient the robot's direction of attention towards the objects that people were referring to. They also evaluated their model in an experiment where subjects were asked to point at objects and evaluate whether the robot could correctly direct its attention to the object. Their results showed that the robot could direct its attention to the object that participants referred to with 92% accuracy.

The last two studies (i.e. Imai et al., 2003; Sugiyama et al., 2006) show that cues from a robot's gaze is instrumental in establishing joint attention. However, whether these cues can lead to significant social outcomes such as stronger associations between the verbal content and the object of mutual attention or better learning of foreign vocabulary remains unexplored.

Next section provides a brief outline of interpersonal differences in gaze behavior with a particular focus on cultural and gender-based differences.

3.7. Gaze Cues and Interpersonal Differences

Gaze behavior is found to be extremely sensitive to individual differences (Argyle and Cook, 1976), particularly gender-based (Hall, 1984) and cultural differences (Watson, 1970). Below, a review of these differences is provided.

3.7.1. Gender-based Differences

In all measures of gaze, women are found to look more than men (Argyle and Cook, 1976; Francis, 1979; Hall, 1984). Argyle and Ingham (1972) found that parties in female dyads looked more (66% vs. 56%) and longer (3.12 vs. 2.45 seconds) at each other than parties in male dyads. They also established more (38% vs. 23%) and longer (1.42 vs. 0.86 seconds) mutual gaze than parties in male dyads. Similarly, parties in female triads were found to look more (37% vs. 23%) and establish more mutual gaze (8% vs. 3%) than parties in male triads do (Exline, 1963). In unfocused interactions, females are looked at more than males are (Coutts and Schneider, 1975). Patterson and his colleagues (2002) found passersby in public look more at female confederates (47%) than at male confederates (30%).

Gender-based differences in gaze behavior appear as early as infancy. Girls at 12 and 24 months build more eye contact with their mothers than boys at that age do (Lutchmaya et al., 2002). Similarly, 12-month-old girls show stronger joint attention abilities than boys at that age do (Olafsen et al., 2006). Furthermore, Kagan and Lewis (1965) found that, at six months, girls attend to faces more than boys do. At birth, boys look longer than girls at a mobile, while girls at this age look longer than boys at a face (Connellan et al., 2000). Similarly, Lutchmaya and her colleagues (2002) showed 12-month-old infants videos of moving cars and faces and found that attention in males was more drawn by moving cars and in females by moving faces. All three studies suggest that men and women respond differently to social and non-social stimuli.

3.7.2. Cultural Differences

People from certain cultures are found to look more than others. Watson (1970) found differences in how much people from "contact" cultures (i.e. Arabs, Latin Americans, South Europeans) and those from "non-contact" cultures (i.e. Asians, Indians-Pakistanis, Northern Europeans) look at their partners and found that the former looked at others more than the latter. Japanese were found to look more frequently but with shorter glances than Australians (Elzinga, 1978). Ingham (1972, as described in Argyle and Cook, 1976) compared Swedes and Englishmen and found that the former looked less often (8 vs. 13 glances per minute) but with longer glances (5 vs. 2.93 seconds).

A consideration of cultural differences in gaze behavior is particularly important in the context of this dissertation as the first empirical study was conducted in the United States with native American-English-speaking subjects and the second study was and third study will be conducted in Japan with native Japanese-speaking subjects.

3.7.3. Individual Differences

While sex and cultural differences account for some of the variability in gaze behavior, greater differences are observed among individuals (Nielsen, 1962; Kendon, 1967, 1978; Ellsworth & Ludwig, 1972). Nielsen (1964) found that the amount of time spent looking at a confederate varied between 8% and 73% with an average of 50%. Argyle and Cook (1976) argue that such large variability might result from differences in need for affiliation (which accounts for some of the sex differences), personality type, and varying levels of tolerance for arousal and intimacy.

This section provided background on social gaze behavior from research on human communication, and related work on the design of social gaze behavior for embodied conversational agents and humanlike robots focusing specifically on the social contexts described in the scenarios presented in Section 2. The next section describes in detail a series of empirical studies that build on these scenarios to explore some of the important aspects of robot gaze that have not been looked at by existing work.

4. Empirical Studies

The scenarios described in the Research Context and Approach section (i.e. Section 2) provide social context and motivation for three empirical studies. Each study involves a human-robot interaction situation where aspects of robot gaze play a central role. The robot's gaze behaviors are created using human communication theory and empirical data collected from human subjects. Designed behaviors are algorithmically implemented on robots and evaluated in controlled experiments where human subjects are asked to perform in a task with the robot. Manipulations are introduced to robot's gaze behavior to understand how elements of the designed behavior affect people's interactions with the robot. This section provides detailed descriptions of these stages for the two completed studies and planned stages for the proposed third study.

4.1. **Study 1:** The Design of Gaze Cues for Communication of Attention

The scenario that motivated the first study posed the research questions: Can robot gaze communicate attention in way that it affects learning? How can we design robot gaze behavior to attract attention and improve learning? What are the design variables? Below, the three stages of the study that addressed these questions are described in detail.

4.1.1. Theoretically and Empirically Grounded Design

To provide a social context for the study, a storytelling experience was designed where ASIMO (Figure 2), a humanoid robot developed by Honda (Sakagami et al., 2002), told a Japanese fairy tale, "The Tongue-Cut Sparrow" (Ozaki, 1970) to two listeners using a pre-recorded voice. ASIMO's gaze behavior was designed based on a human-like gaze model that used existing theory and empirical data collected from a human storyteller. Arm and body gestures were added to enrich the naturalness of ASIMO's behavior.

4.1.1.1. Theoretical Grounding

Theoretical grounding is established by extending findings by Cassell and her colleagues (1999). They developed an empirical model of gaze behavior during turn-taking and within a turn based on the structure of the information conveyed by the speaker (Cassell et al., 1999). Because the task involved storytelling, which follows the structure of an

Figure 2. Honda's ASIMO (Sakagami et al., 2002).

oratory instead of a conversation and does not involve turn-taking, the model was used to determine gaze shifts within the utterances of robot's speech. Their model follows the English sentence structure suggested by Halliday (1967), who describes the two main structural components of an utterance using the terms "theme" and "rheme." The theme refers to the part of an utterance that sets the tone of the utterance and connects the previous utterance to the next one. The rheme contains the new information that the utterance intends to communicate. For instance, in the sentence "In the evening the old man came home." "In the evening the old man" is the theme while "came home" is the rheme of the utterance. In their model, speakers look away from their listeners at the beginning of a theme with 0.70 probability and look at their listeners at the beginning of a rheme with 0.73 probability. They suggested the following algorithm to simulate natural gaze behavior using a randomized function, *distribution(x)*, that returns true with probability *x*.

```
for each proposition do
    if proposition is theme then
        if beginning of turn or distribution(0.70) then
            attach a look-away from the listener
        end if
    else if proposition is rheme then
        if end of turn or distribution(0.73) then
            attach a look-toward the listener
        end if
    end if
end for
```

4.1.1.2. Empirical Grounding

Empirical data collected from a professional storyteller was used to determine locations and frequencies for the algorithm proposed by Cassell and her colleagues (1999). The professional storyteller was videotaped relating two stories to a two person audience. Figure 3 illustrates the spatial configuration of the storyteller, her audience, and data collection equipment. 30 minutes of video data was used to analyze where in the environment and for how long each gaze shift executed by the storyteller was directed.

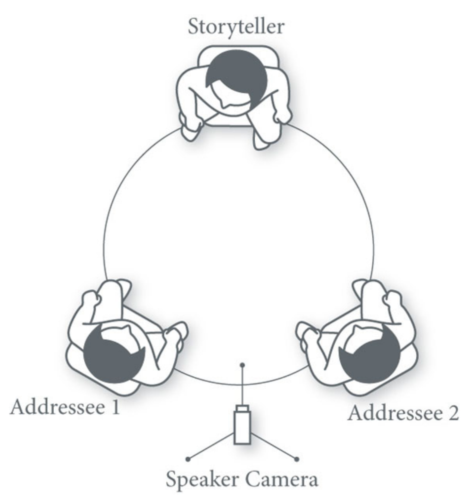


Figure 3. Spatial configuration of the data collection.

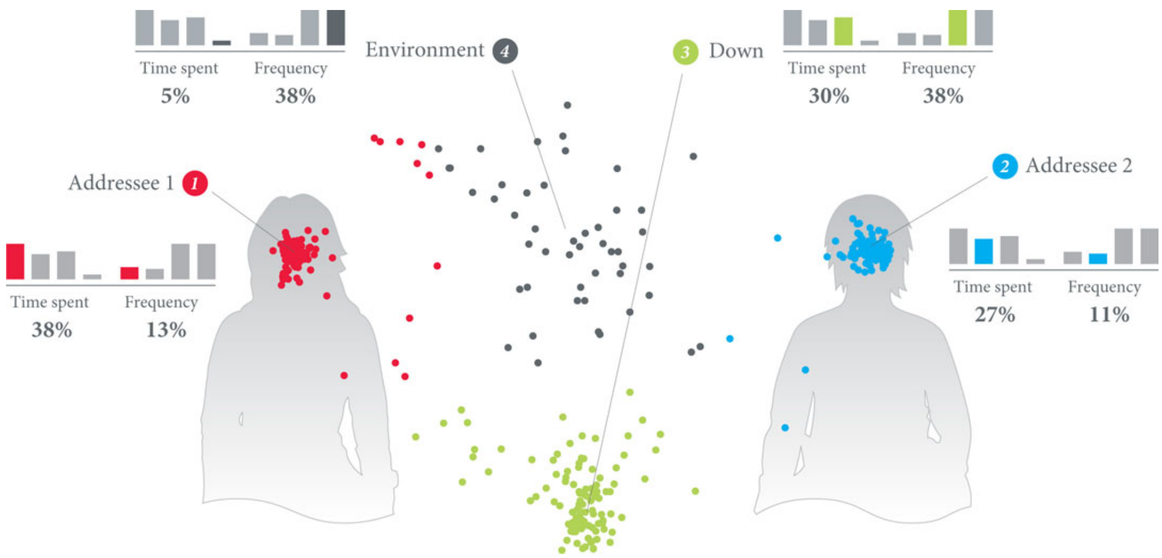


Figure 4. Clusters of speaker's gaze targets and time spent and frequency of gaze at each cluster.

The results showed that the storyteller gazed at four different kinds of locations: the two members of the audience, a fixed spot on the table in front of her, and a set of spots in the environment. Figure 4 shows a k-means clustering of these four locations.

"Looking at" was defined as keeping ASIMO's gaze on one listener once it was fixated there. "Looking away" meant looking at the other listener or looking at a spot in the environment or the fixed location. When the gaze was not currently directed at a listener, "looking at" meant looking at one of the listeners, while "looking away" meant looking at any four of the targets with predetermined probabilities. These probabilities were derived from an analysis of the frequencies of the storyteller's gaze at each location. The duration of the gaze at each location was represented by a normal distribution, which was used to determine the length of the simulated gaze. Table 1 shows these values for each gaze location.

	Listener 1	*Listener 2*	*Fixed spot*	*Environment*
Frequency (%)	13	11	38	38
Time spent (%)	38	27	30	5
Min time (ms)	477	484	242	360
Max time (ms)	15,324	5,914	13,674	4,383
Mean time (ms)	2,400	2,262	2,640	1,072

Table 1. Amount, frequency, and length of gaze at each location.

4.1.2. Implementation

This gaze model was used with a hand-coded script of the information structure of the fairy tale to simulate human-like gaze behavior. The script marked the start of each theme and rheme and pauses between utterances. Below is the pseudo-code for an algorithm developed by extending the algorithm proposed by Cassell and her colleagues. In the designed algorithm, *distribution(x)* produces a uniform randomized function that returns true with the probability derived from Cassell and her colleagues' (1999) algorithm (e.g. 0.70) and from the empirical data. For example, *probability(environment)* is 38% from Table 1. Function *length(x)* generates a duration for the gaze over a normal distribution with mean and standard deviation values from the empirical results (~Normal(Mean(x), StDev(x))).

```
for each part of the utterance (theme/rheme/pause) do
    while the duration of the part do
        if current part is pause then
            if distribution(probability(environment)) then
                gaze at environment with length(environment)
            else
                gaze at fixed spot with length(fixed spot)
            end if
        else if current part is theme then
            if distribution(0.70) then
                if distribution(probability(environment) then
                    gaze at environment with length(environment)
                else
                    gaze at fixed spot with length(fixed spot)
                end if
            else
                if distribution(probability(listener 1)) then
                    gaze at listener 1 with length(listener 1)
                else
                    gaze at listener 2 with length(listener 2)
                end if
            end if
        else if current part is rheme then
            if distribution(0.73) then
                if distribution(probability(listener 1)) then
                    gaze at listener 1 with length(listener 1)
                else
                    gaze at listener 2 with length(listener 2)
                end if
            else
                if distribution(probability (environment)) then
                    gaze at environment with length(environment)
                else
                    gaze at fixed spot with length(fixed spot)
                end if
            end if
        end if
    end while
end for
```

The gaze algorithm was implemented on ASIMO by using a hand-coded script of the story and synchronizing ASIMO's gaze behavior with a pre-recorded voice. Ten simple arm gestures were automatically added for long utterances (greater than the mean length of 2,400 ms for gaze at a listener). Six special gestures such as bowing, crying or acting angry were added by hand when they were semantically appropriate. The location of the participants was not sensed but was determined by placing two chairs at known locations and programming ASIMO to look in those two directions. The initiation of the robot's movement was controlled by the experimenter. The robot then introduced itself to the participants, told the story, and ended the interaction.

4.1.3. Experimental Evaluation

Drawing from existing theory, two hypotheses were formulated about responses to a manipulation in the amount of gaze behavior:

Hypothesis 1. - Participants who are looked at more will perform better in the recall task than participants who are looked at less.

Hypothesis 2. - Participants who are looked at more will evaluate the robot more positively than participants who are looked at less.

To test these hypotheses, a between-subjects experiment was conducted where participants listened to ASIMO while it told a Japanese fairy tale in English. ASIMO's gaze behavior was manipulated to gaze at one of the participants with 20% frequency and the other participant with 80% frequency. Participants were placed at the same distance from ASIMO and space was left between them so that they would not interact with each other and the robot's gaze at each participant would be easily distinguishable (Figure 5).

4.1.3.1. Experiment Procedure

Participants were first given a brief description of the experiment procedure. After the introduction, participants were asked to answer a pre-experiment questionnaire and then provided with more detail on the task. ASIMO then introduced himself and performed the storytelling task. After listening to ASIMO's story, participants performed a distractor task, where they listened to another story on tape ("The Flying Trunk" by Hans Christian Andersen, 2001). Before listening to either story, they were told that they would be asked

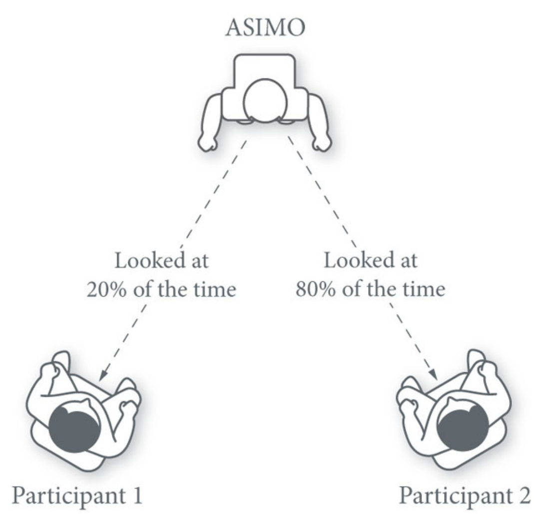


Figure 5. Spatial configuration of the robot and participants.

questions regarding one of the stories. All participants were asked questions regarding ASIMO's story. After completing the task, participants answered a post-experiment questionnaire regarding their affective state, their perceptions of the robot, and their demographic information. ASIMO's story, the story on tape, and the whole experiment took an average of 17.5 minutes, 7.5 minutes, and 35 minutes respectively. The experiment was run in a dedicated space with no outside distraction. A male and a female experimenter were present in the room during the experiment. All participants were paid $10 for their participation.

4.1.3.2. Measures and Participation

All factors in the experiment were identical for each participant except for two controlled factors: the frequency of the robot's gaze at each participant (a manipulated independent variable) and the participant's gender (a measured independent variable). The dependent variables measured were task performance, the participant's own affective state, their positive evaluation of the robot, their perceptions of the robot's physical, social, and intellectual characteristics, and their involvement in and enjoyment of the task. The post-experiment questionnaire included a question as a manipulation check, "How much did the robot look at you?" Seven-point Likert scales were used for all scales.

Twenty (12 males, 8 females) undergraduate and graduate students from Carnegie Mellon University participated in the experiment. Figure 6 shows participants in the experiment. Ten participants were assigned to the "looked at 80% of the time" condition. The other ten participants were assigned to the "looked at 20% of the time" condition. All participants were native English speakers and their ages ranged from 19 to 33. Participants were chosen to have a variety of majors including management sciences, social sciences, art, and engineering. Four male and three female participants had technical majors such as computer science, electrical engineering and information systems, while eight males and five females came from non-technical fields including english, business/management, writing, and psychology. On average, male participants had more video gaming experience and more familiarity with robots than female participants did.

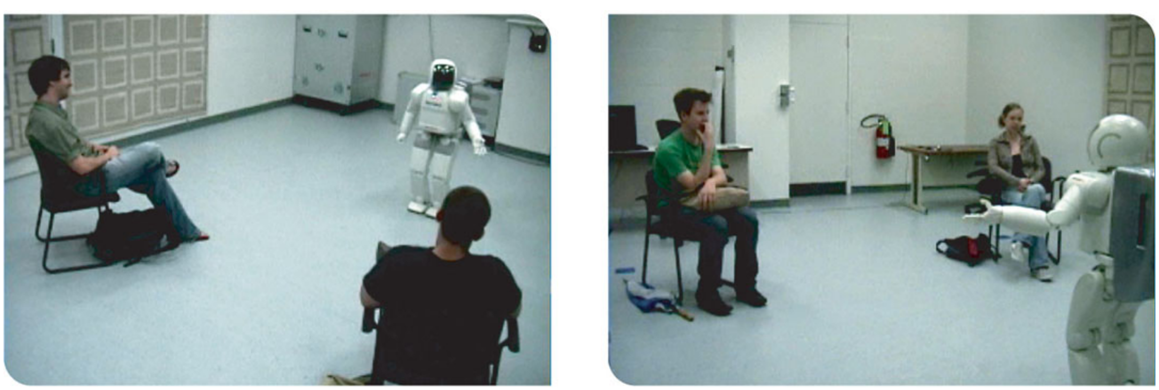


Figure 6. Participants in the experiment.

4.1.4. Results

Three methods were used in the data analysis; repeated measures analysis of variance (MANOVA), regression (Least Squares Estimation), and multivariate correlations. The first method applied an Omnibus F-Test to see if the difference between pre-experiment and post-experiment measurements was significant across the two conditions and/or genders. The second technique used a linear regression on the variables that were significant across conditions to identify the direction of main effects and interactions. The last method looked at how these variables correlated with each other. Reliability tests and factor analyses were also ran on the scales that were used for measurement.

