Conversational Gaze Mechanisms for Humanlike Robots

BILGE MUTLU, University of Wisconsin–Madison
TAKAYUKI KANDA, ATR
JODI FORLIZZI and JESSICA HODGINS, Carnegie Mellon University
HIROSHI ISHIGURO, Osaka University

During conversations, speakers employ a number of verbal and nonverbal mechanisms to establish who participates in the conversation, when, and in what capacity. Gaze cues and mechanisms are particularly instrumental in establishing the participant roles of interlocutors, managing speaker turns, and signaling discourse structure. If humanlike robots are to have fluent conversations with people, they will need to use these gaze mechanisms effectively. The current work investigates people's use of key conversational gaze mechanisms, how they might be designed for and implemented in humanlike robots, and whether these signals effectively shape human-robot conversations. We focus particularly on whether humanlike gaze mechanisms might help robots signal different participant roles, manage turn-exchanges, and shape how interlocutors perceive the robot and the conversation. The evaluation of these mechanisms involved 36 trials of three-party human-robot conversations. In these trials, the robot used gaze mechanisms to signal to its conversational partners their roles either of two addressees, an addressee and a bystander, or an addressee and a nonparticipant. Results showed that participants conformed to these intended roles 97% of the time. Their conversational roles affected their rapport with the robot, feelings of groupness with their conversational partners, and attention to the task.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation / methodology, user-centered design*

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Human-robot interaction, humanlike robots, conversation, gaze, turn-taking, conversational roles, discourse structure

ACM Reference Format:
Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., and Ishiguro, H. 2012. Conversational gaze mechanisms for humanlike robots. ACM Trans. Interact. Intell. Syst. 1, 2, Article 12 (January 2012), 33 pages.
DOI = 10.1145/2070719.2070725 http://doi.acm.org/10.1145/2070719.2070725

12

A less comprehensive version of this manuscript has appeared in *Proceedings of the 4th ACM / IEEE International Conference on Human-Robot Interaction and* was awarded Best Paper.
This research was supported by NSF grant IIS-0624275 and the Ministry of Internal Affairs and Communications of Japan.
Authors' Addresses: B. Mutlu, Department of Computer Sciences, University of Wisconsin–Madison, 1210 W. Dayton St., Madison, WI 53706; email: bilge@cs.wisc.edu; T. Kanda, Intelligent Robotics and Communication Laboratories, ATR, 2-2-2 Hikaridai, Soraku-gun, Seika-cho, Kyoto, 619-0224, Japan; email: kanda@atr.jp; J. Forlizzi, Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA; email: forlizzi@cs.cmu.edu; J. Hodgins, Robotics Institute, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA; email: jkh@cs.cmu.edu; H. Ishiguro, Department of System Innovation, Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama, Toyonaka, Osaka; email: ishiguro@irl.sys.es.osaka-u.ac.jp.

DOI 10.1145/2070719.2070725 http://doi.acm.org/10.1145/2070719.2070725

1. INTRODUCTION

Humanlike robots promise to provide social and intellectual assistance in our daily lives as information booth attendants, museum guides, shopkeepers, storytellers, educational assistants, and companions for children and the elderly. In these positions, robots will need to establish and maintain conversations with the people they are assisting. Achieving fluent conversations will require that robots effectively employ verbal and nonverbal conversational mechanisms that people use in day-to-day interactions. Research in human communication suggests that successful conversations depend on the effective use of gaze cues, which play a particularly important role in establishing the conversational roles of interlocutors [Bales et al. 1951; Schegloff 1968; Goodwin 1981], managing speaker turns [Duncan 1972; Sacks et al. 1974], and signaling topic changes in verbal discourse [Grosz and Sidner 1986; Cassell et al. 1999a; Quek et al. 2000; Quek et al. 2002a].

While research to date has emphasized the role of key gaze cues and mechanisms in conversations and provided clues about their functioning, how these mechanisms might be computationally reconstructed, assembled to function harmoniously, and implemented into humanlike robots to achieve fluent human-robot conversations is unknown. Furthermore, research has not yet explored how these mechanisms might allow humanlike robots to achieve high-level social outcomes such as building rapport with people. The current research seeks to answer the following questions. What are the key gaze mechanisms humanlike robots should use to effectively engage in conversations with people? How can we gain a computational understanding of these mechanisms? How can we integrate these mechanisms into coherent models of gaze behavior for humanlike robots? How might robots use these mechanisms to evoke positive social and cognitive outcomes?

This work seeks to answer these questions through three phases: (1) *computational modeling* of key conversational gaze mechanisms, (2) *interaction design* of these mechanisms for humanlike robots, and (3) *experimental evaluation* of the social outcomes of manipulations in these mechanisms. The computational modeling phase involves formal observations of human conversations and detailed analyses of the gaze cues of the conversational participants, producing hybrid rule-based and stochastic models of gaze behavior. The interaction design phase seeks to combine a set of mechanisms into a coherent conversational behavior that can be implemented in available humanlike robot platforms. Finally, the evaluation phase involves creating human-robot conversation scenarios and manipulations in the designed mechanisms to better understand their conversational and social outcomes.

The remainder of this section outlines the key conversational mechanisms on which this research focuses. The remainder of the article will describe the three research phases above and discuss their findings and limitations.

1.1. Conversational Gaze Mechanisms

In conversations, gaze cues serve a number of functions—from helping a speaker hold the floor [Kendon 1967] to clarifying who is being addressed [Schegloff 1968]. We refer to behavioral cues that serve particular communicative functions as *mechanisms*. Research on conversations highlights three key gaze mechanisms: (1) cues that enable speakers to signal conversational roles, (2) signals that facilitate the taking and passing of speaking turns, and (3) gaze shifts that provide information on the structure of the speaker's discourse. The following paragraphs provide background on these mechanisms.

1.1.1. Role-Signaling Mechanism. Interlocutors of a conversation engage in the ongoing discourse at varying levels of involvement that determine their "participant roles"

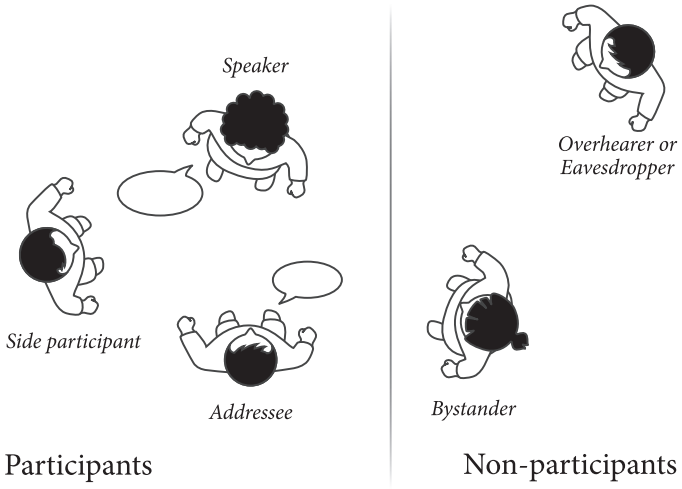

Fig. 1. Levels of conversational participation (adapted from Goffman [1979] and Clark [1996]).