Item reliabilities for all partner (robot), task, and self evaluation scales except the mutual liking scale (alpha=0.54) were high. However, because our scales for partner evaluation were created to evaluate human-like interface agents, we ran a factor analysis of all the items that we used for evaluating the robot and created a highly reliable (alpha=0.91), 8-item scale for positive evaluation. An analysis of the manipulation check showed that the participants were aware that they were looked at more or less by the robot ($F[1:18]=4.29$, $p=0.05$).

Consistent with the first hypothesis, a regression on the performance measure showed that participants who were looked at more performed significantly better in the recall task (answering questions regarding ASIMO's story) than those who were looked at less ($F[1:16]=5.15$, $p=0.03$). When participant's gender was included in the statistical model, the effect was significant only in females ($F[1:16]=8.58$, $p<0.01$) while men did not show any significant difference across conditions ($F[1:16]=0$, $p=1$) (Figure 7).

The analysis of the ratings of the positive evaluation scale showed no significant main effect but a significant interaction of experimental condition and participant gender (Figure 8). Men rated ASIMO more positively when they were looked at more while women's evaluations were higher when they were looked at less ($F[1:16]=5.62$, $p=0.03$). Although this result reveals significant interactions with participant's gender, it is not consistent with the prediction in the second hypothesis. Analysis of scales of participant's affect, task enjoyment, and task involvement did not show any significant effects or interactions.

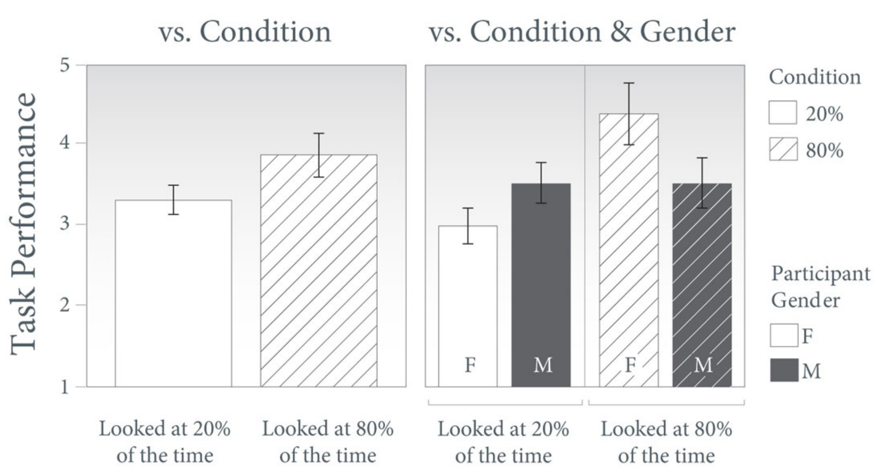


Figure 7. Task performance across conditions and condition & participant gender.

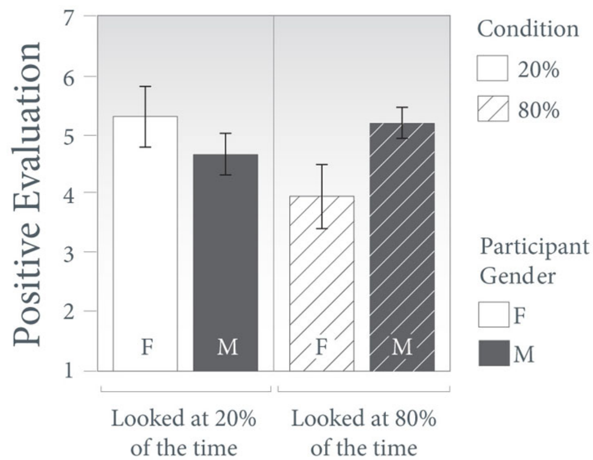


Figure 8. Positive evaluation (liking) across condition & participant gender.

The last analysis looked at how the scales correlated with participant's computer use, their familiarity with robots, and video gaming experience. A multivariate analysis using Pearson's correlation coefficient showed that ratings of the positive evaluation scale was highly correlated with video gaming experience ($r=0.65$, $p<0.01$), while not correlated with computer use or familiarity with robots. This correlation held for both genders although it was stronger in men. Video gaming experience was also correlated with task enjoyment ($r=0.53$, $p=0.02$).

4.1.5. Discussion

The results supported the first hypothesis: the frequency of the robot's gaze affected performance on the recall task. This result suggests that Melanie (the teacher in the first scenario) can instruct ASIMO to direct Justin's (the distracted student in the scenario) attention to itself and its story, which will lead to better recall of the story and eventually improved learning.

The second hypothesis, that participants who are looked at more will evaluate the robot more positively, was not supported. When gender was included as a variable in the analysis, women appeared to like the robot more when they were looked at less. This result can be explained by the differences in men's and women's perceptions of personal space based on the amount of mutual gaze established with a partner (Argyle and Dean, 1965; Hayduk, 1983). Bailenson and his colleagues (2001) showed that these differences appeared in people's interactions with virtual agents. They found that female participants maintained more interpersonal distance between themselves and agents who engaged them in eye contact than with agents who did not. Male participants did not show similar changes in behavior. This finding implies that because participants were not allowed to control their distance to the robot, females perhaps felt uncomfortable and evaluated the robot negatively when the robot gazed at them more. Lack of control over their distance with the robot did not affect men and they evaluated the robot more positively when it looked at them more.

In correlation analyses, positive evaluations of ASIMO were found to be highly correlated with participant's video gaming experience and not with their computer use, which suggests that people might perceive ASIMO as more like a video-game character or avatar than like a computer. This result suggests that people's interactions with robots and their interactions with embodied agents fall into the same ontological category of social responses to technology supporting the earlier argument that social responses to robots are evoked by humanlike cues instead of the automaticity of social behavior in the presence of minimal social cues.

4.1.5.1. Limitations

There are a number of limitations of the first study, some of which are addressed in the second and third studies. These limitations are discussed below. Limitations that are not addressed by the other two studies are discussed in the general discussion at the end of this document.

The design of the gaze behavior did not account for some elements of the professional storyteller's gaze. For example, she occasionally switched from looking at one listener to looking at the other listener during a theme or rheme, but the analysis did not identify a pattern in this behavior. The second study will address this limitation by comparing gaze shifts across situations with one and two listeners.

While this study looked at how cues from a robot's gaze affected the communication of attention, other non-verbal elements such as arm gestures and postural changes were used to make the experience as fluid and natural as possible. However, ASIMO's arm gestures were found distracting by some participants, perhaps because of the servo motors that generate noise while moving the robot's arms. Furthermore, the robot used arm gestures based on its orientation of attention. When it looked at the participant on the right, it gestured with its right arm meaning that the robot directed more gestures at the person to whom it looked at more. This behavior might have heightened participants' feelings of being attended to by the robot. To isolate this possible effect, arm gestures are eliminated in the next two studies.

Another limitation to the human-likeness of ASIMO's gaze model was due to the physical design of the robot. When humans direct their gaze, their movement combines movement of the eyes, the head, and the upper torso, whereas ASIMO only used head movement to shift its gaze because it's design does not include visible, controllable eyes and movement of the upper torso requires lifting and placing of the feet repeatedly, which were found to be time consuming and distracting in the pilot study. However, the results showed that this simple head movement was sufficient to create the experimental manipulation. Participants were asked to rate the amount of gaze they received from the robot. People who were looked at more thought the robot looked at them more ($M=56$, $SD=19$) and those who were looked at less thought ASIMO looked at them less ($M=38$, $SD=20$). The difference was marginally significant ($F[1:18]=4.29$, $p=0.05$). While the results suggest that head movement is sufficient to evoke the feeling of being looked at, robots with visible, controllable eyes are used in the second and third study to achieve more natural humanlike gaze behavior.

Some participants found ASIMO's story to be too long (17.5 minutes). While it is important for participants to immerse into the experience with sufficiently long interactions, the second and third studies involve relatively shorter tasks.

A post-hoc analysis of the data on storyteller's gaze behavior showed that the length of gaze shifts followed a positively skewed distribution instead of a normal distribution. This caused ASIMO's gaze shifts to be longer than the human storyteller's gaze shifts. This error is corrected in studies two and three by representing gaze lengths with two-parameter continuous distributions with values θ and k (Gamma).

This study looked at how cues from a robot's gaze can communicate attention and found that attention communicated through increased gaze led to better recall of the verbal content presented by the robot. The next study looks at how a robot's gaze cues can lead to different levels of participation in a conversation. In the next subsection, the design, implementation, and evaluation stages of the study are described in detail.

4.2. **Study 2:** The Design of Gaze Cues for Regulation of Conversational Participation

The scenario that motivated the second study posed the following research questions: Can simple cues from a robot's gaze lead to different forms of conversational participation? How can we design gaze behavior that leads to such outcome? And what are the design variables? Below, the three stages of the study that addressed these questions are described in detail.

4.2.1. Theoretically and Empirically Grounded Design

The goal of the second study is to gain a deeper understanding of how gaze cues are used in regulating conversational participation, use this understanding to create robot gaze, and study how human-robot conversations might be affected by the use of these regulative cues in a robot's gaze. While the literature on nonverbal behavior provides extensive knowledge on some of these cues (e.g. how much people look at and away from their addressees and speakers), there remains a number of other cues that might play an important role in regulating conversations (e.g. how much people look at non-participants, where people look when they look away from their partners, gaze durations at these locations, patterns of gaze shifts induced by the structure of spoken discourse, etc.). In order to gain a more extensive understanding of these cues, the first stage of this study involved creating conversational situations with different levels of participation and conducting a detailed analysis of speakers' gaze behavior. Different from the modeling in the first study, analyses were done for three kinds of gaze cues: (1) gaze shifts induced by information structure, (2) cues that signal conversational turn-taking, and (3) cues that signal participation structure. Theoretical knowledge from literature and analyses of the empirical data were combined to model both kinds of gaze behaviors.

4.2.1.1. Empirical Grounding Methodology

To study gaze cues that speakers use to regulate conversational participation, the following conversational scenarios with different configurations of participation were created (see Figures 9.a, 9.b, and 9.c for illustrations of the spatial configurations):

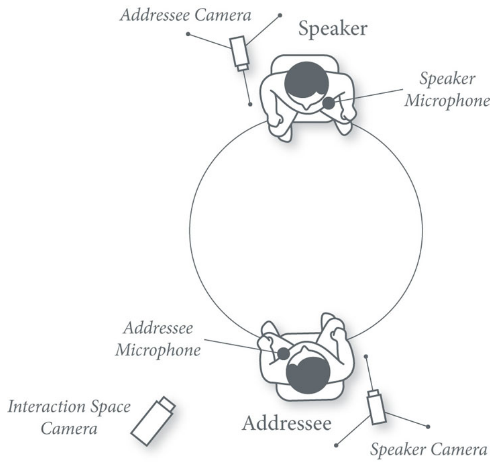


Figure 9.a. Spatial configuration and setup of the data collection equipment for the two-party conversation.

A two-party conversation with a speaker and an addressee:

> *Aiko (speaker) is a student at Osaka University and an active member of a student club. Takeo (addressee) is a new student at the same university. He is looking for a student club to join, therefore, attends the university club fair where he meets Aiko who is volunteering at her club's information booth. Aiko asks Takeo questions about his interests and provides him with information on club activities that might suit his interests.*

A two-party conversation with a speaker, an addressee, and a bystander:

> *Hiromi (speaker) is a resident of Shinsaibashi town in Osaka and attends Osaka University. He has lived in Osaka for a few years and is familiar with the town. Yoshi (addressee) is from Hokkaido and will be attending the same university. He is in town with his older brother Akira (bystander) to look for an apartment. A friend from high school connected him with Hiromi so that he can provide Yoshi with information on living in Osaka and local attractions that he might be interested in. As Hiromi tells Yoshi about the local attractions, Akira listens to their conversation.*

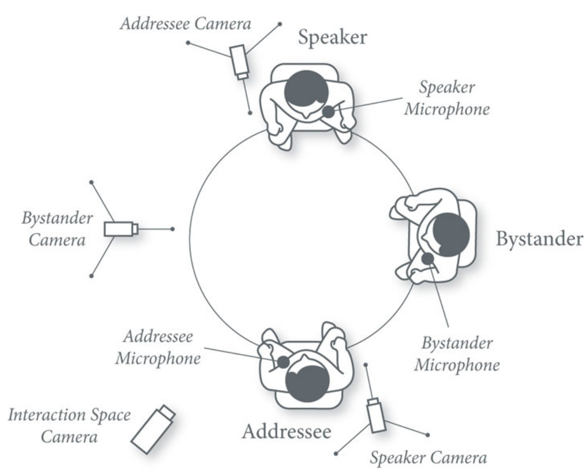


Figure 9.b. Spatial configuration and setup of the data collection equipment for the two-party conversation in the presence of a bystander.

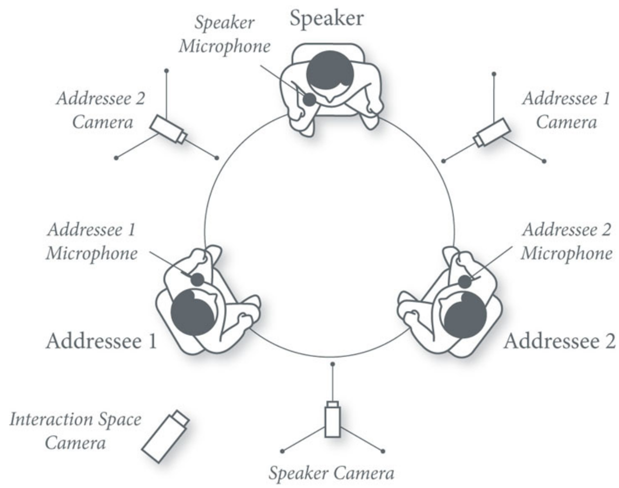


Figure 9.c. Spatial configuration and setup of the data collection equipment for the three-party conversation.

A three-party conversation with a speaker and two addressees.

> *Mika (speaker) is a student at Osaka University and works part-time at a local travel agency in Osaka. Toshi (addressee) and Jiro (addressee) are two friends attending Kinki University in Osaka. They plan to go on a trip together and are shopping for an affordable vacation package that they both find interesting. Mika inquires about their budget and shared interests and provides information about his company's travel packages.*

Participation

Defining the participant profile for the observation was particularly challenging. Research on nonverbal behavior reports strong effects of group composition on the production and perception of gaze, particularly of gender (Exline, 1963; Argyle and Dean, 1965; Exline et al., 1965; Argyle and Ingham, 1972), age (Efran, 1968; Libby, 1970), and personality traits such as extroversion (Mobbs, 1968). The first study in this dissertation also found strong gender effects on how robot's gaze affected people's performance and their perceptions of the robot. To control for these effects, the participant profile was limited to male university students from the Osaka area with little variance on age.

Four all-male triads (12 participants) were hired to perform in the three scenarios described above. The experimental procedure was as follows: Before their participation, all participants were asked to review and sign consent forms. Next, they were asked to provide demographic information and fill in a questionnaire that measured introversion-extroversion (Whitman, 1929). All triads performed in each scenario for fifteen minutes in the same order listed above. At the beginning of each scenario, an experimenter provided each participant with a description of the scenario and his role. Participants were given five minutes to ask questions and adapt to their roles. Between performing in the scenarios, they were asked to solve ten-minute-long crossword puzzles to distract them from their roles in the previous scenario. The roles of Aiko, Hiromi, and Mika (speakers), of Takeo, Yoshi, and Toshi (addressees), and of Akira and Jiro (bystander in scenario 2 and addressee in scenario 3) were performed by the same member of each triad. At the end of their participation, each participant was paid 3,000 ¥ (roughly $28 or €18).

Figure 10. Video data samples from the cameras that captured (a) the speaker, (b) addressee 1, (c) addressee 2, and (d) the interaction space.

Measurement

Gaze behavior was captured using high-definition cameras at 1080i resolution placed across from each participant's seat. Speech was captured using stereo microphones attached on each participant's collar. The cameras provided video sequences of each participant's face (from hair to chin). A fourth camera was placed on the ceiling to capture the interaction space (see image (d) in Figure 10). For both practical and ethical reasons, cameras remained visible to the participants. The resulting data was 45 minutes of video for each participant and 180 minutes of data for each triad. Data from one of the triads was used for the detailed analysis and 10% the data from the other triads was used to verify the findings from the detailed analysis. Only speaker gaze was analyzed in detail. Addressee gaze is only looked at to analyze turn-taking behavior. Interruptions and backchannel responses are omitted in the analysis.

4.2.1.2. Theoretical and Empirical Grounding

Existing theory was used in designing two main aspects of conversational gaze behavior. First, theory on conversational turn-taking informed the design of turn-yielding gaze signals. Second, in identifying reoccurring patterns in gaze shifts induced by information structure, unitization of discourse segments was informed by linguistic theory. The design of the gaze behavior was also informed by analyses of empirical data, particularly in identifying gaze targets, total time spent looking at each target, length distributions of glances at each target, turn-yielding and floor-holding signals, and patterns in gaze shifts among targets (e.g. the likelihood that a target is looked at next, given the current target). Below, a detailed description of the design process and how theoretical knowledge and empirical findings informed this process is provided.

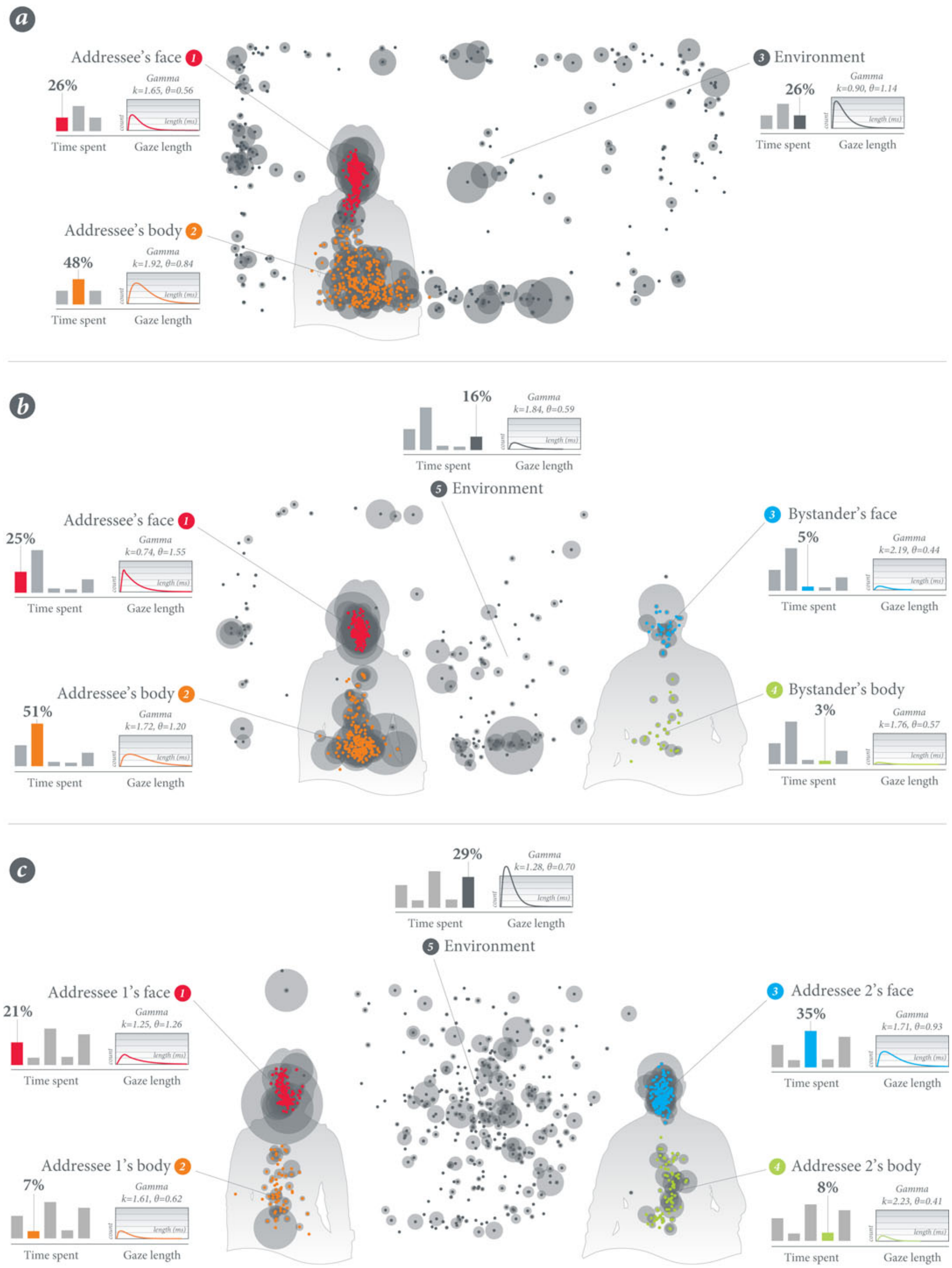


Figure 11.a, b, c. Gaze targets and target clusters for (a) the two-party conversation, (b) the two-party conversation with bystander, and (c) the three-party conversation. Points represent gaze targets and areas around points represent time spent at targets. Graphs illustrate the length distribution for and the amount of time spent looking at each cluster.