[Goodwin 1981] or "footing" [Goffman 1979] or the "participation structure" of the conversation [Levinson 1988]. These participant roles not only determine how the conversation is organized (who is speaking and when) [Goffman 1979; Clark 1996], but also shape spoken discourse (what is being said) [Hymes 1972; Hanks 1996]. The core conversational roles are those of the *speaker* and the *addressee* [Clark 1996]. Conversations with more than two participants may also involve side participants who are the "unaddressed recipients" of the speech at that moment [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 the participants acknowledge and who observe the conversation without participating in it [Goffman 1979; Clark and Carlson 1982; Clark 1996]. There might also be hearers whose presence the participants do not acknowledge, but who follow 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 these different levels of participation.

While participant roles might be fixed in some social settings (e.g., lectures), most conversational settings allow for the shifting of roles over the course of the conversation. For instance, the role of the speaker or the addressee in a two-party conversation might change at any moment in the conversation [Goffman 1979]. The direction of gaze plays an important role in establishing and exchanging conversational participant roles. In conversations that involve more than two people, the gaze of a speaker toward 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 single 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 whom is being addressed, breakdowns might occur in the organization of the conversation [Schegloff 1968].

1.1.2. Turn-Taking Mechanism. Role shifts among conversational participants are enabled by a *turn-taking* mechanism, which allows interlocutors to seamlessly exchange

speaking turns [Yngve 1970; Goffman 1971]. Research in conversational organization suggests that turn-exchanges follow one of two models: (1) "one part speaks at a time" model in which participants sequentially take turns to share a single floor [Sacks et al. 1974], and (2) a model with more than one floor and "overlapping talk" in which participants collaboratively develop the floor and take turns when functionally appropriate [Edelsky 1981; Schegloff 2000]. Gaze direction serves as a key signal in managing turns in both single-floor turn-exchanges [Nielsen 1962; Duncan 1972; Sacks et al. 1974; Goodwin 1980; 1981] and in 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 and look at their addressees to signal the end of a remark and the passing of the floor to another participant [Nielsen 1962]. In this context, the participant 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 by Kendon [1967]). Exchanging of turns might be delayed when remarks do not end with gazing at another participant [Kendon 1967; Vertegaal et al. 2000].

1.1.3. Topic-Signaling Mechanism. Participants in a conversation form a *discourse*—a composition of discourse segments in particular discourse structures [Grosz and Sidner 1986]. These structures signal shifts in topic in the discourse or how information is organized [Halliday 1967; Brown et al. 1980]. Speakers produce a number of cues that signal these structures to enable contributions from other participants or to direct attention to important information [Goodwin 1981; Whittaker and Stenton 1988; Cassell et al. 1999b]. These signals include verbal, vocal, and nonverbal cues, particularly verbal prompts, repetitions, and summaries [Whittaker and Stenton 1988], discourse markers [Schiffrin 1988; Clark 1996], vocal intonation [Hirschberg and Pierrehumbert 1986; Hirschberg and Grosz 1992], and nonverbal cues such as gaze and gestures [Cassell et al. 1999b; Quek et al. 2002a; Quek et al. 2002b]. Gaze cues serve as a particularly important cue for signaling discourse structure; research on the relationship between speaker gaze shifts and the thematic structure of utterances show that 73% of all gaze shifts toward the addressee correspond with the shifts in the thematic structure toward providing new information [Cassell et al. 1999b]. In this context, gaze might serve as a mechanism to draw the addressee's attention to important new information in spoken discourse.

Research on the conversational role of gaze cues suggests that these three gaze mechanisms are inseparable. For instance, the proper functioning of the role-signaling mechanism depends on the functions afforded by a turn-taking mechanism [Bales et al. 1951; Schegloff 1968; Bales 1970]. Similarly, speaker gaze cues that signal discourse structure, particularly looking toward an addressee while delivering new information, might also signal the availability of the floor for the next speaker [Cassell et al. 1999b]. Furthermore, human conversations involve several other mechanisms that serve specific functions in specific conversational contexts. For instance, gaze cues might help disambiguation when the conversation involves referring to information, artifacts, or other people in the environment [Hanna and Brennan 2007]. These mechanisms are not considered in the work described here.

Existing knowledge on conversational gaze mechanisms have been primarily based on English conversation and our understanding of these mechanisms in Japanese conversations is sparse. While not on gaze, comparisons of linguistic and other conversational mechanisms such as turn-taking, simultaneous talk, and backchannels indicate that these mechanisms are employed in both English and Japanese conversations but appear at significantly different rates [Maynard 1986; Hayashi 1988; Tanaka 1999]. Therefore, we expect that these conversational gaze mechanisms generalize across the

two languages in functioning but vary in how much and under what circumstances interlocutors employ them.

1.2. Related Work on Conversational Gaze Mechanisms

Conversational gaze mechanisms have been explored by different research communities in order to design embodied conversational agents, build immersive virtual environments, and create embodied social cues for humanlike robots. The discussion below provides an overview of related work in the embodied conversational agents, immersive virtual environments, and human-robot interaction communities.

1.2.1. Embodied Conversational Agents. The majority of the prior research on designing conversational gaze mechanisms has been done with the goal of designing embodied conversational agents and intelligent virtual agents [Cassell et al. 1994; Garau et al. 2001; Thórisson 2002; Rehm and André 2005; Heylen et al. 2005]. Researchers in this area have explored how verbal and nonverbal cues might facilitate a number of conversational functions including turn-taking, feedback, repair, and synchronized speech [Cassell et al. 1994; Vilhjálmsson and Cassell 1998; Cassell et al. 1999a]. They have built systems that combined nonverbal cues such as gaze, gestures, facial expressions, and postural shifts [Cassell et al. 1999a; Cassell et al. 2001; Lee et al. 2002]. These systems primarily used gaze cues to achieve conversational functions such as establishing participant roles, facilitating turn-exchanges, and signaling discourse structure [Cassell et al. 1999a; Vertegaal et al. 2001]. Evaluations of these systems showed that by employing signals that resemble human gaze cues (as opposed to randomly generated signals), conversational agents can achieve more efficient conversations, increase task performance, and improve people's evaluations of them [Colburn et al. 2000; Garau et al. 2001; Lee et al. 2002; Heylen et al. 2005]. Using these signals effectively, agents can also improve how much people conform to the intended participant structure of a conversation [Rehm and André 2005].

1.2.2. Immersive Virtual Environments. Conversational gaze cues have also been explored in the context of building immersive virtual environments (IVEs), particularly to improve the communicative effectiveness of avatars toward gaining positive social and cognitive outcomes [Garau et al. 2003; Bailenson et al. 2005; Murray et al. 2007; Steptoe et al. 2008]. These studies have shown that when the gaze cues of an avatar are designed to match the conversational role of the avatar as opposed to random gaze, people perceive the avatar to have a stronger social presence and conversational involvement and evaluate the avatar and the overall quality of the interaction more positively [Garau et al. 2003]. Studies in this area have also explored how speaker gaze cues might be "augmented" to create the impression in two listeners that they are being addressed simultaneously [Bailenson et al. 2005]. This work compared participants' evaluations of the speaker across augmented and typical gaze conditions and found that augmented gaze improved the overall agreement with the speaker's message.

1.2.3. Human-Robot Interaction. A more recent but growing number of studies in human-robot interaction have explored conversational gaze cues in the context of designing humanlike robots that show appropriate social behavior [Sidner et al. 2004; Mutlu et al. 2006; Bennewitz et al. 2006; Kuno et al. 2007; Trafton et al. 2008; Yamazaki et al. 2008; Yamazaki et al. 2010; Staudte and Crocker 2011]. These studies show that robots elicit more appropriate responses from conversational partners when they display appropriately timed gaze cues such as looking toward their addressees at "turn-relevant places" [Sacks et al. 1974] during turn-exchanges as opposed to producing random

gaze shifts [Bennewitz et al. 2006; Kuno et al. 2007; Yamazaki et al. 2008; Yamazaki et al. 2010]. Such appropriate timing of gaze cues also improves how conversational partners evaluate robots [Trafton et al. 2008]. The appropriate direction of gaze cues also affects conversational outcomes in human-robot interaction; robots can facilitate comprehension when their gaze direction is congruent with their verbal reference and disrupt comprehension when the information from their gaze and speech are incongruent [Staudte and Crocker 2011]. Finally, robots interacting with large groups can more accurately indicate the target of a hand-over by directing gaze toward the target participant [Kirchner et al. 2011].

These studies provide considerable evidence that displaying appropriate humanlike gaze behavior improves people's perceptions of the conversational effectiveness of artificial agents, including embodied conversational agents, avatars in immersive virtual environments, and humanlike robots. However, how robots might use specific gaze mechanisms to achieve more fluent conversations is still unknown. The next section describes our process of modeling three conversational mechanisms that are key to achieving fluent conversational functioning.

2. PHASE I: MODELING OF CONVERSATIONAL GAZE MECHANISMS

A central problem in the study of social behavior for interactive computer interfaces—including humanlike robots—is determining how valid communicative signals that people recognize and reciprocate might be designed for these interfaces. Approaches to this problem involve designing behaviors that follow a set of principles developed and refined over time (e.g., Thomas and Johnston [1995]) or metaphors that scaffold the interaction (e.g., Laurel [1991]). A promising approach is to guide the design of behavioral cues using validated theories of human communication, such as the use of *Politeness Theory* [Brown and Levinson 1987] in the design of the conversational characteristics of pedagogical agents [Wang et al. 2005; Wang and Johnson 2008]. In human-robot interaction, previous studies have drawn on theory on conversational turn-taking [Sacks et al. 1974] to design the conversational gaze behavior of a robot [Kuno et al. 2007]. However, due to their explanatory character, these theories fail to provide all the specifications needed to recreate behavioral mechanisms and serve only as scaffolding for design.

The current study further formalizes this approach into a design process guided by theories of human communication and analyses of empirical observations of social situations in the design of behavioral mechanisms for humanlike robots. Below we describe the theoretical scaffolding and empirical specification of the conversational gaze mechanisms that the current study investigates.

2.1. Participation

Research in nonverbal behavior reports strong effects of group composition on both the production and the perception of gaze, particularly of gender [Exline 1963; Argyle and Dean 1965; Argyle and Ingham 1972] and age [Efran 1968; Libby 1970]. Our previous work also found gender effects on how the robot's gaze affected people's performances and their perceptions of the robot [Mutlu et al. 2006]. One of the limitations of this work was that we used observations of a female speaker in an all-female triad to design the gaze behavior of the robot and evaluated the designed gaze behavior with a mixed-gender population. We suspect that our results might have been affected by gender-based differences in the production and perception of gaze behavior. Therefore, we decided to control for these group composition effects, testing our hypotheses in a smaller population of male native-Japanese-speaking college students from the Osaka area of Japan between the ages of 18 and 24. We recruited four all-male triads

(12 subjects[1]) to perform the tasks in the modeling study. Accordingly, we also limited our subject profile for the observation to an all-male triad and chose a male experimenter to administer the study. Subjects were paid ¥1,500 (roughly $14 or €9) for their participation in the study.

2.2. Setup, Task, and Procedure

A broad survey of discourse theory suggests that gaze cues are strongly associated with the structure of the discourse, particularly with information, conversation, and participation structures [Bales et al. 1951; Kendon 1967; Schegloff 1968; Duncan 1972; Sacks et al. 1974; Goodwin 1981; Cassell et al. 1999b] and that they might serve as three key conversational mechanisms.

Information Structure and Topic-Signaling Mechanism. Patterns in gaze shifts temporally aligned with the structure of the speaker's discourse, particularly changes in topic or shifts between thematic units of utterances [Cassell et al. 1999b].

Conversation Structure and Turn-Taking Mechanism. Speaker gaze cues that facilitate turn-exchanges (producing turn-yielding, turn-taking, and floor-holding gaze signals) [Kendon 1967; Duncan 1972; Sacks et al. 1974; Goodwin 1981].

Participation Structure and Role-Signaling Mechanism. Speaker gaze cues that signal the roles of interlocutors [Bales et al. 1951; Schegloff 1968; Goodwin 1981].