Gaze Targets

A frame-by-frame analysis of the speaker's gaze behavior was conducted. Each gaze shift was coded for its target and time of execution by a single coder. The targets are qualitatively estimated from speaker's gaze direction and marked on an image representation of the speaker's field of view. Coded locations were both qualitatively and quantitatively analyzed for creating clusters. In the qualitative analysis, three clusters in the first scenario and five clusters in the second and third scenarios were identified. A Gaussian mixture model estimation algorithm (Bouman, 1997) was used to quantitatively determine the number of clusters and identify the cluster to which each gaze target belonged. The algorithm confirmed the number of clusters in the qualitative analysis for the first and third scenarios. However, it produced eight clusters in the second scenario, four of which were grouped together to match the qualitative analysis. Figures 11.a, 11.b, and 11.c illustrate gaze locations, identified clusters, and length distribution for and time spent looking at each cluster for each conversational setup.

Target Clusters

The three clusters identified in the two-party scenario corresponded to listener's face, looking down at the area of listener's body, and looking away at spots in the environment, mainly aligned with or above eye level. The five clusters identified in the two-party-with-bystander scenario corresponded to listener's face, looking down at the the area of listener's body, bystander's face, down at the area of bystander's body, and looking away at spots in the environment, mainly aligned with or above eye level. The five clusters identified in the three-party scenario corresponded to the first listener's face, down at the area of the first listener's body, second listener's face, down at the area of second listener's body, and looking away at spots in the environment, mainly located in the area between the first and the second speaker. Figures 11.a, 11.b, and 11.c illustrate these clusters as they are mapped on the speaker's field of vision.

Gaze Amounts and Durations

The amount of looking at each target cluster was calculated as percentages. Parameters for the distribution of gaze durations for each cluster was obtained by fitting a two-parameter continuous Gamma distribution with values θ for shape and k for scale to the data for the cluster. Figures 11.a, 11.b, and 11.c provide the amount of time spent at and the distribution parameters for each cluster in the three scenarios.

Gaze Cues that Signal Turn-Taking

Several researchers have shown that gaze behavior is instrumental in turn management in conversation and follows a common pattern (Kendon, 1967; Duncan, 1972; Sacks et al., 1974; Goodwin, 1981). To identify how gaze cues were used for turn apportionment by the speaker, turn-relevant places were marked based on the systematics suggested by Sacks and his colleagues (1974). The analysis focused on turns that were initiated by an explicit turn-yielding signal of the speaker, therefore omitting interruptions and simultaneous turns (Duncan, 1974). The analysis showed that almost all of the turns that did not involve interruptions or simultaneous speech (a total of 8, 9, and 20 turns with explicit turn-yielding signals in the two-party conversation, two-party-with-bystander conversation, and three-party conversation scenarios respectively) followed the general pattern below that is also proposed by Kendon (1967) and Duncan (1974):

Turn-yielding: The speaker looks at the addressee at the end of a turn accompanied by an evaluative remark or question signaling that he is ready to pass the floor to his addressee (the next speaker).

Turn-taking: The addressee looks at his speaker at the end of the speaker's turn signaling that he is open to taking the floor. As floor is passed to the addressee (now, the current speaker), he looks away signaling that he is keeping the floor until his turn is complete.

Question-answer pairs: If the speaker asks a question to his addressee, he looks at his addressee (the next speaker) for the duration of his turn. When the answer is complete, he responds to the answer with a minimal response (McLaughlin and Cody, 1982) such as an acknowledgement, mirror response, or laughter, during which the speaker looks at his addressee. When he starts a new utterance, he looks away from his addressee. Figure 12 illustrates the speaker's turn-yielding, turn-taking, and floor-holding gaze signals during a question-answer pair.

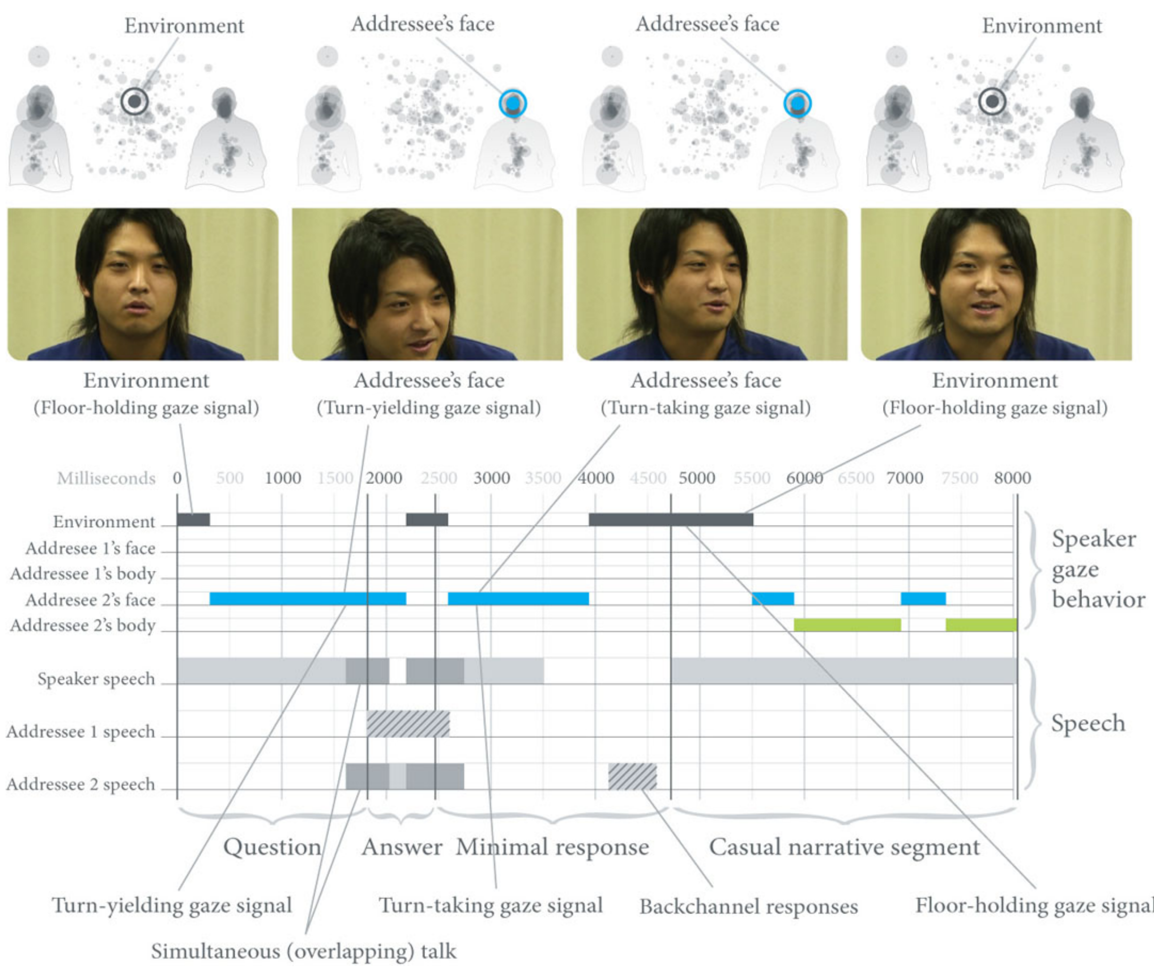


Figure 12. The speaker's turn-yielding, turn-taking, and floor-holding signals during a question-answer pair. Horizontal lines illustrate the timeline of an 8000-ms speech segment from data. The lower half of the timeline shows the speaker's and the addressees' speech and the upper half illustrates the speaker's gaze targets. The pictures show the speaker's turn-related gaze signals.

Gaze Shifts Induced by Information Structure

Unitization of Talk - As suggested by Cassell and her colleagues (1999), the structure of speaker's discourse accounts for a large part of gaze shifts within the course of a turn. Cassell and her colleagues (1999) unitized speech into utterances and identified theme-rheme progressions within each utterance. Because the data in this study showed the characteristics of a "casual narrative" (Tannen, 1984) where the speaker holds the floor for longer periods, a different unit of analysis was sought. Casual narratives are sequences of narrative segments where each segment performs as a pragmatically-functional speech act and causes a shift in the participants' points of view (Maynard, 1989). Hinds (1976) calls these segments "paragraphs" of a discourse, where each paragraph represents a distinct discourse topic and consists of sentences that are more closely related to each other than to other sentences in the discourse. Maynard (1989) describes a similar unit of analysis called "thematic fields" bounded by topic shifts marked by linguistic and interactional expressions. These expressions include:

- Substantial lapse, often filled with back-channel-like utterances,
  The speaker pauses as a result of hesitation that might signal a move from one focus to another or from one thought to another (Chafe, 1979).
- Formulation and evaluative comments,
  The speaker uses a part of the conversation to summarize, characterize, translate, or explain what is being said (Garfinkel and Sacks, 1970).
- Minimal responses,
  Responses such as acknowledgements, mirror responses, and laughter that do not contribute to the advancement of the topic and are often followed by pauses (McLaughlin and Cody, 1982). In Japanese, minimal responses that mark the transition of a thematic field are conclusive remarks that are pronounced with a finalizing tone such as "naruhodo ne," "I see" (Maynard, 1989).
- Sentence adverbs and conjunctions.
  Transitional adverbs and conjunctions that fill the gap between two themes to minimize disruptions in conversational flow such as "tokorode," "by the way" (Maynard, 1989).

The data was unitized by having a native Japanese speaker mark these expressions in the transcript. The following example provides an example segment of speech split into two thematic fields with the identified topic shift marker (in this case, a substantial lapse induced by hesitation).

> ウチはその、ホントあの、基本的にはオールラウンドサークル的な感じで、あの、スポーツなんかも結構やるし、▼ 月に*1*・・、月に*1*回は言いすぎかな。*2*ヶ月に*1*ぺんくらいは、フットサルの、結構、みんなでやる大会とか開いたりとか、あの、たまにソフトボールやったり・・みたいな感じで、結構体も動かすし。
>
> *Uchi wa sono honto ano kihonteki niwa o-ruraundosa-kuru tekina kanji de ano supotsu nankamo kekkou yarushi, ▼ tsuki ni 1.. tsuki ni 1 kai ha ii sugikana. 2 kagetsu ni 1 penkuraiwa futtosaru no kekkou minnade yaru taikai toka hirai taritoka ano tamani sofutobo-ru yattari.. Mitaina kanji de kekkou karada mo ugoka sushi.*

> *Our club, uh... basically does many activities, uh... for instance, we do sports activities often ▼ Once a month... once a month is maybe too much. Every two months, we do an indoor soccer tournament and sometimes we play softball.. You can see that we have a lot of chance to get some exercise.*

The unitization of data produced 181, 146, and 155 thematic fields in the two-party conversation, two-party-with-bystander conversation, and three-party conversation scenarios respectively. For each conversation, each thematic field was visually mapped onto the speech timeline along with gaze shifts that take place within the thematic field and 5000 ms before the beginning and after the end of the thematic field. Visualization software was developed to qualitatively identify the patterns of gaze shifts that occurred at the onset of each thematic field.

The analysis identified two main patterns of gaze shifts in the two-party conversation and the two-party-with-bystander conversation and another set of two patterns in the three-party conversation. Figure 13 provides an example of the most frequent pattern identified in the two-party conversation. Table 2 provides the frequencies for each pattern for each conversation. All frequent patterns in the two-party and three-party conversations are illustrated in Appendix 1.

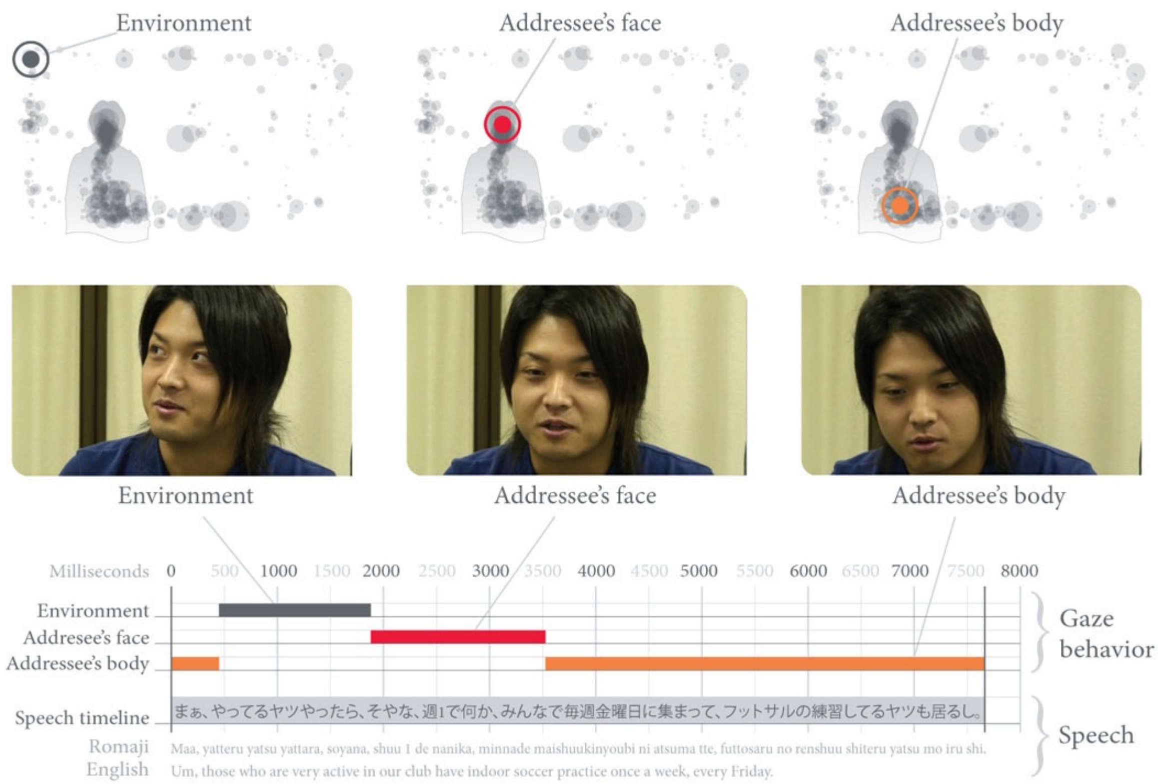


Figure 13. An example of the gaze pattern Environment > Face > Body in the two-party conversation. Other frequent patterns are illustrated in Appendix 1.

	Two-party conversation and two-party conversation with bystander	*Three-party conversation*
Look away > Look at > Look down	*25% at thematic field beginnings* *63% at turn beginnings*	*29% at thematic field beginnings* *7% at turn beginnings*
Look at > Look down > Look at	*30% at thematic field beginnings* *17% at turn beginnings*	*Not observed*
Look away > Look at > Look away	*Not observed*	*47% at thematic field beginnings* *60% at turn beginnings*
Pattern continuing from the previous thematic field	*22% at thematic field beginnings* *0% at turn beginnings*	*22% at thematic field beginnings* *0% at turn beginnings*
No recurring pattern	*22% at thematic field beginnings* *21% at turn beginnings*	*2% at thematic field beginnings* *33% at turn beginnings*

Table 2. Frequencies of the patterns identified in the two- and three-party conversations. Frequencies from two-party and two-party-with-bystander conversations are combined because similar patterns with similar frequencies were observed in these two conversations.

Gaze Cues that Signal Participation

Greetings serve to clarify and establish the roles that participants will take in a conversation (Goffman, 1955). An analysis of how the speaker established the roles of other participants during greetings showed that, in all three conversations, the speaker acknowledged all the participants (i.e. the addressees) and non-participants (i.e. the bystander) by following common conversational rituals such as exchanging names and ending the greeting with the expression "nice to meet you." During these exchanges, the speaker's gaze was directed at the person whom he was greeting. Turns during greetings often ended with an explicit head nod. The main differences in speaker's gaze behavior across the three conversational configurations appeared at the transitions from greetings to casual conversations. Below are descriptions of the speaker's gaze behavior at these transitions in the three conversations.

- In the **two-party conversation**, after the greeting, the speaker directed his attention the addressee, keeping his attention at the addressee at all times and gazing at the addressee at turns.
- In the **two-party conversation with bystander**, after the greeting, the speaker directed his attention towards the addressee, keeping his gaze at the addressee most of the time and glancing shortly at the direction of the bystander at different points of the conversation. While no patterns or triggers were identified in the glancing behavior, these glances never occurred at the end of speaker's turns because looking at the bystander at the end of a turn could be interpreted as a turn-yielding signal (Duncan, 1972; Sacks et al., 1974). Instead, the speaker gazed at the addressee at the end of turns signaling to the addressee that he could take the floor.

- In the **three-party conversation**, after the greeting, the speaker divided his attention between the two addressees, switching his gaze from one addressee to the other and waiting for one of the addressees to take the floor. Once the floor was taken, the conversation roughly followed the pattern of a sequence of two-party conversations. The speaker addressed and looked mostly at one of the addressees at a time and switched his focus when the other addressee interrupted with an attempt to take the floor or when his questions were directed at both addressees and were answered by the other addressee.

4.2.2. Algorithmic Implementation

The model developed in the design stage was used to create gaze behaviors for Robovie R-2 (Figure 14), a humanoid robot developed by ATR (Ishiguro et al., 2001). Because Robovie has controllable eyes, the gaze shifts were divided into eye and head movements with a 1:1 vertical ratio and a 4:1 horizontal ratio. These ratios were determined based on the robot's pan and tilt ranges, motor speeds, and smoothness of motion to optimize for speed of gaze shifts and naturalness of the behavior. When the robot looked at participants, each eye was given a 1.5-degree horizontal angle to converge the eyes at the participants who were seated at a two-meter distance.

4.2.2.1. Gaze Behavior for Turn-taking, Regulating Participation, and Information Structure

Robovie's gaze behavior was designed to adapt to the three configurations of conversational participation analyzed in the design stage. Its gaze behavior in two-party conversations involved acknowledging (i.e. looking at an addressee during greeting to signal that the addressee's presence is acknowledged), looking at, and producing turn-yielding signals for a single addressee. In two-party conversations with the presence of a bystander, Robovie acknowledged the presence of both the addressee and the bystander, but looked at the addressee for most of the conversation and produced turn-yielding signals only for the addressee. It also glanced at the bystander at random points in the

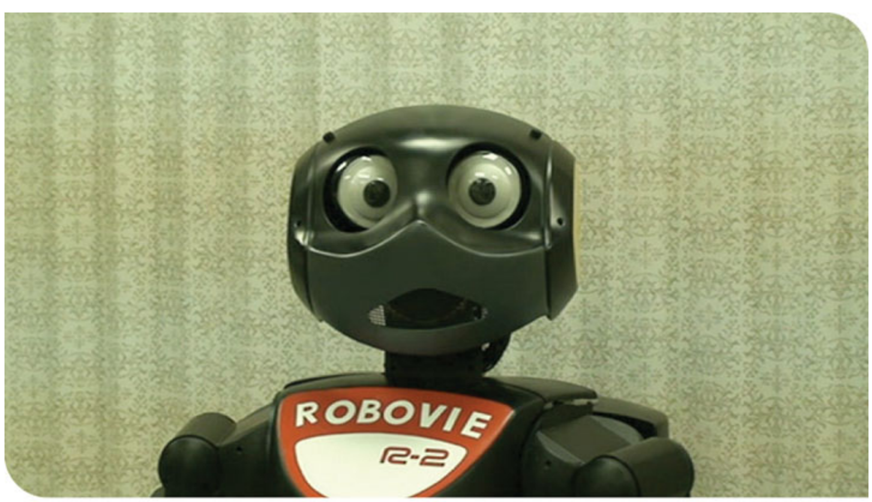


Figure 14. Robovie R-2 developed by ATR (Ishiguro et al., 2001).

conversation to reaffirm the role of the bystander but did not produce turn-yielding signals directed at him. The gaze behavior in three-party conversations involved acknowledging and looking at the two addressees equally and producing turn-yielding signals for both addressees. In looking at addressees, Robovie reproduced gaze behavior that followed the patterns identified for each configuration of conversational participation. Table 3 summarizes Robovie's gaze behavior in the three conversational configurations.

	Two-party conversation	*Two-party conversation with bystander*	*Three-party conversation*
Greeting	*Acknowledge the addressee*	*Acknowledge the addressee and then the bystander*	*Acknowledge one of the addressees and then the other addressee*
Participation structure (footing)	*Direct attention at the addressee at the transition from greeting to casual conversation and keep direction of attention at the addressee at all times*	*Direct attention at the addressee at the transition from greeting to casual conversation and keep direction of attention mostly at the addressee occasionally glancing at the bystander for short periods*	*Divide attention at both addressees at the transition from greeting to casual conversation producing turn-yielding signals for both addressees and wait for one of the them to take the floor* *Switch speakers at paragraphs (Hinds, 1976)*
Conversational structure (turn-taking)	*Turn-yielding: Look at the addressee at the end of a turn* *Turn-taking: Look at the addressee during minimal responses and look away from the addressee at the beginning of the turn*	*Turn-yielding: Look at the addressee at the end of a turn* *Turn-taking: Look at the addressee during minimal responses and look away from the addressee at the beginning of the turn*	*Turn-yielding: Look at the one of the addressees at the end of a turn* *Turn-yielding with speaker change: Look at one of the addressees and then the other and wait for one of them to take the floor* *Turn-taking: Look at the addressee who just passed the floor during minimal responses and look away at the beginning of the turn*
Information Structure	*Look in pattern "Look away > Look at > Look down" at the addressee* *Look in pattern "Look down > Look at > Look down" at the addressee*	*Look in pattern "Look away > Look at > Look down" at the addressee* *Look in pattern "Look down > Look at > Look down" at the addressee* *Short glances at the bystander at random intervals*	*Look in pattern "Look away > Look at > Look away" at one addressee at a time but at both addressees* *Look in pattern "Look away > Look at > Look down" at one addressee at a time but at both addressees*
Leave-taking	*Acknowledge the addressee*	*Acknowledge the addressee and then the bystander*	*Acknowledge one of the addressees and then the other addressee*

Table 3. A summary of Robovie's gaze behavior in the three conversational configurations.

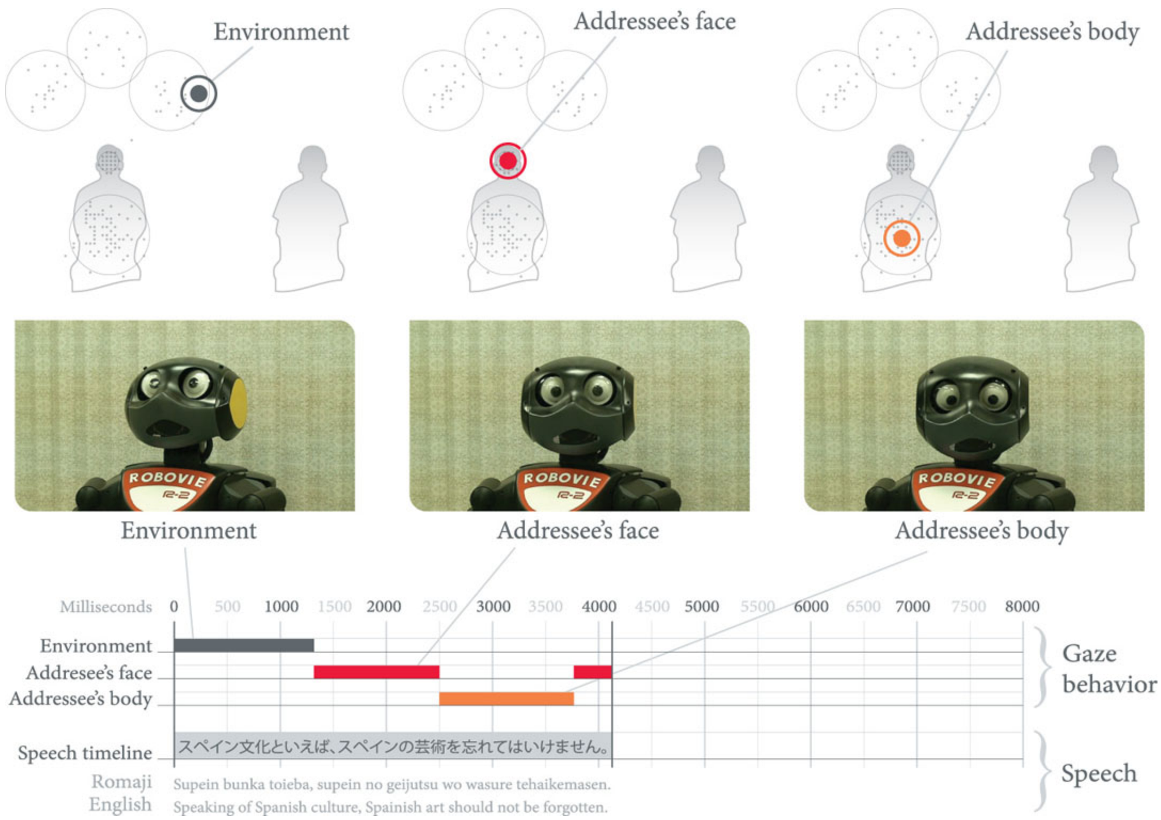


Figure 15. An example of Robovie's production of the gaze pattern "Environment > Addressee's face > Addressee's body" for a two-party conversation.

Robovie's speech was marked for turn-relevant places and thematic fields. The robot automatically produced the appropriate gaze behavior at these markers using probabilities of occurrence for gaze patterns and length distributions for each gaze target. For instance, in producing a turn-yielding signal, the robot looked at its addressee at the end of a question, while it listened to the answer, and while producing minimal responses. At this point, a new thematic field started and Robovie produced the appropriate pattern of gaze based on the probability of occurrence for the gaze pattern and calculated the length of the gaze shifts in the pattern based on the length distributions for each gaze. Figure 15 illustrates an example of Robovie's gaze behavior at the start of a thematic field.

4.2.2.2. Conversational Scenario

A conversational scenario was choreographed where Robovie played the role of a "travel agent robot" and provided people with travel information. To enrich conversational flow, the robot provided people with options of travel packages and destinations and adapted its information to people's choices. It also assessed people's knowledge of the countries to which they showed interested in traveling by asking them factual questions such as "Are you familiar with Picasso?" or "Did you know that Spain is this year's World Champion in basketball?" Wizard-of-Oz techniques were used in processing participant responses.

The robot followed common interaction rituals of a conversation. When people were introduced to Robovie, it introduced itself, asked participants for their names, and told them that it was happy to meet them. During leave-taking, it told participants that it had to talk to another customer, but it was nice meeting them, and thanked them for their interest. A prerecorded non-gendered voice was used for Robovie's speech.

4.2.3. Experimental Evaluation

Three hypotheses were developed drawing from existing theory on conversational participation, person perception, and group formation. To distinguish conversational participants from experimental participants, the former will hereafter be referred to as "participants" and latter as "subjects."

Greetings clarify the roles in a conversation. Acknowledging a person's presence during a greeting assigns the status of either a participant (e.g. an addressee in a conversation) or a non-participant (e.g. audience in a theatrical play). Not acknowledging a person's presence might lead to discomfort and negative perceptions of the speaker.

- *Hypothesis 1.* - Subjects whose presence is acknowledged will evaluate the robot more positively than those whose presence is not acknowledged. Therefore, a significant difference in subjects' liking between the first condition and the other conditions (aggregated) is predicted.

More participation in the conversation will lead to more attention and involvement in the task, therefore better task performance.

- *Hypothesis 2.* - Subjects who are treated as addressees will recall the details of the information presented by the robot better than those who are treated as bystanders, and subjects who are ignored.

More participation in the conversation will lead to stronger feelings of groupness.

- *Hypothesis 3.* - Subjects who are treated as addressees and take speaking turns will express stronger feelings of groupness (with the robot and the other subject) than those who are treated as non-participants or those whose presence is ignored. Therefore, a significant difference in subjects' feelings of groupness between the third condition and other conditions (aggregated) is predicted.

4.2.3.1. Experimental Design

To test these predictions, a between-subjects experiment was conducted where Robovie acted as a travel agent to give subjects information on travel packages. The robot first greeted subjects and introduced itself. It asked subjects for their names and told them that there are two special packages that they can choose from. After subjects chose a package, Robovie provided them with details of that package. Throughout the interaction, Robovie asked subjects questions regarding their travel preferences and their knowledge of the travel destination. Robovie's gaze behavior was manipulated in three conditions:

- In **condition 1**, one of the subjects was treated as an addressee and the other was ignored. The robot's gaze followed the patterns of a two-party conversation.

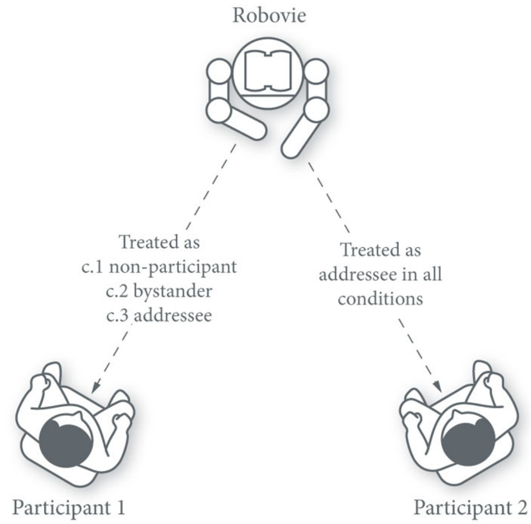


Figure 16. Spatial configuration of subjects and the robot.

- In **condition 2**, one of the subjects was treated as an addressee and the other was treated as a bystander. The robot's gaze followed the patterns of a two-party-with-bystander conversation.
- In **condition 3**, both subjects were treated as addressees. The robot's gaze followed the patterns of three-party conversation.

Robovie's speech was identical across conditions. Speech recognition was not used during the experiment. Instead, the experimenter initiated the robot's turns in the conversation and selected a preset sequence of utterances from a library. Figure 16 illustrates the spatial configuration of the robot and subjects.

4.2.3.2. Experiment Procedure

Subjects were first given a brief description of the purpose and the procedure of the experiment. After the introduction, they were asked to review and sign a consent form. Subjects were then provided with more detail on the task and asked to answer a pre-experiment questionnaire. Both subjects were told that researchers were developing a travel agent robot and would like to test their design with them. Subjects were provided with identical instructions and randomly assigned to the conditions in the experiment. They were told that after their interaction with the robot, they will be asked to answer a questionnaire on their experience and their recall of the material presented by the robot. After completing the task, subjects answered a post-experiment questionnaire on the information presented by the robot, their affective state, their perceptions of the robot, the group, and the task, and their demographic information. The experimental task and the whole experiment took an average of 7.5 minutes and 25 minutes respectively. The experiment was run in a dedicated space with no outside distraction. A male native-Japanese-speaking experimenter was present in the room during the experiment. All subjects were paid 3,000 ¥ (roughly $28 or €18) for their participation.

4.2.3.3. Measurement and Participation

The manipulation on the robot's gaze was the only independent variable. The dependent variables involved three kinds of measurements.

Objective - Subjects' recall of the information presented by the robot was measured using a post-experiment questionnaire.

Subjective - Subjective measures evaluated subjects' affective state using the PANAS scale (Watson et al., 1988), perceptions of the robot's physical, social, and intellectual characteristics using a scale developed to evaluate humanlike agents (Parise et al., 1998), feelings of closeness to the robot (Aron et al., 1992), feelings of groupness and ostracism Williams et al., 2000), perceptions of task characteristics, and demographic information.

The subjective evaluation also included a question for manipulation check. Subjects were asked how much they thought the robot looked at them and how much they thought the robot looked at the other subject. Seven-point Likert scales were used in all questionnaire items.

Behavioral - Subjects' gaze behavior and speech were captured using high-definition cameras at 1080i resolution. The cameras captured subjects' faces from upper torso to the top of their hair. The behavioral measures included the amount of mutual gaze with the robot and whether subjects took turns to answer Robovie's questions.

A total of 36 dyads (72 subjects) participated in the experiment. All subjects were native Japanese speakers. The experiment was limited to male subjects to eliminate gender effects. This limitation is further considered Discussion section of the study description (i.e. section 4.2.5). The ages of the subjects varied between 18 and 24 with an average of 20.8. All subjects were university students and were chosen to represent a variety of majors. Of all the subjects, 26 studied management sciences, 23 studied social sciences & humanities, 16 studied engineering, and 5 studied natural sciences. Subjects were randomly assigned to the experimental conditions. The computer use among subjects was very high (M=6.27, SD=0.98) on a scale from 1 to 7. Their familiarity with robots was

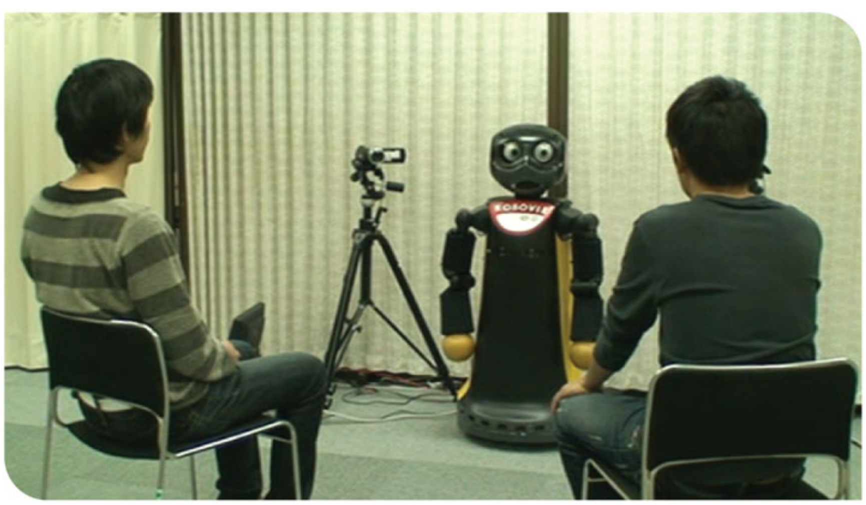


Figure 17. Subjects and Robovie in the experiment.

relatively low (M=2.97, SD=1.67), so was their video gaming experience (M=2.92, SD=1.91). Five (out of 72) subjects had toy robots and 23 owned pets. Figure 17 shows subjects in the experiment.

4.2.4. Results

Objective and subjective measures were analyzed using an analysis of covariance (ANCOVA). This method, similar to analysis of variance (ANOVA), applies a linear regression on the dependent variables that are significant across conditions to identify the direction of main effects and interactions while taking covariates into consideration that can account for some of the variance in data. This method was chosen to account for possible interactions between the two subjects in each trial. For instance, when a dependent variable for a set of subjects (e.g. those who are ignored, treated as bystanders or as addressees across the three conditions, marked with red, orange, and lime green colors respectively in the figures below) is analyzed, ratings from the other set of subjects (e.g. those who are treated as addressees across all conditions, marked with gray color in the figures below) are included in the model as a covariate. An ID number for each dyad is also included in the model as a random effect. Item reliabilities for scales and correlations across dependent measures were also calculated

Item reliabilities for the three-item scale that measured how much subjects liked the robot (alpha=0.76) and the six-item scale for measuring feelings of groupness (alpha=0.92) were high. An analysis of the manipulation check showed that subjects were aware of how much the robot looked at them. Figure 18 shows results for all pairwise tests on the manipulation check.

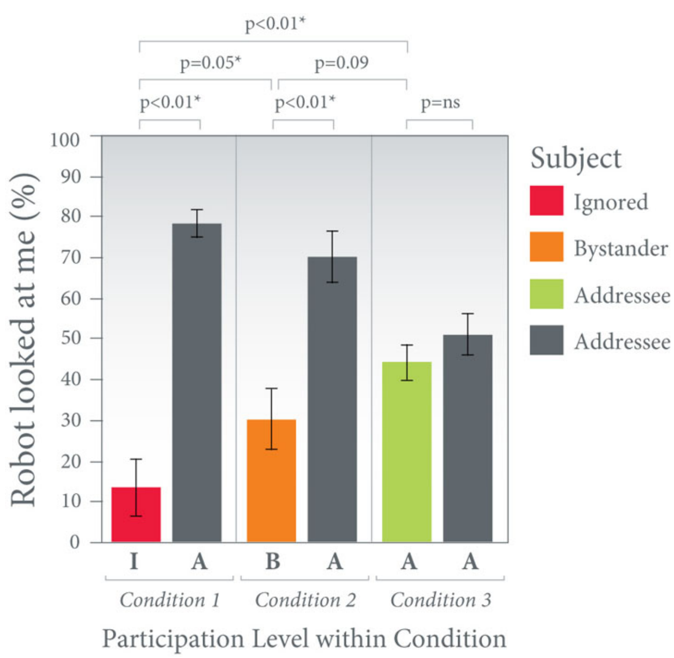


Figure 18. Manipulation check across levels of participation and condition. (*) indicates statistically significant probabilities below 0.05.