To limit the scope of our investigation, we focused on the following three key participant roles described by Goffman [1979], Levinson [1988], and Clark [1996]: *addressees*, *bystanders*, and *overhearers*. In the context of this research, addressees include participants who take speaking turns to contribute to the conversation and to whom the speaker addresses while speaking. Bystanders are acknowledged nonparticipants who do not take speaking turns and whom the speaker does not address while speaking, but whose presence is acknowledged during the conversation, particularly during greetings and leave-taking. Overhearers involve unacknowledged nonparticipants who do not take speaking turns, whom the speaker does not address while speaking, and whose presence is not acknowledged at any point in the conversation. Here, it is important to note that we chose the role of overhearer to refer to the general category of unacknowledged non-participants for purposes of consistency. In the context of our study, this role is considered interchangeable with an *eavesdropper* or *ignored participant*.

To further specify the communicative mechanisms described in the previous section, we conducted formal observations guided by existing theory on conversational organization. In these observations, we provided naive participants with conversational scenarios involving different role structures, asked them to converse according to these roles, and observed how speakers used gaze cues to signal these roles. We created three role structures: (1) *two-party conversation* with a speaker, an addressee, and an overhearer, (2) *two-party-with-bystander conversation* with a speaker, an addressee, and a bystander, and (3) *three-party conversation* with a speaker and two addressees (illustrated in Figure 2).

The study procedure was as follows: Before their participation, all subjects 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

[1]To distinguish conversation participants (those who participate in a conversation by taking speaking turns) from study participants (those whom we recruited to participate in our investigation), the latter will hereafter be referred to as "subjects."

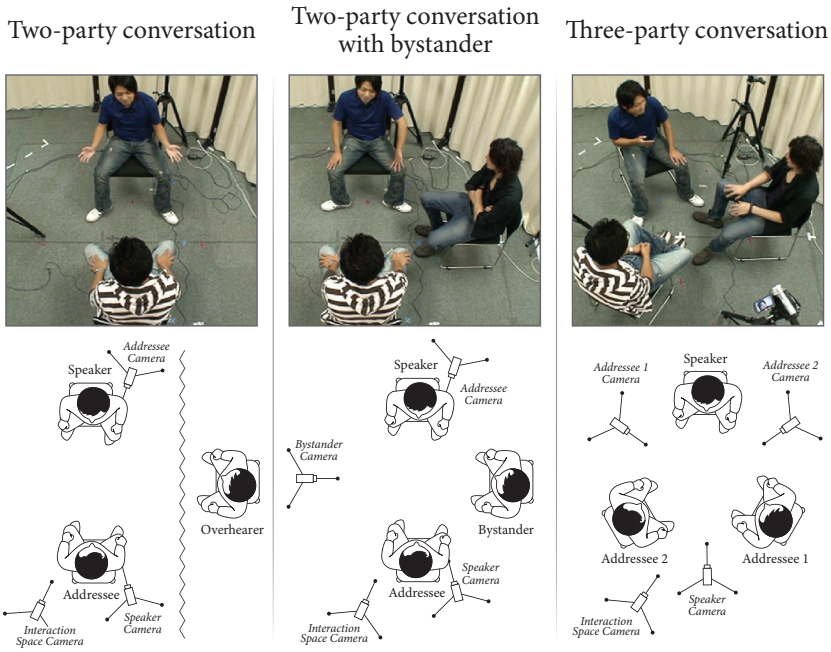

Fig. 2. The data collection setup for the three conversational structures studied: (top) a two-party conversation, (middle) a two-party conversation with a bystander, and (bottom) a three-party conversation.

[Goldberg et al. 2006]. All triads performed all three scenarios for fifteen minutes in the order listed above. At the beginning of each scenario, subjects were provided with a description of the scenario and their roles. They were first given five minutes to ask questions and adapt to their roles. Between each pair of scenarios, subjects were asked to take ten-minute-long breaks and solve crossword puzzles in order to distract them from their roles in the previous scenario. Subjects were compensated at end of their participation. The study procedure was reviewed and approved by the ATR Institutional Review Board.

2.3. Measurement and Analysis

We captured subjects' gaze behavior using high-definition cameras placed across from their seats. Subjects' speech was captured using stereo microphones attached to their collars. The cameras provided video sequences of subjects' faces (from hair to chin). We placed an additional camera on the ceiling to capture the interaction space. For both practical and ethical reasons, cameras remained visible to the study subjects. In total, we captured approximately 45 minutes of video for each subject and 180 minutes of data for each triad from four cameras.

An important limitation of modeling the behavior of naive subjects role-playing in predetermined scenarios is the low fluency of interaction led by factors such as certain aspects of personality (such as introversion), experimenter effects (performance anxiety), and unfamiliarity with the topic of the scenario and with conversational partners. In order to base our inquiry on data that is minimally affected by these factors, we used data from the triad that exhibited the most fluent interaction for our detailed analysis (determined based on the total amount of pauses in the conversation that was calculated from the videos and subjects' levels of extroversion/introversion that were measured through questionnaires). To limit our focus to how speakers use gaze cues

to signal participant roles, we analyzed only the speaker's gaze behavior. However, because certain conversational mechanisms such as turn-taking necessarily involve reciprocity, we also analyzed addressee gaze and speech, but only at turn-exchanges. The final analysis included an close examination of 45 minutes of video data from a single speaker and segments of the video data from each addressee or bystander at turn-exchanges. For purposes of simplicity, interruptions and backchannel responses—short utterances, such as "uh-huh," "yeah," and "okay," produced by one conversational participant while the other is talking [Ward and Tsukahara 2000]—were omitted in the analysis.

The analysis of the video data involved data coding and descriptive statistics, in particular coding of speech and gaze events from the video, calculating frequencies of and co-occurrences among these events, and computing the distribution parameters for the temporal and spatial properties of these events.

2.4. Results

Our modeling process sought to capture the basic spatial and temporal parameters of gaze cues and aspects of conversational mechanisms that signal information, conversation, and participation structures. In this subsection, we describe the results of our analysis for each mechanism and elements of our model.

2.4.1. Spatial and Temporal Parameters of Gaze Cues. To identify the spatial and temporal characteristics, our analysis of the speaker's gaze behavior tried to answer the following questions. Where do speakers look in different conversational structures? How much time do they spend looking at these targets? The answers that our analysis provided directly informed our design of the basic spatial and temporal parameters of the robot's gaze behavior. In the following, we describe the results of the analysis and elements of the design.

Where Do Speakers Look?. To understand where our speaker looked while speaking, we conducted a frame-by-frame analysis of his gaze behavior. We coded the target and time of execution of each gaze shift. The coding was done by qualitatively estimating the speaker's approximate gaze targets and marking them on an image representation of the speaker's field of view. We clustered the final set of gaze targets both qualitatively and quantitatively. In the qualitative analysis, we identified three clusters in the first scenario and five clusters in the second and third scenarios through visual inspection. We then used a Gaussian Mixture Model (GMM) estimation algorithm [Bouman 1997] to quantitatively determine the number of clusters and identify the cluster to which each gaze target belonged. The quantitative analysis confirmed the numbers of clusters in our qualitative assessment of the first and third scenarios. However, the clustering algorithm identified eight clusters in the second scenario, four of which included gaze shifts away from conversational partners and towards the environment. We combined these four clusters to make up a single cluster, as they all defined areas in the environment instead of conversationally significant gaze targets such as an addressee's face. Figure 3 illustrate gaze targets and the identified clusters.

We conducted an inter-coder reliability analysis to ensure the objectivity of our coder's analysis of the speaker's gaze direction. We asked both the primary and the secondary coder to categorize a randomly selected sample from the video data (90 clips with single gaze shifts, 30 for each conversational scenario) into the clusters identified by our qualitative and quantitative assessment. Cohen's Kappa (κ) was calculated to measure the agreement between the two raters, producing substantial agreement ($Cohen's\ \kappa = .78$).

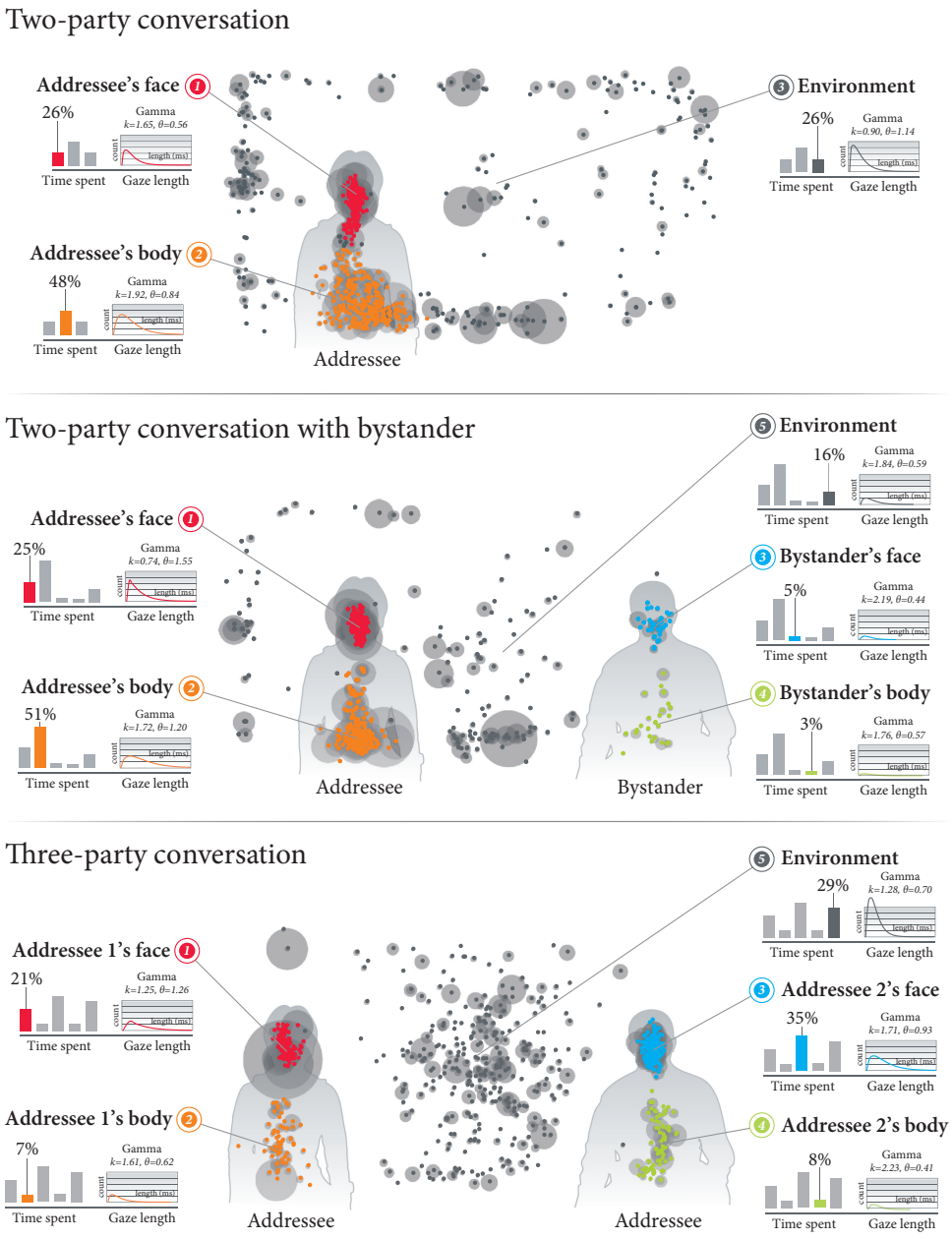

Fig. 3. The gaze clusters that our analysis identified in the two-party conversation (top), two-party conversation with bystander (middle), and three-party conversation (bottom). The illustration includes total amount of gaze directed toward each cluster and fitted Gamma distribution parameters for gaze durations.

How Much Time Do Speakers Spend Looking at Each Target?. In all scenarios, the speaker looked at his addressees for the majority of the time; 74%, 76%, and 71% in the two-party, two-party-with-bystander, and three-party scenarios respectively. However, in the first two scenarios, the speaker looked at the bodies of his addressees more

than he looked at their faces (26% and 25% at the faces and 48% and 51% at the bodies). We speculate that this behavior is a form of intimacy regulation—to reduce the arousal increased by establishing eye contact with a conversational partner as suggested by Argyle and Dean [1965]. We also observed that the gaze toward the body often followed shifts towards the face, providing further support for the intimacy-regulation explanation. This behavior was less prominent in the third scenario (56% at the faces of the two addressees and 15% at their bodies), perhaps because shifting gaze to look towards another interlocutor serves as an alternative way of regulating intimacy (similar results were obtained by Vertegaal et al. [2001]). Gaze breaking (i.e., avoiding eye contact) was also a common behavior that we observed in our data. In all scenarios, the speaker spent a significant amount of time looking away from addressees (26%, 16%, and 29% of the time in the three conversational situations respectively).