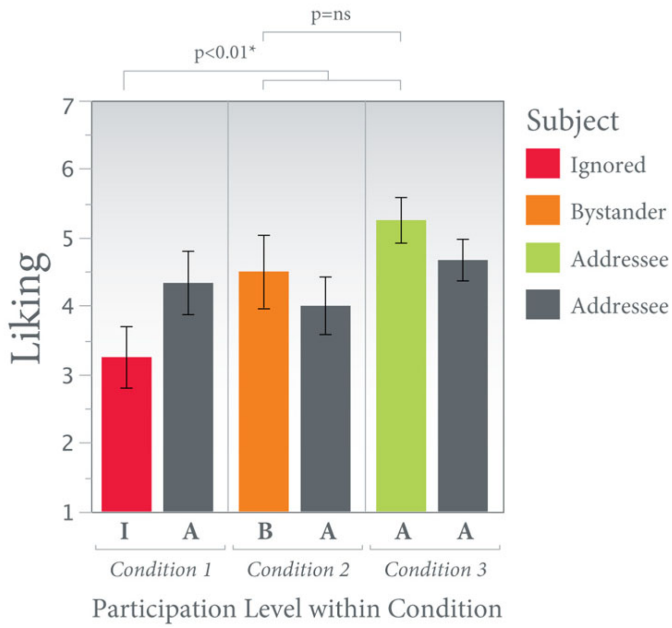


Figure 19. Participants' liking of the robot across levels of participation and condition. (*) indicates statistically significant probabilities below 0.05.

An analysis of covariance on subjects' liking of the robot supported the first hypothesis. Subjects whose presence was acknowledged (i.e. addressees and bystanders in condition 2 and 3) liked the robot more than those whose presence was not acknowledged (i.e. subjects who were ignored in condition 1) ($F[1:30]=9.03$, $p<0.01$). The difference between subjects who were treated as bystanders and those who were treated as addressees in how much they liked the robot was not significant ($F[1:30]=1.58$, p=ns) (see Figure 19).

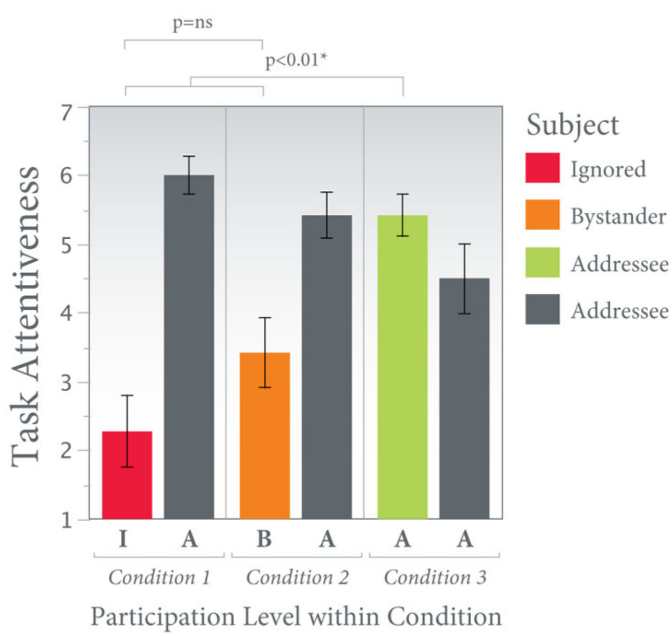


Figure 20. Task attentiveness across levels of participation and condition. (*) indicates statistically significant probabilities below 0.05.

The second hypothesis was not supported by results in task performance. All pairwise tests showed that subjects' recall of the information presented by Robovie was not affected by their participation (F[2:29]=1.14, p=ns). However, an analysis of subjects' ratings of their attentiveness to the task partially supported the second hypothesis. Subjects who were treated as addressees and who participated in the conversation attended to the task more than those who were treated as bystanders and those who were ignored (F[1:29]=15.00, p<0.01) (see Figure 20).

The third hypothesis was supported by the analysis. Subjects who were treated as addressees and who participated in the conversation rated their feelings of groupness significantly higher than people who were treated as bystanders or who were ignored (F[1:30]=12.92, p<0.01). The difference between subjects who were ignored and those who were treated as bystanders in their feelings of groupness was not significant (F[1:30]=0.34, p=ns) (see Figure 21).

Analyses on single-item scales showed that the experimental manipulation led people to feel ignored by the robot and that the robot did not consider their preferences in providing travel information. Subjects at whom the robot did not look felt ignored more than those whose presence was acknowledged and who were treated as bystanders (F[1:30]=5.33, p=0.03). Those who were treated as bystanders felt that the robot ignored them more than people who were treated as addressees and who participated in the conversation (F[1:30]=4.04, p=0.05). Participation led subjects to think that the robot considered their preferences in providing information. Those who participated in the conversation and who were treated as addressees felt that the robot considered their preferences more than subjects who were treated as bystanders or who were ignored (F[1:30]=8.77, p<0.01).

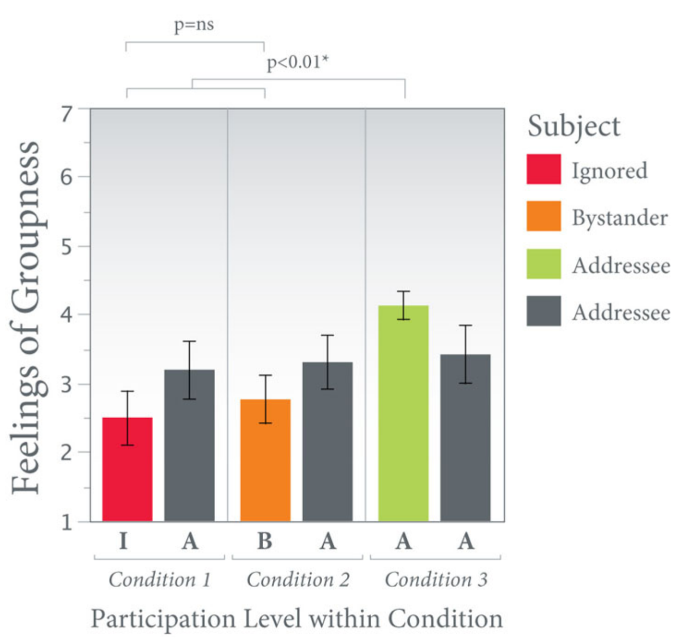


Figure 21. Feelings of groupness across levels of participation and condition. (*) indicates statistically significant probabilities below 0.05.

Finally, Pearson product-moment correlations were calculated to understand how dependent variables related to each other. These analyses showed small but statistically significant correlations between familiarity with robots and liking ($r=0.26$, $p=0.03$), task attentiveness ($r=0.25$, $p=0.04$), and feelings of groupness ($r=0.37$, $p<0.01$).

4.2.5. Discussion

The results on subjective evaluations showed strong support for all hypotheses. Subjects whose presence was acknowledged either as bystanders or addressees liked the robot more than people whose presence was not acknowledged. More interestingly, a simple acknowledging gaze at a partner during greeting led to their liking the robot more than they would if they were not acknowledged. The results also showed that more participation led to stronger feelings of groupness. Subjects who were treated as addresses and who participated in the conversation with the robot felt that they were part of the group more than those who did not participate.

Contrary to the prediction, subjects' participation in the conversation with the robot and whether the robot acknowledged their presence did not affect their recall of the information presented by the robot. However, subjective measures showed that participation in the conversation led to higher task attentiveness. That higher attentiveness in the task did not lead to better recall might be explained by subjects' established prior knowledge of the topic of the conversation. In support of this explanation, while participation or acknowledgement did not affect subject' recall performance, about which travel destination subjects chose to hear had a significant effect on the number of correct answers ($F[1:29]=11.10$, $p<0.01$). Administering a pre-experiment questionnaire to measure subjects' prior knowledge of the conversational topic would have helped in identifying how much new information is learned during the experiment.

4.2.5.1. Limitations

This second study has a number of limitations, some of which will be addressed in the third study. These limitations are discussed below. Other limitations will be discussed in the General Discussion section (i.e. Section 5) as a part of future work.

The design of the gaze behavior did not account for the coordination of the head and eyes for gaze shifts, mostly because of the method used to collect data for the design of gaze behavior. While direction of attention can be reliably coded from video data, how much changes in the head orientation and eye movements account for gaze shifts cannot be estimated accurately. The collection of the data for empirical grounding in Study III is done using eye-trackers and a motion capture system to precisely identify the use of eye movements and changes in head orientation in gaze behavior.

In understanding the factors that accounted for changes in the speaker's gaze, addressees' speech (during turns) was taken into consideration, but their behavior was not analyzed. In the third study, how the listener's gaze and nodding behavior affects speaker's gaze behavior will be analyzed and included in the gaze model.

Finally, how gaze behavior might change over the course of interaction was not examined in this study. In the third study, time will be considered as a factor in analyzing speakers' gaze behavior.

This study looked at how aspects of a robot's gaze can lead to different levels of participation and found that gaze cues that signal conversation structure (i.e. turn-taking) and participation structure (i.e. footing) led to different levels of conversational involvement, attentiveness, liking, and feelings of groupness. The next study will explore how a robot's gaze cues can direct its partner's attention towards an object in the environment. In the next subsection, a detailed plan of the study is provided.

4.3. **Proposed Study 3:** The Design of Gaze Cues for Joint Attention

The scenario that motivated the third experiment posed the following research questions: Can cues from a robot's gaze help in establishing joint attention in a way that it leads to better language learning? How can we design such cues and what are the design variables? Below is a preliminary description of the proposed third study.

4.3.1. Theoretically and Empirically Grounded Design

The goal of this study is to gain a deeper understanding of the role of gaze cues in establishing joint attention, to use existing theory and empirical data to design joint attention abilities in a robot, and to evaluate how these designed abilities affect human-robot communication. A review of related theory on joint attention, particularly conversational aspects of deictic gaze, will provide the design process with theoretical grounding. Empirical grounding will be achieved through a detailed analysis of how gaze cues are used in joint attention situations from data collected from human dyads. The data collection is completed. Details of the data collection setup is provided below. The analysis of the data will be similar to that of the second study with additional focus on the aspects of gaze behavior described below.

4.3.1.1. Changes in Gaze Patterns over the Course of the Interaction

Gaze behavior changes over the course of a conversation (Abele, 1986). While the first two studies did not take this factor into consideration, an understanding of how speakers' use of gaze cues might change over time is particularly important in deictic gaze. For instance, when parties talk about an object in their environment (e.g. a painting at a museum or a sales item at a department store), how much joint attention is established relative to the object might change as parties become more familiar with the object within the course of the conversation. Another possible scenario is that as parties become more familiar with each other, they look more at each other, which leads to a decrease in the amount of deictic gaze. An analysis of changes over time in the amount of deictic (verbal and gaze) references to the object of mutual attention and the total amounts and length distributions of gaze at the partner and at the object will be conducted.

4.3.1.2. The Effect of Addressee's Behaviors on Speaker's Gaze Production

Another factor that affects speaker gaze behavior is the gaze and gestures of addressees. Addressee gaze was only analyzed in designing turn-taking behavior in the second study.

However, cues from an addressee's gaze might account for a significant part of speaker's gaze shifts, particularly in joint attention situations. For instance, gaze following, the behavior of following the line of sight of another person (Emery, 2000), which is how people establish joint attention, can only be understood by looking at the gaze behavior of both parties in a dyad. Similarly, addressee head nods might account for changes in conversation structure and speaker gaze behavior. An analysis of this interaction is particularly important in the context of this study, because Japanese speakers are found to use head nods frequently (four times as much as Americans do) and for semantical, syntactical, and interactional purposes (Maynard, 1987).

4.3.1.3. Co-occurence of Verbal and Nonverbal Actions

Research has shown that head movements co-occur with verbal actions, particularly in producing deictic references (McClave, 2000). In the first study, theme-rheme and discourse boundaries were used to trigger gaze shifts. In the second study, gaze behavior was designed to coincide with turn and thematic field boundaries. In this third study, an analysis of the co-occurence between verbal and nonverbal deictic references within discourse will be conducted to develop a more accurate coupling between the robot's speech and gaze behavior in establishing joint attention.

4.3.1.4. Head-Eye Coordination

Gaze shifts are produced by a dynamic and socially significant interaction between changes in eye movements and head and body orientation (Emery, 2000; Frischen et al., 2007). The analyses of speaker gaze in the first two studies did not account for how eye, head, and body orientation contributed to gaze shifts. While the speakers in both studies were seated, meaning that their body orientation was mostly fixed, gaze shifts were produced by a combination of eye and head movements. To gain a rough understanding of how cues from the eyes and head are used in establishing joint attention, a pretest study was conducted where a male dyad was seated across from each other with an object on one side. One of the parties was asked to describe the object to the other. An analysis of video data showed that the speaker most often positioned his head between the direction of his addressee and that of the object and mostly moved his eyes to look at the addressee and at the object. This behavior was not predicted by the literature on gaze and suggests that head position can be informative about how a speaker's attention is divided between an addressee and environmental stimuli. The third study will try to distinguish head and eye movements in gaze production and identify patterns in how these cues are combined at syntactically, semantically, and interactionally significant points of the conversation.

4.3.2. Empirical Grounding Setup

To conduct an analysis of how gaze cues are used to establish joint attention in dyadic interaction, two interaction scenarios were created and subjects were asked to perform the scenarios while data was collected on their gaze behavior. In the first scenario, subjects took turns and told each other about their hometowns. In the second scenario, subjects were asked to try to sell an object familiar to them (i.e. their cellular phones) to each other. This task was observed to facilitate fluent and engaging conversations among tasks that were explored during the pretest. In both scenarios, subjects were seated across each

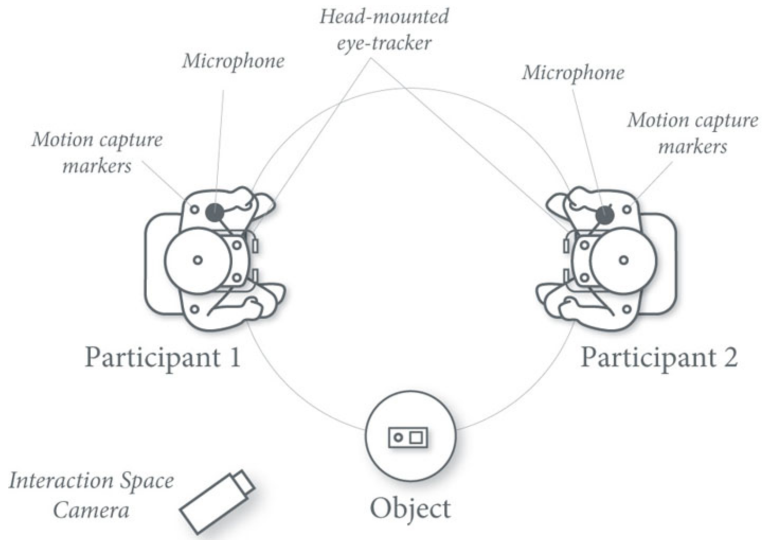


Figure 22. Data collection setup and equipment.

other. The object in the second scenario was placed on one side of the subjects as illustrated in Figure 22.

4.3.2.1. Measurement

Four kinds of measurements were conducted (see Figure 22). The interaction space was captured with a high-definition video camera at 1080i resolution. To capture eye movements, both subjects were equipped with head-mounted eye-tracekers. The eye-tracker data was coupled with head position data captured by a 12-camera Vicon motion-capture system. Finally, speech was recorded using stereo microphones.

4.3.2.2. Participation

Twenty male dyads (40 participants) participated in the data collection (e.g. Figure 23). Subjects were first provided with a description of the study, including an overview of the eye-tracking and motion-capture systems. They were then asked to review and sign a consent form. Next, subjects were provided with a pretest questionnaire that measured

Figure 23. Sample video from collected data. Images show (a) participant 1's (in Figure 22) field of vision from the eye-tracker camera mounted on the participant's head, (b) participant 2's field of view from his eye-tracker camera, and (c) the interaction space.

introversion-extroversion. They were then taken to a dedicated laboratory space with no distractions. At the end of their participation, each subject was paid 3,000 ¥ (roughly $29 or €19).

Both subjects were asked to perform each scenario for five minutes. Between taking turns and the two scenarios, they were asked to solve ten-minute-long crossword puzzles as a distraction from their previous task. The overall study took 35 minutes. Two male experimenters were present in the room during the study, but were seated behind a cloth screen to minimize visual distractions.

4.3.3. Algorithmic Implementation

The designed gaze behaviors will be implemented on ATR's Geminoid robot (Nishio et al., 2007). The robot is designed with a highly humanlike appearance and controllable facial features including head and eye orientation. Figure 24 illustrates a pretest implementation of the gaze behavior on Geminoid using (a) only eye cues, (b) only head cues, and (c) eye and head cues combined.

4.3.4. Experimental Evaluation

Two hypotheses were generated drawing from existing theory on joint attention and learning. Research on language learning has shown that teacher's nonverbal behaviors play an important role in learning new vocabulary (Lazaraton, 2004). While no research has looked at the role of diectic gaze in adult learning, infants' learning of new vocabulary is facilitated by following a caretaker's gaze towards objects while hearing the names of the objects (Baldwin, 1995; Morales et al., 2000) implying that people make associations between visual stimuli and verbal content.

- *Hypothesis 1.* - Robot's deictic references will facilitate the development of similar associations: Subjects with whom the robot uses deictic gaze references to look at an object in one of the two displayed colors will more strongly associate the object with the color at which the robot looked than participants with whom the robot does not use these references.

While effect of gaze following is only studied in the context of infant language development, Lazaraton (2004) found frequent use of deictic references by foreign

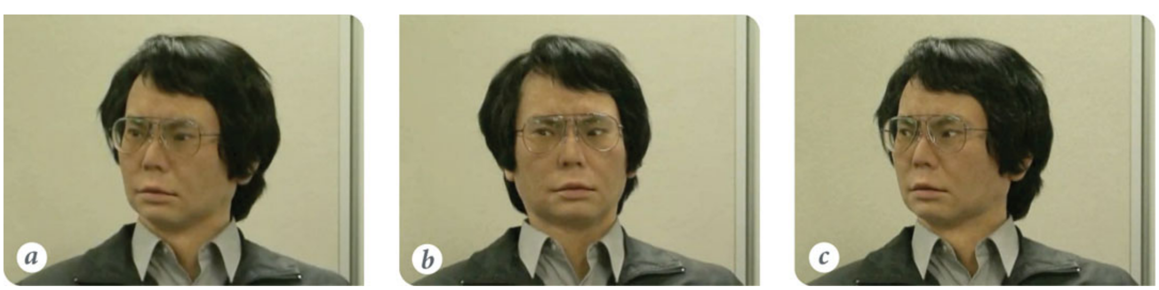


Figure 24. Geminoid developed by ATR (Nishio et al., 2007). Images show the robot's gaze behavior (a) using only head movements, (b) using only eye movements, and (c) head and eye movements combined.