Our analysis of the duration of gaze toward each target showed that these durations closely resembled exponential distributions with varying parameters for each cluster. We obtained estimates for these parameters by fitting two-parameter continuous Gamma distributions (defined by parameters θ for shape and k for scale) to the data.[2] Figure 3 provides the fitted distribution parameters values for each gaze cluster.

2.4.2. Topic-Signaling Gaze Mechanism. Research in computational linguistics suggests that information structure—the relationship between the context of utterances or clauses and the emerging discourse context—accounts for a large portion of speaker gaze shifts within the course of a turn [Cassell et al. 1999b]. Based on the "information structure" suggested by Halliday [1967], utterances in English consist of two segments: the "theme" and "rheme." The theme (also termed "link" or "given") connects the utterance to the previous discourse and provides context for the new information to be presented. The rheme represents the new information that hearers could not have predicted from the context of the previous discourse. Cassell et al. [1999b] found that speakers mostly looked away from their conversational partners at the beginning of a theme and mostly looked toward their partners at the theme-rheme junction.

We followed a similar approach to model the relationship between the speaker's gaze shifts and the information structure of his speech. We started this process by using discourse theory in Japanese to identify the most appropriate unit of analysis for our speaker's speech. The conversational style of this data can be characterized as a "casual narrative" [Tannen 2005] in which the speaker holds the floor for longer periods. In these sequences of narrative segments, 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. Following Maynard's description, we unitized our speech data into thematic fields, producing a total of 181, 146, and 155 units in our three conversational scenarios respectively. Here is the English translation

[2]A distribution that better fits the gaze data was necessary, as gaze literature only reports mean and standard deviations for fixation durations and, without information on the data distribution, it is impossible to synthesize fixations in the robot. For instance, when the distributions shown in Figure 3 are considered, drawing samples from a normal distribution with a given set of mean and standard deviation values can generate negative duration values.

of an example segment of speech split into two thematic fields with the identified topic shift marker.

> *Speaker:* Our club, uh... basically does many activities, uh... for instance, we do sports activities often *[topic switch]* Once a month... once a month is maybe too much. Every two months, we do an indoor soccer tournament and sometimes we play softball.

Using a data visualization tool that we developed for this analysis, we mapped each thematic field onto the speech timeline along with gaze shifts that took place within the thematic field and 4000-millisecond periods before the beginning and after the end of the thematic field. This mapping allowed us to identify patterns in gaze shifts that occurred at the onset of each thematic field and quantify the frequency of occurrence for each pattern. The analysis identified two main recurring 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 4 illustrates these patterns for two- and three-party conversations. Table I shows the frequencies of occurrence for each pattern.

2.4.3. Turn-Taking Gaze Mechanism. Research in conversational organization and nonverbal behavior has shown that gaze behavior is also instrumental in managing turns in conversations and follows a common pattern [Kendon 1967; Duncan 1972; Sacks et al. 1974; Goodwin 1981]. To identify how gaze cues facilitated turn-exchanges, we identified the "turn-relevant places" in our speech data based on the set of rules suggested by Sacks et al. [1974] and analyzed speaker's gaze direction at these moments. In our analysis, we focused on turns that the speaker initiated with an explicit "turn-yielding" signal (as described by Duncan [1972]), omitting interruptions and simultaneous turns. Our analysis showed that 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 our three conversational scenarios respectively) involved the following three turn-management signals (also proposed by Kendon [1967] and Duncan [1972] for conversations in English):

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

Turn-taking. The addressee looks at the speaker at the end of the speaker's turn signaling to him that he is open to taking the floor.

Floor-holding. As the new speaker takes the turn, he looks away from his partner, signaling that he is keeping the floor until his turn is complete.

The majority of these turns showed the structure of "question-answer pairs" [Clark 1992], a specific prototypical case of "adjacency pairs" observed in conversations [Schegloff and Sacks 1973]. In these sequences, the speaker (1) produces a turn-yielding signal at the end of a question, (2) looks continuously at his partner during the response, (3) when his partner's response is complete, he produces a "minimal response" [McLaughlin and Cody 1982] such as an acknowledgement, mirror response, or laughter, during which he looks at his partner, and finally (4) produces a floor-holding signal by looking away from his partner when he starts his discourse. Figure 5 illustrates the speaker's turn-yielding, turn-taking, and floor-holding gaze signals during a question-answer pair that we observed in our data.

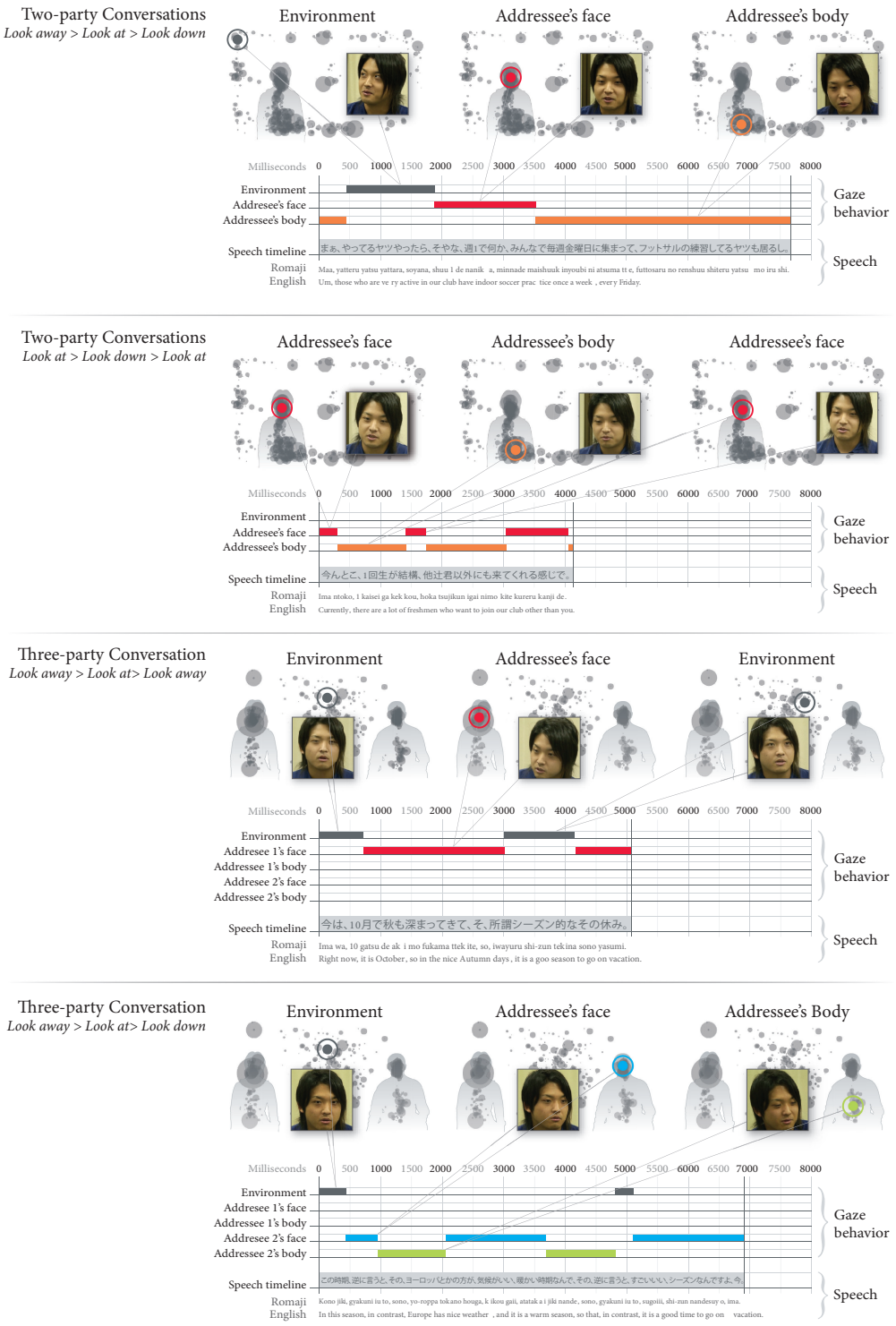

Fig. 4. Examples of the two most recurring patterns in the two- and three-party conversations.

Table I.
The frequencies of the topic-signaling patterns identified in the two- and three-party conversations. Data from two-party and two-party-with-bystander conversations are combined, because the patterns varied minimally across the two conversational structures with the exception of short, occasional glances toward the bystander.

Topic-Signaling Gaze Pattern	Two-party conversations	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 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

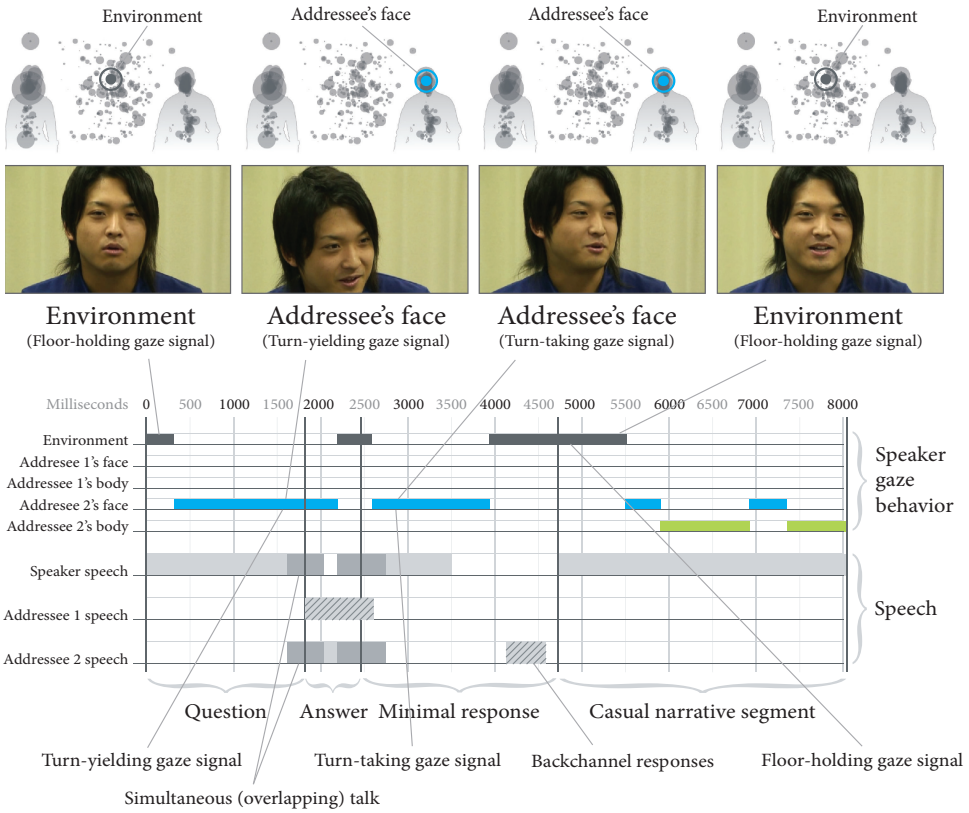

Fig. 5. The turn-exchange signals the speaker produced at one of the question-answer pairs.

2.4.4. Role-Signaling Gaze Mechanism. In our data, we identified three gaze cues that our speaker produced to signal the participant roles of his interlocutors. These cues are described below and are categorized based on where in the conversation they occurred.

Greetings and summonses. An important point where speakers signal the roles of their conversational partners (and others signal their availability for these roles) is the opening of a conversation such as greetings where one welcomes and acknowledges another or summonses where one attracts the attention of another to start a conversation. 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." Speakers rely primarily on gaze cues to signal these roles [Bales et al. 1951; Bales 1970]. Schegloff [1968] depicts an observation where the lack of gaze cues during a summons leads to ambiguity in who is being addressed in a crowd of bus-riders. In our observation, the speaker greeted and directed his gaze towards individuals in the roles of both addressee and bystander. However, in the second conversational scenario, at the point of transition from greetings to the body of the conversation, the speaker diverted his gaze towards the addressee and away from the bystander, providing a significant cue for their participant roles.

The body of the conversation. In our observation, the speaker spent the majority of his speaking time looking at addressees. In our first conversational scenario, he looked towards the addressee 74% of the time and the environment 26% of the time. In the second scenario, the speaker allocated some of his gaze towards the bystander (8%), mostly in short acknowledging glances averaging nearly half the average length of the gazes at his addressee. He looked towards the addressee, bystander, and the environment 76%, 8%, and 16% of the time, respectively. Finally, in the last scenario, the speaker looked towards his addressees and the environment at 71% and 29% of the time. The bottom row in Table II provides the length distribution parameters for each cluster of gaze targets.

Turn-exchanges. Another important point in conversations where participant roles are renegotiated is the turn-exchanges. For instance, Weisbrod [1965] (as described by Kendon [1967]) observed in seven-party conversations that the person at whom the speaker looked at the end of a turn was more likely to take the next speaking turn. In our observation, addressees received all turn-yielding gaze signals and bystanders received none, suggesting that the turn-yielding gaze cues are also important signals for establishing the participation structure of a conversation. 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, when his questions were directed at both addressees and were answered by the other addressee, or at points of significant shift in the topic of the conversation.