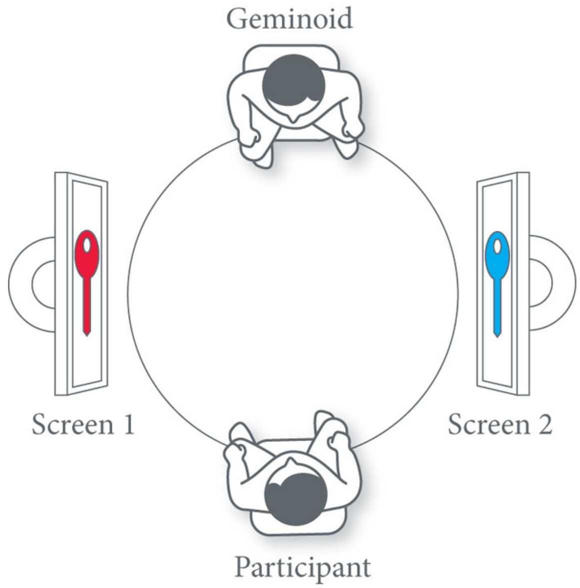


Figure 25. Spatial configuration of the task space, the robot, the participant, and the screens.

language teachers in teaching adults, which might play an important role in language learning.

- *Hypothesis 2.* - Subjects with whom the robot uses deictic gaze references while teaching new words in a foreign language will have better recall of the words than those with whom the robot does not use these these references.

4.3.4.1. Experimental Design

To evaluate these predictions, a between-subjects-design experiment will be conducted with Geminoid. Subjects will be asked to perform in a task where Geminoid will teach Japanese participants English names for a number of objects. Subjects will be seated across from Geminoid. Figure 25 provides an illustration of the experimental setup. Two computer screens will be placed on the two sides of the dyad. As objects appear on the screen, Geminoid will pronounce the English names of the objects, ask subjects to pronounce them, describe their Japanese meanings, and give examples of how the words are used in casual speaking. The two screens will show pictures of the same object in two distinct colors (e.g. blue and red). Geminoid's deictic gaze references will be directed at one of the screens. The prediction is that subjects will associate the name of the object with the color on the screen at which the robot gazes. The robot's gaze will be manipulated to either use or not use deictic gaze references. The use of referential gaze cues is expected to significantly improve subjects' recall of the names in the post-experiment questionnaire.

4.3.4.2. Experimental Procedure

Subjects will be provided with a description of the purpose of the experiment. They will be told that researchers are developing an English teacher robot and would like their help in testing their design. Subjects will be asked to review and sign a consent form, which will be followed by pre-experiment questionnaire that will test subjects' English proficiency, particularly vocabulary. During the experimental task, 25 objects will be displayed on the screen. Geminoid will first ask whether the subject knows the English word for the object. If the subject is familiar with the word, a new object will be displayed. Otherwise, the robot will proceed to teaching the English name of the object. The task is expected to take at an average of 12 minutes. After completing the task, subjects will be asked to answer a post-experiment questionnaire. They will be paid 3,000 ¥ (roughly $28 or €18) for their participation.

4.3.4.3. Measurement

As in the second experiment, objective, subjective, and behavioral measures will be used in assessing the social outcome.

Objective - Subjects will be provided with two questionnaires at the end of the experiment. First, they will be given a set of pictures of objects (i.e. those that appeared in the experimental task) and a set of English names and asked to match objects with the correct English names. In the second questionnaire, they will be given the same set of pictures of objects and a set of colors and asked to associate the object with a color.

Subjective - Before subjects perform the experimental task, they will be given a pre-experiment questionnaire that will measure affective state using the PANAS scale (Watson et al., 1988). After the experimental task, subjects will be provided with post-experiment questionnaires that will measure post-experiment affective state using the PANAS scale (Watson et al., 1988), perceptions of the robot's physical, social, and intellectual characteristics using the scale that was administered in the first two studies (Parise et al., 1998), perceptions of task characteristics, and demographic information.

Behavioral - Subjects' gaze behavior and speech will be recorded during the experiment. The data will be used to analyze gaze following behavior, the total amount of joint attention and mutual gaze established with the robot, and conversational flow.

This section provided detailed descriptions of the two completed studies and a planned study. In the next section, a brief discussion of the findings from the the completed studies and the expected outcome of the third study is provided. The limitations of these studies and future work are also discussed.

5. General Discussion

The primary goal of the proposed dissertation is to understand how robot gaze might serve as a communicative mechanism in a variety of social situations. Study I showed that by looking at a person longer, a robot can lead to better recall of information. It also showed that women felt uncomfortable with being looked at too much by a robot while men liked the robot more when it looked at them more. In Study II, a robot was shown to regulate who participates in a conversation using simple gaze cues. Participation in a conversation with a robot was also shown to heighten task attentiveness, positive evaluations of the robot, and feelings of groupness with the robot. Study III will show whether deictic gaze references could lead to better learning of names through facilitating associations between words and pictures of objects.

The process of designing gaze behaviors has also led and will lead to new knowledge about the design variables for social gaze behavior. In Study I, the analysis of the professional storyteller's gaze behavior revealed where and how long oratory speakers look at in the presence of two addressees. Study II allowed for a comparison of gaze targets and durations across different participation structures. The analysis also confirmed existing knowledge about gaze behavior during conversational turn-taking. Additionally, speakers were found to not only look at and away from their addressees, but also do so in patterns. Study III is expected to inform the research on how verbal and nonverbal deictic gaze references relate to each other and how head and eye movements are coordinated in gaze production. These findings are expected to contribute to the design of humanlike robots and our understanding of human communication.

5.1. Limitations

The studies described here also have a number of limitations that restrict the findings to the tasks and research platforms used, populations, cultural contexts, and languages studied, and the design decisions that were made in creating gaze behaviors and experimental scenarios.

5.1.1. Gender

Because gender has a significant effect on the production and perception of gaze behavior, the studied populations were restricted to minimize gender bias, which in return limits the generalizability of the results of the studies. In Study I, the design of ASIMO's gaze behavior was based on an all-female triad (a female speaker with two female addressees). Therefore, whether the results could be replicated with a design based on a male speaker is unknown. In Study II, both the design and evaluation of Robovie's gaze behaviors were based on all-male triads (a male speaker and two male addressees). The results of this study can only be generalized to male populations. Furthermore, the design of the gaze behavior in Study III will be based on male dyads. While limiting population gender is common in gaze research (Exline, 1963; Argyle and Ingham, 1972), it places restrictions on the generalizability of the results.

5.1.2. Culture and Language

Gaze behavior is also found to be sensitive to cultural context and language. For instance, Ingham (1972, as described in Argyle and Cook, 1976) found significant differences in how much, how long, and how often Swedes and Englishmen looked at their partners during conversation. Therefore, designed behaviors as well as the social outcomes that they lead to are limited to the cultural context and language of the study. In Study I, ASIMO's gaze behavior was designed based on data collected from an English-speaking Icelander speaker and two English-speaking American addressees. Whether using data from a speaker with a different origin would lead to differences in the design of the behavior is unknown. Also, American, native-English-speaker participants were hired to evaluate the gaze behavior. Whether the results from the experiment would generalize to other populations is not known. Study II involved all native-Japanese speakers for both the design and evaluation of the gaze behavior. Study III will also involve all native-Japanese speakers. Whether results from these experiment would apply to non-Japanese populations needs further investigation.

5.1.3. Tasks

The tasks used in these studies also places some limitations on the generalizability of their results. For instance, topic of conversation is found to affect how much people look at each other (Abele, 1986). Study I used storytelling as the context of the study. In Study II, Robovie provided travel information. In Study III, Geminoid will teach English vocabulary. Whether the results from these studies would hold with different tasks and conversation topics is unknown.

5.1.4. Research Platforms

Another important limitation of this research is imposed by the designs of the research platforms used in the studies. Study I used ASIMO, which has a dark screen instead of discrete eyes. That gaze production on the robot relies only on head movements might have intensified or reduced the feeling of being looked at. In contrast, Study III will use Geminoid, which has a highly humanlike appearance, which may strengthen referential signals. Alternatively, the designed gaze behavior may be perceived as crude because of the mismatch between appearance and behavior, which might reduce the outcome. Furthermore, while theoretical grounding provides the designed gaze behavior a certain level of generalizability, how these findings might be replicated with other robotic systems is unknown.

5.1.5. Limited Interactivity

An important limitation of this research is the controlled and limited interaction people have with robots. While designed gaze behaviors were implemented algorithmically and gaze was produced automatically and adaptively to robots' speech, other aspects of the interaction relied on the use of Wizard-of-Oz techniques. For instance, in the first two studies, robots did not sense subjects' locations. Instead, they were seated at designated locations and the robots were programmed to look at these locations. Similarly, speaking turns in the second study were controlled by an operator behind the robot. Robust vision and natural language techniques are required to address these issues and allow the construction of a truly interactive experience for the participants.

5.1.6. Singling Gaze Out

Perhaps the most important limitation of all three studies is that gaze is singled out of other visible behavior. In Study I, arm and body postures were used to enrich ASIMO's behavior as a storyteller. Because adding these gestures might have confounded the results of the study, gestures were eliminated in the second and third studies. However, other forms of nonverbal expressions such as facial expressions, arm, head, and body gestures, and posture co-construct visible behavior and work in harmony to produce human language. Therefore, future work should look at integrating different forms of nonverbal behavior. Furthermore, when highly humanlike research platforms are used, such as the robot used in Study III, idle behaviors such as breathing and small, patterned movements might be required to achieve lifelike behavior.

5.1.7. Design Approach

The particular design approach taken in this research also introduces certain limitations. The design process followed by this research aims at creating natural, humanlike social gaze behaviors that people interpret as valid communicative stimuli. With this goal, the process involved rigorous steps to ground design decisions in theory and empirical findings. However, other approaches that produce a similar outcome might exist. For instance, animation artists are trained to follow a set of guidelines to create lifelike behaviors for animated characters. Whether these behaviors create a particular social outcome is not tested. Furthermore, whether communicative behaviors created using this approach would produce stronger (or weaker) results than those of this research is unknown.

5.2. Future Work

Some of the limitations discussed above provide a roadmap for future work on gaze and the design of humanlike behavior for robots. Below, a discussion of possible future directions that stem from the limitations of current research is provided.

5.2.1. Platforms and Modalities

Future work needs to provide an understanding of the extent that these findings would carry over to interactions with other robots to ensure generalizability across platforms. This understanding could be achieved through replicating these studies on other platforms or conducting comparative studies on multiple platforms that look at how the physical design of the robot affects the way that human communicative mechanisms respond to the robot. Furthermore, future research should look at the extent that these findings would carry over to other modalities (i.e. interactions through video, with on-screen agents, etc.) and levels of agency (i.e. autonomous agents vs. avatars).

5.2.2. Culture, Language, and User Attributes

Studies that compare these results across cultures, languages, and user attributes (e.g. gender, age, personality, social status, occupations, etc.) would also significantly improve the generalizability of the results of this research. Future work should look at how designed behaviors could be extended to robots that work in different cultural contexts, use different languages, and interact with people at different ages, with different personality attributes, and so on.

5.2.3. Design Approaches

Future research should also compare the outcomes of different approaches to designing humanlike behaviors. Furthermore, while the approach taken in this research might create behaviors that better fit to a robot with highly humanlike appearance, an animation artist's approach that is grounded in communication theory might create behaviors that are more appropriate for a robot with an abstract design.

5.2.4. Gaze as a Part of Visible Behavior

Finally, future work should look at how different nonverbal behaviors could be combined to create visible behavior. Behaviors of particular communicative importance are body orientation, head, arm, and hand gestures and facial expressions.

This section briefly discussed the findings of the completed studies and the expected outcome of the planned study and described the limitations of the studies and future work. Next section will outline the proposed schedule of the dissertation.

6. Proposed Schedule

My proposed schedule is shown in Figure X. I plan to spend the next two months finalizing the analysis of the evaluation data from Study II, publishing my results from Study II, and completing the behavioral modeling for Study III. From early August to late September, I plan to spend seven weeks at ATR in Japan implementing and conducting Study III. When I return from ATR, I will spend the next two months analyzing the evaluation data from Study III. I plan to spend the Spring of 2009 writing the dissertation and publishing my results from Study III. I hope to defend my dissertation early in May 2007 and spend the next month completing revisions on the dissertation document. Figure 26 shows the timeline for the proposed work.

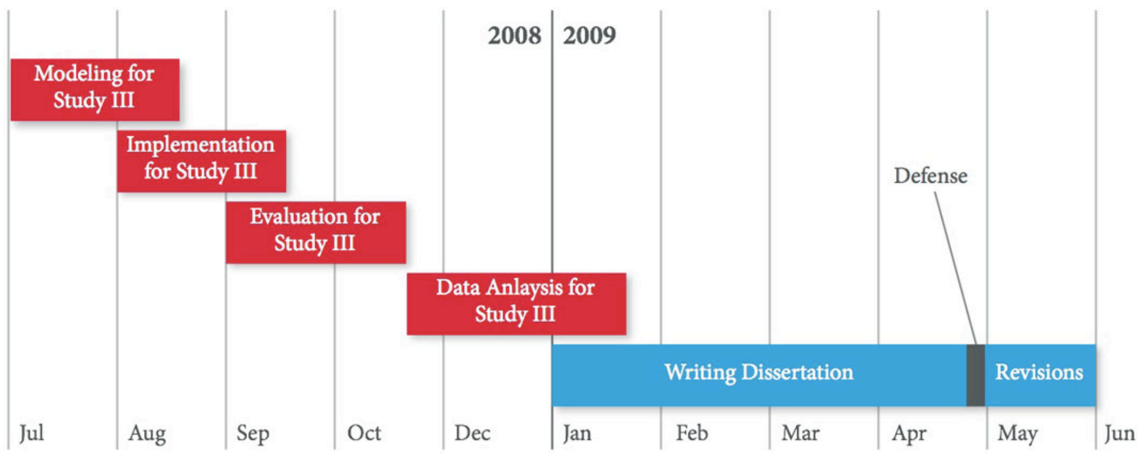


Figure 26. Timeline for the proposed work.

Acknowledgements

First and foremost, I would like to thank my advisors, Jodi Forlizzi and Jessica Hodgins, for joining forces to help me with the innumerable facets of my work and for mentoring me towards becoming an academic. I am incredibly grateful to Sara Kiesler for her guidance throughout my doctoral education. I would also like to thank Justine Cassell for her valuable advice and feedback as well as for her work that has been an inspiration in devising my own.

I am thankful to all my co-authors and collaborators for their hard work and valuable contributions including Jeremy Bailenson, Susan Fussell, Norohiro Hagita, Hiroshi Ishiguro, Takayuki Kanda, Toshiyuki Shiwa, and Fumitaka Yamaoka. Several others have provided help and contributed to discussions around this work including Nathan Frier, Darren Gergle, Peter Kahn, Clifford Nass, Leila Takayama, and Cristen Torrey, as well as organizers and participants of DIS 2006 and CHI 2008 Doctoral Consortiums, particularly Judith Olson and Alistair Sutcliffe.

I would also like to thank all those who provided support and feedback on this work including Sonya Allin, Laura Dabbish, Matthew Easterday, Rachel Gockley, Scott Hudson, Amy Hurst, Andy Ko, Queenie Kravitz, Johnny Lee, Joonhwan Lee, Ian Li, Marek Michalowski, Selma Sabanovic, Irina Shklovski, Karen Tang, and John Zimmerman. I would also like to thank my partner Lindsay Jacobs, my brother Ilker Mutlu, and my parents Hanife and Mustafa for their continuing support of my career.

This work is funded in part by National Science Foundation grants NSF IIS-0121426 and NSF HSD IIS-0624275. Fellowships and equipment grants from ATR International, Japan and Honda R&D Co., Ltd. have also made this work possible.

References

Abele, A. (1986). Functions of gaze in social interaction: Communication and monitoring. Journal of Nonverbal Behavior, 10 (2), 83-101.

Andersen, H. C. (2001). Tales. Vol. XVII, Part 3, The Harvard Classics. P.F. Collier & Son, 1909-14; Bartleby.com.

Argyle, M. (1972). The Psychology of Interpersonal Behavior. Harmondsworth: Penguin Books, 2nd edition.

Argyle, M. and Dean, J. (1965). Eye-Contact, Distance and Affiliation. Sociometry, 28 (3), pp. 289-304.

Argyle, M. and Ingham, R. (1972). Gaze, mutual gaze and proximity. Semiotica, 6, 32-49.

Argyle, M. and Cook, M. (1976). Gaze and mutual gaze. Cambridge: Cambridge University Press.

Argyle, M. and Graham, J. A. (1976). The Central Europe experiment - looking at persons and looking at objects. Journal of Environmental Psychology and Nonverbal Behaviour, 1 (1), 1-16.

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. Journal of Personality and Social Psychology, 63 (4), 596-612.

Bailenson, J.N., Blascovich, J., Beall, A.C. and Loomis, J.M. (2001). Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. Presence: Teleoperators & Virtual Environments, 10 (6), 583-598.

Bailenson, J.N., Beall, A.C., Loomis, J., Blascovich, J., and Turk, M. (2005). Transformed Social Interaction, Augmented Gaze, and Social Influence in Immersive Virtual Environments. Human Communication Research, 31 (4), 511--537.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), Joint attention: Its origins and role in development (pp. 131–158). Hillsdale, NJ: Erlbaum.

Bales, R. F. (1970). Personality and Interpersonal Behavior. New York: Holt, Rinehart and Winston.

Bales, R. F., Strodtbeck, F. L., Mills, T. M. and Roseborough, M. E. (1951). Channels of communication in small groups. American Sociological Review, 16 (4), pp. 461-468.

Bargh, J. A., Chen, M., and Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. Journal of Personality and Social Psychology, 71 (2), 230-244.

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. J. Dunham (Eds.), Joint attention: Its origins and role in development (pp. 41–59). Hillsdale, NJ: Erlbaum.

Bateson, M., Nettle, D., and Robeerts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. Biology letters, 2 (3), 412-412.

Bayliss, A. P., Pellegrino, G., and Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. The Quarterly Journal of Experimental Psychology Section A, 58 (4), 631-650.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., and Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. Infant Behavior and Development, 23, 223–229.

Beattie, G. (1980). The role of language production processes in the organization of behavior in face-to-face interaction. In B. Butterworth (Ed.) Language Production, Vol. 1, 69-107.

Bennewitz, M., Faber, F., Joho, D., Schreiber, M. and Behnke, S. (2005). Towards a humanoid museum guide robot that interacts with multiple persons. In Proceedings of 5th IEEE-RAS International Conference on Humanoid Robots.

Bouman, C. A. (1997). Cluster: An unsupervised algorithm for modeling Gaussian mixtures. Available from www.ece.purdue.edu—\string~bouman.

Breazeal, C. (1998). A motivational system for regulating human-robot interaction. In proceedings of AAAI, pp. 436-477.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G. and Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005, pp. 708-713.

Brooks, D. M. (1985). The Teacher's Communicative Competence: The First Day of School. Theory into Practice, 24 (1), 63-70.

Brooks, R. A. and Breazeal, C. and Marjanovic, M. and Scassellati, B. and Williamson, M. (1999). The Cog Project: Building a Humanoid Robot. Computation for Metaphors, Analogy and Agents, 1562, pp. 52--87.

Brown, P. and Levinson, S. C. (1987). Politeness: Some universals in language use. New York: Cambridge University Press.

Bruce, A., Nourbakhsh, I., and Simmons, R. (2002). The role of expressiveness and attention in human-robot interaction. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA '02), pp. 4138-4142.

Burnham, T. C. and Hare, B. (2007). Engineering Human Cooperation: Does Involuntary Neural Activation Increase Public Goods Contributions in Adult Humans?. Human Nature, 18 (2), 88-108.

Butterworth, G. (1991). The ontogeny and phylogeny of joint visual attention. In A. Whiten (Ed.), Natural theories of mind: Evolution, development, and simulation of everyday mindreading. Oxford England: Blackwell.

Calinon, S. and Billard, A. (2006). Teaching a Humanoid Robot to Recognize and Reproduce Social Cues. In Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 346-351.

Callan, H., Chance, M., Pitcairn, M. (1973) Attention and advertance in human groups. Social Science Information 12: 27-41.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., and Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monographs of the Society for Research in Child Development, 63 (4), 1-143.

Cassell, J. (2007). Body Language: Lessons from the Near-Human. Genesis redux: essays in the history and philosophy of artificial life, Riskin, J. (Ed), pp. 346-374, University of Chicago Press.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In Computer Graphics Proceedings, Annual Conference Series, pages 413-420.

Cassell, J., Bickmore, T. W., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., and Yan, H. (1999a). Embodiment in Conversational Interfaces: Rea. In Proceedings of the Conference on Human Factors in Computing (CHI'99), pp. 520-527, Pittsburgh, PA, 1999.

Cassell, J, Torres, O., and Prevost, S. (1999b). Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation. In Machine Conversations. pp. 143-154. Norwell, MA: Kluwer.

Cassell, J. and Nakano, Y.I. and Bickmore, T.W. and Sidner, C.L. and Rich, C. (2001). Non-verbal cues for discourse structure. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, pp. 114--123. Association for Computational Linguistics Morristown: NJ, USA.

Cassell, J., Kopp, S., Tepper, P., Ferriman, K., and Striegnitz, K. (2008). Trading spaces: How humans and humanoids use speech and gesture to give directions. T. Conversational Informatics: An Engineering Approach, Nishida (Ed.), Wiley.

Chafe, W. (1979). The flow of thought and the flow of language. Syntax and semantics, 12, 159-181.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., and Drew, A. (2001). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind, Cognitive Development, 15 (4), 481-498.

Clark, H. H. (1996). Using Language. Cambridge: Cambridge University Press.

Clark, H. H. and Carlson, T. B. (1982). Hearers and Speech Acts. Language, vol. 58 (2), pp. 332-373.

Colburn, A. and Cohen, M.F. and Drucker, S. (2000). The Role of Eye Gaze in Avatar Mediated Conversational Interfaces, Microsoft Research Report, 81.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., and Ahluwalia, J. (2000). Sex differences in human neonatal social perception. Infant Behavior and Development, 23, 113–118.

Cook, M. (1977). Gaze and Mutual Gaze in Social Encounters. The American Scientist, 65 (3), 328-333.

Cook, M. and Smith, J. M. C. (1975). The role of gaze in impression formation. British Journal of Social and Clinical Psychology, 14, 19-25.

Coutts, L. M. and Schneider, F. W. (1975). Visual behavior in an unfocused interaction as a function of sex and distance. Journal of Experimental Social Psychology, 11, 64–77.

Dautenhahn, K., Werry, I., Rae, J., and Dickerson, P. (2002). Robotic playmates: Analysing interactive competencies of children with autism playing with a mobile robot. In: K. Dautenhahn, et al. (Eds.), Socially Intelligent Agents: Creating Relationships with Computers and Robots, Dordrecht: Kluwer Academic Publishers.

Duncan, S. (1974). Some signals and rules for taking speaking turns in conversations. In Nonverbal Communication: Readings with Commentary, Oxford University Press.

Ellsworth, P. C., Carlsmith, J. M., and Henson, A. (1972). The stare as a stimulus to flight in human subjects: A series of field experiments. Journal of Social Psychology, 21, 302-311.

Ellsworth, P. C. and Ludwig, L. M. (1972). Visual behavior in social interaction. Journal of Communication, 22, 375-403.

Elzinga, R. H. (1978). Temporal organization of conversation. Sociolinguistics Newsletter, 9(2), 21–29.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience and Biobehavioral Reviews, 24 (6), 581-604.

Exline, R. V. (1963). Explorations in the process of person perception: visual interaction in relation to competition, sex and need for affiliation. Journal of Personality, 31, 1-20.

Exline, R., Gray, D., and Schuette, D. (1965). Visual Behavior in a dyad as affected by interview content and sex of respondent. Journal of Personality and Social Psychology, 95, pp. 201-209.

Exline, R. V. and Winters, L. C. (1966). Affective relations and mutual glances in dyads. In S. S. Tomkins and C. Izard (eds.), Affect, Cognition, and Personality, London: Tavinstock.

Exline, R. V. and Eldridge, C. (1967). Effects of two patterns of a speaker's visual behavior upon the perception of the authenticity of his verbal message. Paper presented at the meetings of the Eastern Psychological Association, Boston, April, 1967.

Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. (2002). Eye contact detection in humans from birth. Proceedings of the National Academy of Sciences, 99, 9602–9605.

Forgas, J.P. (1979). Social episodes: the study of interaction routines. London; New York: Published in cooperation with the European Association of Experimental Social Psychology by Academic Press.

Francis, S. J. (1979). Sex differences in nonverbal behavior. Sex Roles, 5, 519-535.

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze Cueing of Attention: Visual Attention, Social Cognition, and Individual Differences. Psychological Bulletin, 133 (4), 694–724.

Fry, R. and Smith, G. F. (1975). The effects of feedback and eye contact on performance of a digit-encoding task. Journal of Social Psychology, 96, 145–146.

Garau, M., Slater, M., Bee, S., and Sasse, M. A. (2001). The impact of eye gaze on communication using humanoid avatars. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Garfinkel, H. and Sacks, H. (1970). On formal structures of practical actions, Theoretical sociology, McKinney, J. C. and Tirayakian, E. A (Eds.), pp./ 338-366, New York: Appleton-Century-Crofts.

Gibson, J. J. and Pick, A. D. (1963). Perception of Another Person's Looking Behavior. The American Journal of Psychology, 76 (3), 386-394.

Gill, M. J., Swann, W. B., and Silvera, D. H. (1998). On the genesis of confidence. Journal of Personality and Social Psychology, 75, 1101-1114.

Goffman, E. (1955), On face-work. Psychiatry, 18 (3), 213-231.

Goffman, E. (1963). Stigma: Notes on the management of spoiled identity. Englewood Cliffs, NJ: Prentice Hall.

Goffman, E. (1971). Relations in public. New York: Harper & Row.

Goffman, E. (1979). Footing. Semiotica, 25 (1/2), 1-29.

Goodwin, C. (1980). Restarts, pauses, and the achievement of mutual gaze at turn-beginning.

Sociological Inquiry 50:272–302.

Goodwin, C. (1981). Conversational organization: Interaction between speakers and hearers. New York: Academic Press.

Goodwin, M. H. (2000). Participation. Journal of Linguistic Anthropology, 9 (1-2), 177-180.

Haley, K. J. and Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evolution and Human Behavior, 26 (3), 245-256.

Hall, J. A. (1984). Nonverbal sex differences: Communication accuracy and expressive style. Baltimore, MD: Johns Hopkins University Press.

Halliday, M. (1967). Intonation and Grammar in British English. The Hague: Mouton.

Hanks, W. (1996). Language and Communicative Practices. Oxford: Westview Press.

Heylen, D., van Es, I., Nijholt, A. and van Dijk, B. (2005). Controlling the Gaze of Conversational Agents. In J. van Kuppevelt, L. Dybkjaer and O. Bernsen, Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems. Kluwer Academic Publishers.

Hietanen, J. K. (2002). Social attention orienting integrates visual information from head and body orientation. Psychological Research, 66, 174–179.

Hinds, J. (1976). Aspects of Japanese Discourse Structure. Tokyo: Kaitakusha.

Hood, B. M., Willen, J. D., and Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. Psychological Science, 9, 131–134.

Hymes, D. (1972). "Models of the Interaction of Language and Social Life," in J. J. Gumperz and D. Hymes (Eds.), Directions in Sociolinguistics: The Ethnography of Communication, pp. 35-77. New York: Holt, Rinehalt & Winston.

Imai, M., Kanda, T., Ono, T., Ishiguro, H. and Mase, K.. (2002). Robot mediated round table: Analysis of the effect of robot's gaze. In proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication, pp. 411-416

Imai, M., Ono, T., and Ishiguro, H. (2003). Physical relation and expression: joint attention for human-robot interaction. IEEE Transactions on Industrial Electronics, 50 (4), 636-643.

Ingham, R. J. (1972). Cross-cultural differences in social behavior. D.Phil. thesis, Oxford University.

Ishiguro, H., Ono, T., Imai, M., Maeda, T., Kanda, T., and Nakatsu, R. (2001). Robovie: an interactive humanoid robot. Industrial Robot: An International Journal, 28 (6), 498-504.

Johnson, W. L. and Rickel, J. (1997). Steve: an animated pedagogical agent for procedural training in virtual environments. ACM SIGART Bulletin, 8 (1-4), 16-21.

Kagan, J. and Lewis, M. (1965). Studies of attention in the human infant. Merrill-Palmer Quarterly, 11 (2), 95-127.

Kanda, T. and Ishiguro, H. and Ono, T. and Imai, M. and Nakatsu, R. (2002). Development and evaluation of an interactive humanoid robot "Robovie." In proceedings of IEEE International Conference on Robotics and Automation.

Kanda, T., Ishiguro, H., Imai, M., and Ono, T. (2003). Body Movement Analysis of Human-Robot Interaction. In proceedings of International Joint Conference on Artificial Intelligence, pp. 177-182.

Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H., and Anzai, Y. (2007). A humanoid robot that pretends to listen to route guidance from a human. Autonomous Robots, 22 (1), 87-100.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. Acta Psychol (Amst), 26 (1), pp. 22-63.

Kendon, A. and Cook, M. (1969). The consistency of gaze patterns in social interaction. British Journal of Psychology, 69, 481-494.

Kendon, A. (1973). The role of visible behavior in the organization of social interaction. Social communication and movement: Studies of interaction and expression in man and chimpanzee, Cranach, M. V. and Vine, I. (Eds.), pp. 29-74.

Kendon, A. and Ferber, A. (1973). A description of some human greetings. Comparative ecology and behaviour of primates, Michael, R. P. and Cook, J. H. (Eds.), pp. 591--668.

Kendon, A. (1978). Looking in conversation and the regulation of turns at talk: A comment on the papers of G. Seattle and D. R. Rutter et al. British Journal of Social and Clinical Psychology, 77, 23-24.

Kiesler, S. (2005). Fostering common ground in human-robot interaction. In Proceedings of IEEE International Workshop on Robot and Human Interactive Communication (ROMAN), pp. 729-734.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. Psychological Bulletin, 100, 78–100.

Kleck, R.E. and Nuessle, W. (1968). Congruence between the indicative and communicative functions of eye-contact in interpersonal relations. British Journal of Social and Clinical Psychology, 7, 241-246.

Kozima, H., Nakagawa, C., and Yano, H., (2003). Attention coupling as a prerequisite for social interaction. In Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN), pp. 109-114.

Kozima, H. and Nakagawa, C. and Yano, H. (2004). Can a robot empathize with people? Artificial Life and Robotics, 8(1), 83-88.

Kuhl, P. K., Tsao, F. M., and Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. Proceedings of the National Academy of Sciences, 100 (15), 9096-9101.

Kuno, Y., Sadazuka, K., Kawashima, M., Yamazaki, K., Yamazaki, A. and Kuzuoka, H. (2007). Museum guide robot based on sociological interaction analysis. In proceedings of the SIGCHI conference on Human Factors in Computing Systems.

Langer, E. J., Blank, A., and Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. Journal of Personality and Social Psychology, 36, 635-642.

Langton, S. R. H. and Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. Visual Cognition, 6, 541–567.

Langton, S. R. H., Watt, R. J. and Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention, Trends in Cognitive Sciences, 4 (2), 50-59.

Lazaraton, A. (2004). Gesture and Speech in the Vocabulary Explanations of One ESL Teacher: A Microanalytic Inquiry. Language Learning 54 (1) , 79–117.

Lee, S. P., Badler, J.B. and Badler, N.I. (2002). Eyes alive. ACM Transactions on Graphics TOG, 21 (3), pp. 637--644.

Lee, S., Lau, I. Y., Kiesler, S., and Chiu, C. Y. (2005). Human Mental Models of Humanoid Robots. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp. 2767-2772.

Lester, J. C., Voerman, J. L., Towns, S. G., and Callaway, C. B. (1999). Deictic believability: Coordinated gesture, locomotion, and speech in lifelike pedagogical agents. Applied Artificial Intelligence, 13 (4), 383-414.

Libby, W. L. (1970). Eye contact and direction of looking as stable individual differences. Journal of Experimental Research in Personality, 4, 303-312.

Lutchmaya, S. and Baron-Cohen, S. (2002). Human sex differences in social and non-social looking preferences, at 12 months of age. Infant Behavior and Development, 25, 319–325.

Lutchmaya, S., Baron-Cohen, S., and Raggatt, P. (2002). Foetal testosterone and eye contact in 12-month-old human infants. Infant Behavior and Development, 25, 327–335.

Mason, M. F., Tatkow, E. P., and Macrae, C. N. (2005). The look of love: Gaze shifts and person perception. Psychological Science, 16, 236–239.

Matsusaka, Y., Fujie, S. and Kobayashi, T. (2001). Modeling of Conversational Strategy for the Robot Participating in the Group Conversation. In Proceedings of Seventh European Conference on Speech Communication and Technology (ISCA).

Maynard, S. K. (1987). Interactional functions of a nonverbal sign: Head movement in Japanese dyadic casual conversation. Journal of Pragmatics, 11, 589-506.

Maynard, S. K. (1989). Japanese Conversation: Self-Contextualization through Structure and Interactional Management. Norwood, New Jersey: Alex.

McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. Journal of Pragmatics, 32 (7), 855-878.

McLaughlin, M. and Cody, J. (1982). Awkward silences: Behavioral antecedents and consequences of the conversational lapse. Human Communication Research, 8, 299-316.

Minato, T., Shimada, M., Ishiguro, H., and Itakura, S. (2004). Development of an android robot for studying human-robot interaction, Lecture Notes In Computer Science: Proceedings of the 17 th international conference on Innovations in applied artificial intelligence, 17 (20), pp. 424-434.

Mobbs, N. A. (1968). Eye-contact in relation to social introversion-extroversion. British Journal of Social and Clinical Psychology, 7, 305-306.

Moon, Y. and Nass, C. (1996). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. Communication Research, 23 (6), 651-674.

Moon, Y. and Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human-computer interaction. International Journal of Human-Computer Studies, 49, 79-94.

Moore, C. and Dunham, P. J. (1995). Joint Attention: Its Origins and Role in Development. Hillsdale, NJ: Lawrence Erlbaum Associates.

Morkes, J., Kernal, H. K., and Nass, C. (1999). Effects of humor in task oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. Human-Computer Interaction, 14, 395-435.

Mundy, P. and Gomes, A. (1998). Individual differences in joint attention skill development in the second year. Infant Behavior and Development, 21 (3), 469-482.

Mundy, P. and Newell, L. (2007). Attention, Joint Attention, and Social Cognition. Current Directions in Psychological Science. 16 (5), 269-274.

Mutlu, B., Forlizzi, J. and Hodgins, J. (2006). A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. In Proceedings of 2006 6th IEEE-RAS International Conference on Humanoid Robots.

Nagai, Y., Hosoda, K., Morita, A., and Asada, M. (2003). A constructive model for the development of joint attention. Connection Science, 15 (4), 211-229.

Nass, C. and Steuer, J. (1993). Computers and Social Actors. Human Communication Research, 19 (4), 504-527.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, C. (1995). Can computer personalities be human personalities? In Proceedings of Conference on Human Factors in Computing Systems, pp. 228-229.

Nass, C., Moon, Y., and Green, T. (1997). Are computers gender neutral? Gender stereotypic responses to computers. Journal of Applied Social Psychology, 27 (10), 864-876.

Nass, C., Moon, Y., and Carney, P. (1999) Are People Polite to Computers? Responses to Computer-Based Interviewing Systems. Journal of Applied Social Psychology 29 (5) , 1093–1109.

Nass, C. and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81-103.

Nichols, K., and Champness, B. (1971). Eye gaze and the GSR. Journal of Experimental Social Psychology, 7, 623–626.

Nielsen, G. (1962). Studies in Self Confrontation. Copenhagen: Monksgaard.

Nishio, S., Ishiguro, H. and Hagita, N. (2007). Geminoid: Teleoperated Android of an Existing Person, in Humanoid Robots: New Developments, de Pina Filho, A.C. eds., I-Tech Education and Publishing, Vienna, Austria.

Noller, P. (1984). Nonverbal Communication and Marital Interaction. Pergamon.

Nourbakhsh, I. (1999). An affective mobile robot educator with a full-time job, Artificial Intelligence, 114 (1–2), 95–124.

Olafsen, K. S., Ronning, J. A., Kaaresen, P. I., Ulvund, S. E., Handegard, B. H., and Dahl, L. B. (2006). Joint attention in term and preterm infants at 12 months corrected age: The significance of gender and intervention.

Otteson, J. P. and Otteson, C. R. (1980). Effects of teacher gaze on children's

story recall. Perceptual and Motor Skills, 50, 35–42, 1980.

Ozaki, Y. T. (1970). The Japanese Fairy Book. Tokyo: Tuttle Publishing.

Parise, S., Kiesler, S., Sproull, L., and Waters, K. (1998). Cooperating with life-like interface agents. IBM WRC Technical Report.

Patterson, M. L. (1976). An arousal model of interpersonal intimacy. Psychological Review, 83, 235–245.

Patterson, M. L., Webb, A., and Schwartz, W. (2002). Passing encounters: atterns of recognition and avoidance in pedestrians. Basic and Applied Social Psychology, 24 (1), 57—66.

Pelachaud, C., Carofiglio, V., De Carolis, B., de Rosis, F., and Poggi, I. (2002). Embodied contextual agent in information delivering application. In Proceedings of the First international Joint Conference on Autonomous Agents and Multiagent Systems: Part 2 (Bologna, Italy, July 15 - 19, 2002). AAMAS '02.

Pelphrey, K. A., Viola, R. J., and McCarthy, G. (2004). When strangers pass: Processing of mutual and averted gaze in the superior temporal sulcus. Psychological Science, 15, 598–603.

Perrett, D. I., Hietanen, J. K., Oram, M. W., Benson, P. J. and Rolls, E. T. (1992). Organisation and functions of cells responsive to faces in the temporal cortex. Philosophical Transactions of the Royal Society of London, Series B, 335, 23–30.

Perrett, D.I. and Emery, N.J. (1994). Understanding the intentions of others from visual signals: Neurophysiological evidence. Current Psychology of Cognition, 13 (5), 683-694.

Peters, C. and O'Sullivan, C. (2003). Bottom-up visual attention for virtual human animation. In Proceedings of 16th International Conference on Computer Animation and Social Agents, pp. 111-117.

Peters, C. (2005). Direction of attention perception for conversation initiation in virtual environments. T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist (Eds.), Lecture notes in computer science, pp. 215-228, Springer-Verlag, London.

Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., and Poggi, I. (2005). A model of attention and interest using gaze behavior. Lecture notes in computer science, T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist (Eds.), pp. 229-240, Springer-Verlag, London.

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results, Robotics and Autonomous Systems 42, 271–28.

Posner, M. I. (1980). Orienting of attention. The Quarterly Journal of Experimental Psychology, 32 (1), 3-25.

Reeves, B. and Nass, C. (1996). The media equation: how people treat computers, television, and new media like real people and places. New York, NY: Cambridge University Press.

Rehm, M. and Andre, E. (2005). Where do they look? Gaze behaviors of multiple users interacting with an embodied conversational agent. In Proceedings of Intelligent Virtual Agents (IVA).

Rickel, J. and Johnson, W. L. (1999). Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. Applied Artificial Intelligence, 13, 343-382.

Rowe, K. J. and Rowe, K. S. (1999). Investigating the relationship between students' attentive-inattentive behaviours in the classroom and their literacy progress,International Journal of Educational Research, 31 (1-2), pp. 81-117.

Ryokai, K., Vaucelle, C., and Cassell, J. (2003) Virtual peers as partners in storytelling and literacy learning. Journal of Computer Assisted Learning 19 (2) , 195-208.

Sacks, H., Schegloff, E. A. and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, vol. 50 (4) pp. 696-735.

Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., Fujimura, K., H.R.D.C., Ltd., and Saitama, J. (2002). The intelligent ASIMO: system overview and integration. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and System.

Scaife, M. and Bruner, J. (1975). The capacity for joint visual attention in the infant. Nature, 253, 265–266.

Scassellati, B. (1999). Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. Computation for Metaphors, Analogy and Agents, 1562/1999.

Scassellati, B. (2001). Foundations for a theory of mind for a humanoid robot, Ph.D. Thesis, Department of Electronics Engineering and Computer Science, MIT Press, Cambridge, MA.

Schegloff, E. A. (1968) Sequencing in Conversational Openings. American Anthropologist, 70 (6) pp. 1075-1095.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation, Language in Society, 29 (1), pp. 1--63.

Schober, M.F. and Brennan, S.E. (2003). Processes of interactive spoken discourse: The role of the partner. In Handbook of discourse processes, pp. 123--64.

Senju, A. and Hasegawa, T. (2005). Direct gaze captures visuospatial attention. Visual Cognition, 12, 127–144.

Sherwood, J. (1987). Facilitative effects of gaze upon learning. Perceptual and

Motor Skills, 2/3, (2), 1275–1278.

Sidner, C. L., Kidd, C. D., Lee, C., and Lesh, N. 2004. Where to look: a study of human-robot engagement. In Proceedings of the 9th international Conference on intelligent User interfaces (Funchal, Madeira, Portugal, January 13 - 16, 2004).

Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., Rich, C. (2005). Explorations in engagement for humans and robots, Artificial Intelligence, 166 (1-2), pp. 140-164.

Sidner, C. L., Lee, C., Morency, L. P., and Forlines, C. (2006). The effect of head-nod recognition in human-robot conversation. In proceedings of International Conference on Human-Robot Interaction, pp. 290-296.

Simmel, G. (1921). Sociology of the senses: Visual interaction. Introduction to the Science of Sociology, Park, R. E. and Burgess, E. W. (Eds.), University of Chicago Press.

Snyder, M., Grather, J., and Keller, K. (1974). Staring and Compliance: A Field Experiment on Hitchhiking. Journal of Applied Social Psychology, 4 (2), 165-170.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., and Waters, K. (1996). When the interface is a face. Human-Computer Interaction, 11 (2), 97-124.

Staudte, M. and Crocker, M. (2008). The utility of gaze in spoken human-robot interaction. In Proceedings of Metrics for Human-Robot Interaction 2008 Workshop held at the 3rd ACM/IEEE International Conference on Human-Robot Interaction.

Strongman, K. T. and Champness, B. G. (1968). Dominance hierarchies and conflict in eye contact. Acta Psychologica, 28, 376-386.

Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., and Hagita, N. (2005). Three-layered draw-attention model for humanoid robots with gestures and verbal cues. In proceedings of 2005 IEEE/RSJ International Conference Intelligent Robots and Systems (IROS), pp. 2423--2428.

Tannen, D. (1984). Conversational Style: Analyzing Talk among Friends. Norwood, New Jersey: Ablex.

Thorisson K. R. (2002) Natural turn-taking needs no manual: Computational theory and model, from perception to action. In B. Granstrom, D. House, and I. Karlsson, editors, Multimodality in Language and Speech Systems, pp. 173–207. Kluwer Academic Publishers.

Tomasello, M. (1995). Joint attention as social cognition. In: Moore, C., and Dunham, P. (Eds.), Joint attention: its origins and role in development (pp. 85-101). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trafton, J. G., Bugajska, M. D., Fransen,, B. R., & Ratwani, R. M. (2008). Integrating vision and audition withing a cognitive architecture to track conversations. In proceedings of the 3rd international conference on Human Robot Interaction.

Vertegaal, R, Van der Veer, G., and Vons, H. (2000). Effects of Gaze on Multiparty Mediated Communication. Proceedings of Graphics Interface, pp. 92-102.

Vertegaal, R. and Slagter, R. and van der Veer, G. and Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 301--308.

Vilhjalmsson, H. H. and Cassell, J. 1998. BodyChat: autonomous communicative behaviors in avatars. In Proceedings of the Second international Conference on Autonomous Agents (Minneapolis, Minnesota, United States, May 10 - 13, 1998).

Von Cranach, M. and Ellgring, J.H. (1973). The perception of looking behaviour. In Von Cranach, M. and Vine, I. (Ed.), Social Communication and Movement. London: Academic Press.

Van Houten, R., Nau, P., Mackenzie-Keating, S. Sameoto, D., and Colavecchia, B. (1982). An analysis of some variables influencing the effectiveness of reprimands. Journal of Applied Behavior Analysis, 15, 65-83.

Wang, N., Johnson, W. L., Rizzo, P., Shaw, E., and Mayer, R. E. (2005). Experimental evaluation of polite interaction tactics for pedagogical agents. In Proceedings of the 10th International Conference on Intelligent User Interfaces, pp. 12-19.

Watson, O. M. (1970). Proxemic behavior: a cross-cultural study. The Hague: Mouton.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive an negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54 (6), 1063-1070.

Weisbrod, R. M. (1965). Looking Behavior in a Discussion Group. Unpublished paper, Department of Psychology, Cornell University.

Whitman, B. H. (1929). A short scale for measuring introversion and extroversion. Journal of Applied School Psychology, 13, 499-500.

Wilkes-Gibbs, D. and Clark, H. H. (1992). Coordinating beliefs in conversation. Journal of memory and language, vol. 31 (2) pp. 183-194.

Williams, K., Cheung, K. T., and Choi, W. (2000). Cyberostracisms: Effects of being ignored over the internet. Journal of Personality and Social Psychology, 79, 748 –762.

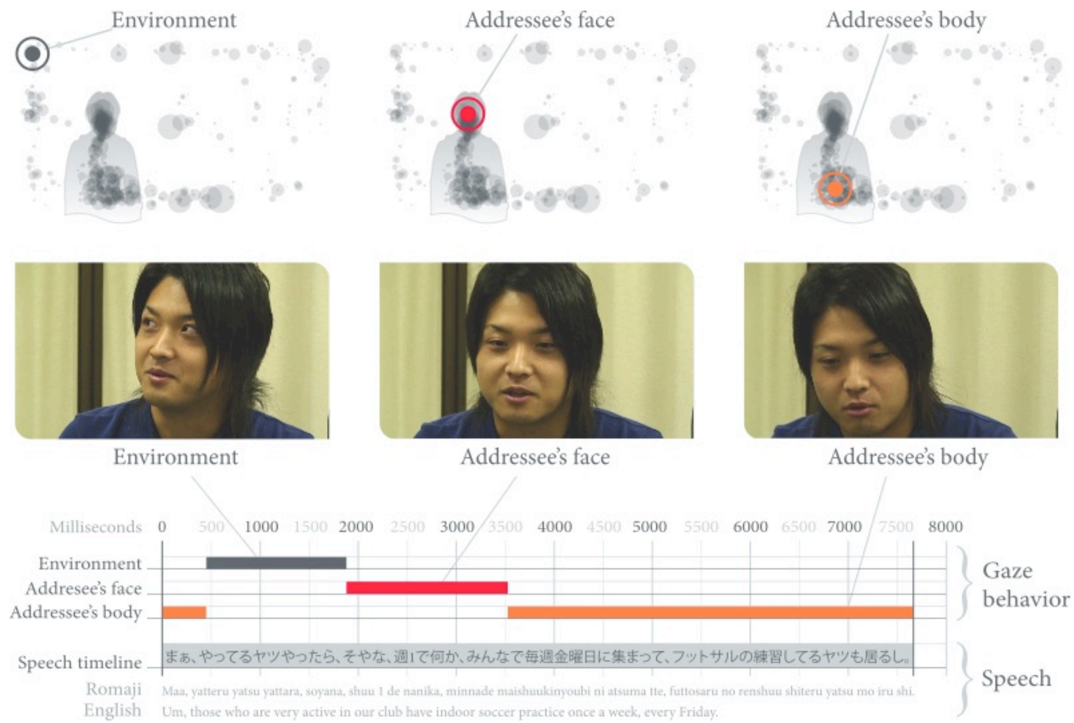
Wolff, P. H. (1963). Observations on the early development of smiling. Determinants of Infant Behavior, Vol 2, Foss, B. M. (Ed.), pp. 113-138. London: Methuen.

Woolfolk, A. E. and Brooks, D. M. (1985). The Influence of Teachers' Nonverbal Behaviors on Students' Perceptions and Performance. The Elementary School Journal, 85 (4), pp. 513-528.

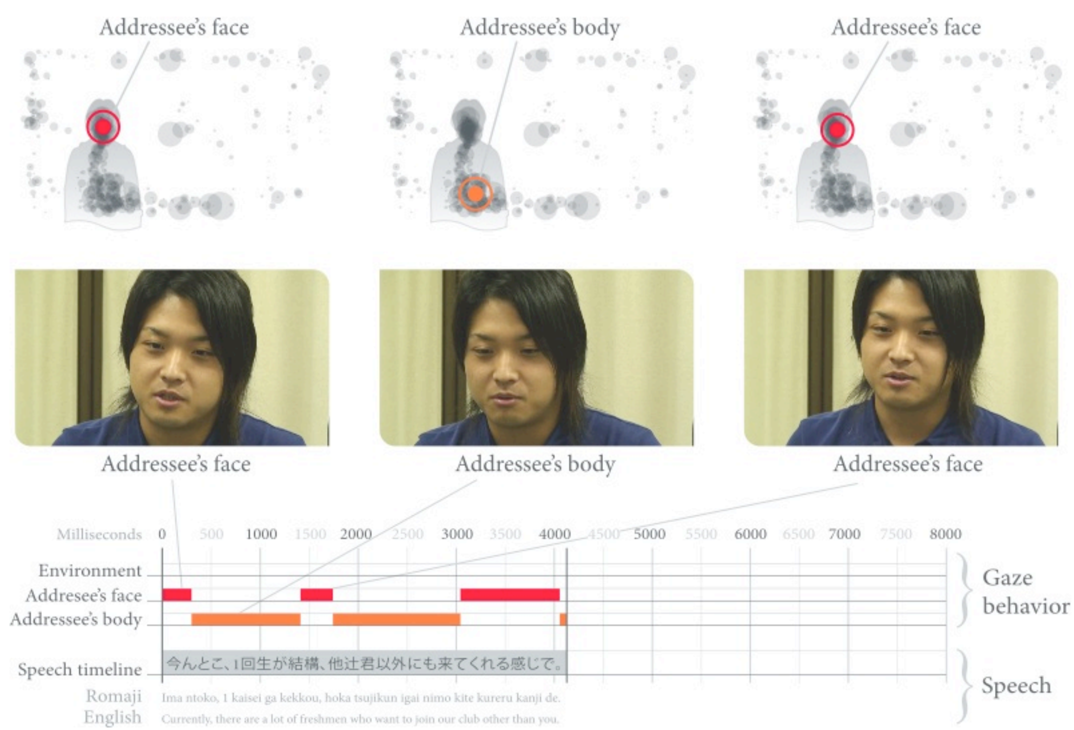
Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M., and Kuzuoka, H. (2008). Precision timing in human-robot interaction: coordination of head movement and utterance. In Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems (Florence, Italy, April 05 - 10, 2008).

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N. and Miyamoto, T. (2006). Responsive robot gaze to interaction partner. In Proceedings of robotics: Science and systems.

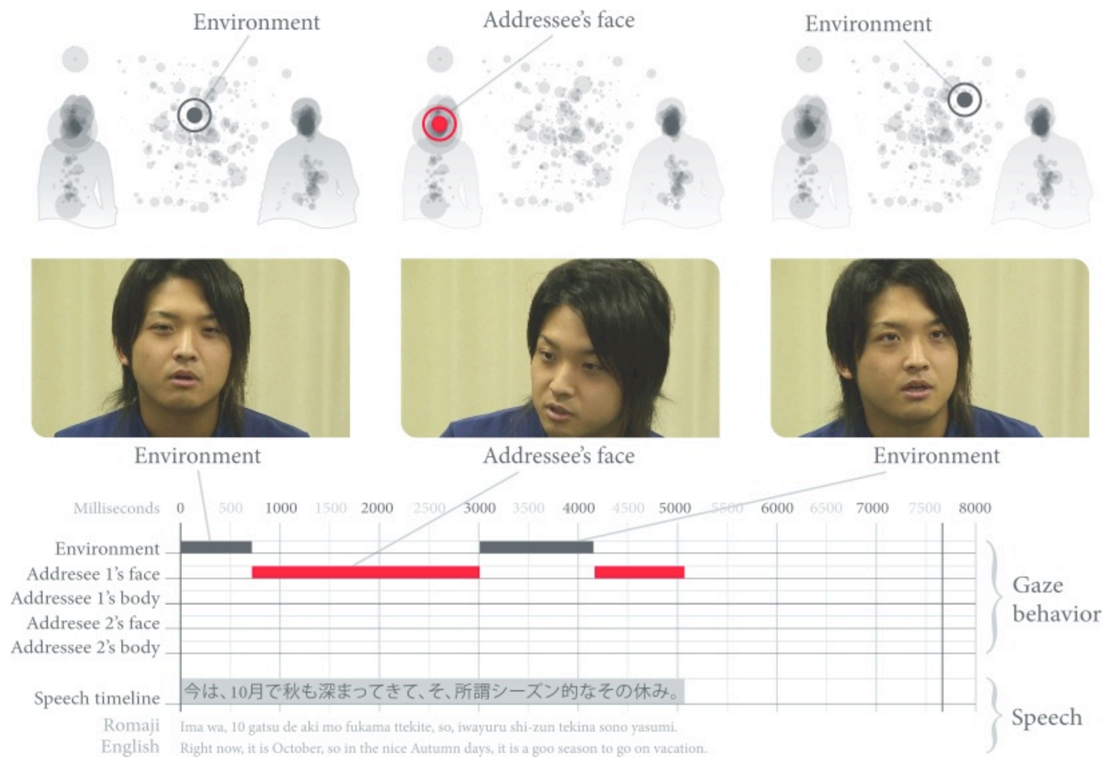
Appendix 1.



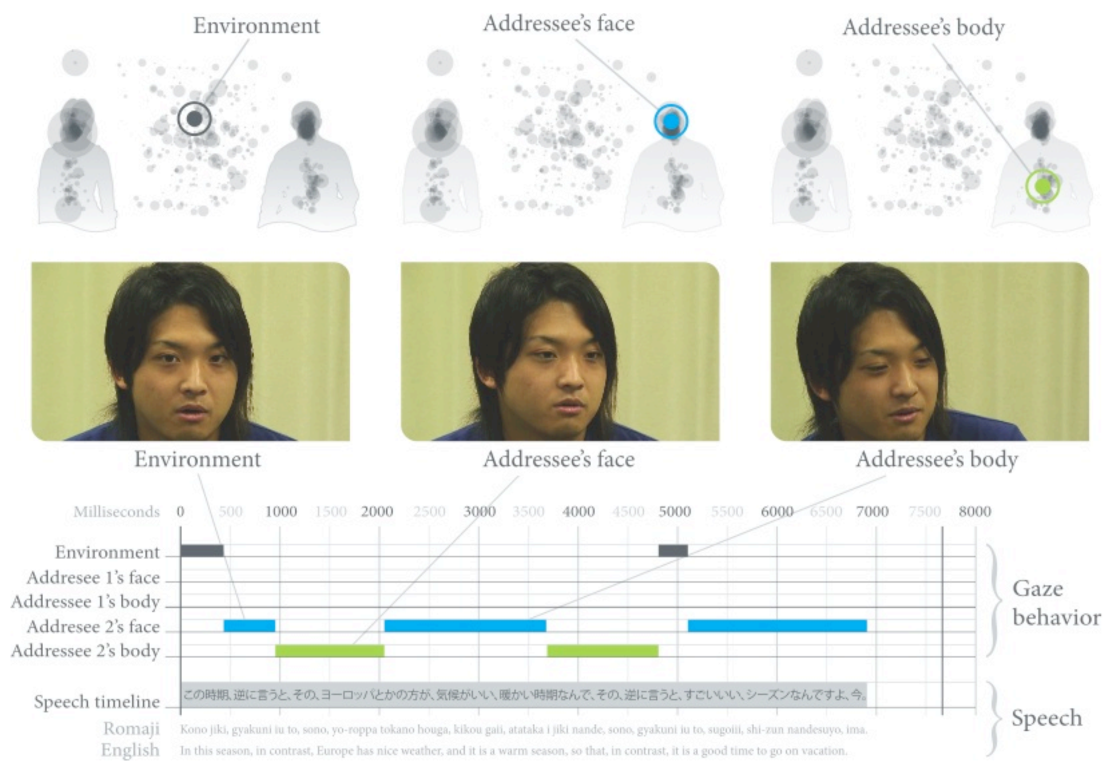
Appendix 1.a. An example of pattern "Environment > Addressee's face > Addressee's body" identified in two-party conversation scenarios.



Appendix 1.b. An example of pattern "Addressee's face > Addressee's body > Addressee's face" identified in two-party conversation scenarios.



Appendix 1.c. An example of pattern "Environment > Addressee's face > Environment" identified in the three-party conversation scenario.



Appendix 1.d. An example of pattern "Environment > Addressee's face > Addressee's body" identified in the three-party conversation scenario.