The findings of the modeling phase of our investigation are summarized in Table II. These findings directly guided our building of robot gaze mechanisms. In the next section, we describe the implementation of these gaze cues and mechanisms on a humanlike robot.

3. PHASE 2: BUILDING GAZE MECHANISMS FOR HUMANLIKE ROBOTS

The next step in our investigation involved building a coherent conversational system using the design elements that we extracted in the modeling phase of our process and implementing this system on a robotic platform that embodied the physical characteristics required by our models by converting the variables described in Table II into an algorithm that controlled the robot. We chose to use ATR's Robovie R-2 robot

Table II.
A summary of our models of role-signaling, turn-taking, and topic-signaling gaze mechanisms and the basic spatial and temporal parameters of gaze.

Conversational Structure *Signaling Gaze Mechanism*	**Two-party conversation**	**Two-party conversation with bystander**	**Three-party conversation**
Participant Structure *Role-Signaling Mechanism*	**At greeting and leave-taking:** Acknowledge the *addressee* **At the transition from greeting to casual conversation:** Maintain gaze toward the *addressee* **During the conversation:** Gaze only toward the *addressee*	**At greeting and leave-taking:** Acknowledge the *addressee* and then the *bystander* **At the transition from greeting to casual conversation:** Gaze toward the *addressee* **During the conversation:** Gaze mostly toward the *addressee,* occasionally glancing toward the *bystander* for short periods	**At greeting and leave-taking:** Acknowledge *one of the addressees* and then *the other addressee* **At the transition from greeting to casual conversation:** Gaze toward either *addressee*, producing turn-yielding signals for *both addressees,* and wait for one of the them to take the floor **During the conversation:** Gaze toward *one addressee* at a time and switch speakers at "paragraphs"
Conversation Structure *Turn-Taking Mechanism*	**Turn-yielding:** Look toward the *addressee* at floor endings **Turn-taking:** Look toward the *addressee* during minimal responses and look away from the *addressee* at floor beginnings	**Turn-yielding:** Look toward the *addressee* at floor endings **Turn-taking:** Look toward the *addressee* during minimal responses and look away from the *addressee* at floor beginnings	**Turn-yielding:** Look toward *one of the addressees* at floor endings **Turn-yielding & speaker change:** Look toward *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 floor
Information Structure *Topic-Signaling Mechanism*	**At thematic field beginnings:** Follow patterns "Look away > Look at > Look down" or "Look down > Look at > Look down"	**At thematic field beginnings:** Follow patterns "Look away > Look at > Look down" or "Look down > Look at > Look down" **At random intervals:** Short glances toward the *bystander*	**At thematic field beginnings:** Follow pattern "Look away > Look at > Look away" or "Look away > Look at > Look down"
Basic Spatial & Temporal Gaze Parameters **Cluster Name** *[Mean (secs); St Dev (secs); Scale (k); Shape (θ)]*	**Addressee's face** *[0.95; 0.91; 1.65; 0.56]* **Addressee's body** *[0.99; 1.03; 1.92; 0.84]* **Environment** *[1.01; 0.98; 0.90; 1.14]*	**Addressee's face** *[1.40; 1.30; 0.74; 1.55]* **Addressee's body** *[1.01; 1.22; 1.72; 1.20]* **Bystander's face** *[0.77; 0.58; 2.19; 0.44]* **Bystander's body** *[0.71; 0.49; 1.76; 0.57]* **Environment** *[0.96; 1.04; 1.84; 0.59]*	**First addressee's face** *[0.98; 1.26; 1.25; 1.26]* **First addressee's body** *[0.80; 0.83; 1.61; 0.62]* **Second addressee's face** *[0.97; 0.83; 1.71; 0.93]* **Second addressee's body** *[0.85; 0.79; 2.23; 0.41]* **Environment** *[0.82; 0.78; 1.28; 0.70]*

Fig. 6. Robovie R-1, the robotic platform we used in the implementation and evaluation of the gaze mechanisms studied in this work.

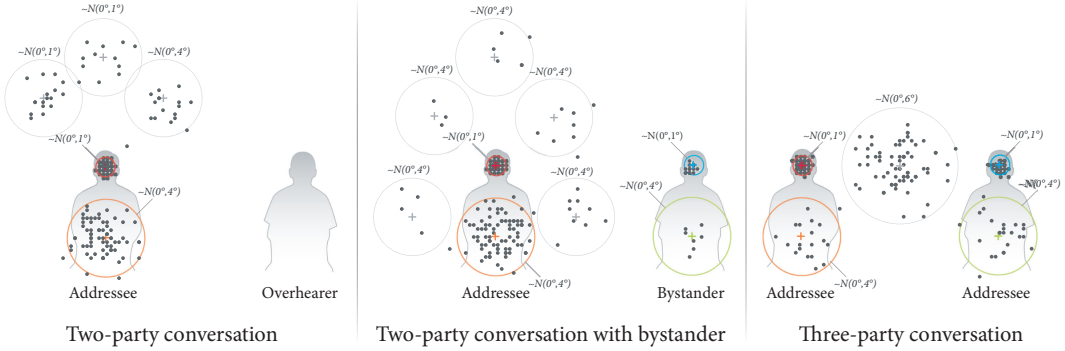

Fig. 7. The gaze targets that the robot produced based on our implementation of the temporal and spatial gaze parameters.

[Ishiguro et al. 2001] for our implementation and evaluation, because its articulate head and eyes allowed us to implement our gaze models and provided salient gaze signals for experimental evaluation. Figure 6 shows the robot used in this work.

3.1. Simulating Spatial and Temporal Gaze Parameters

The results from our analysis of the spatial and temporal parameters of the gaze cues directly informed the design of where the robot should look in the three conversational situations that are described above. The gaze clusters were represented as two-dimensional Normal distributions defined by their centers and spreads and in gaze rotation angles (in degrees). The environment clusters in the two-party and two-party-with-bystander conversations were made up of three and five clusters, respectively, to achieve gaze target distributions similar to those observed in the human data. The exact target of each gaze shift was randomly generated using the parameters of the fitted Normal distributions. We used the gaze duration distribution parameters that we calculated for each cluster to determine how long the robot should spend looking at each target. Figure 7 shows the gaze targets generated by the robot for the three conversational situations.

We divided the gaze shifts of the robot 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. Also, we gave each eye the

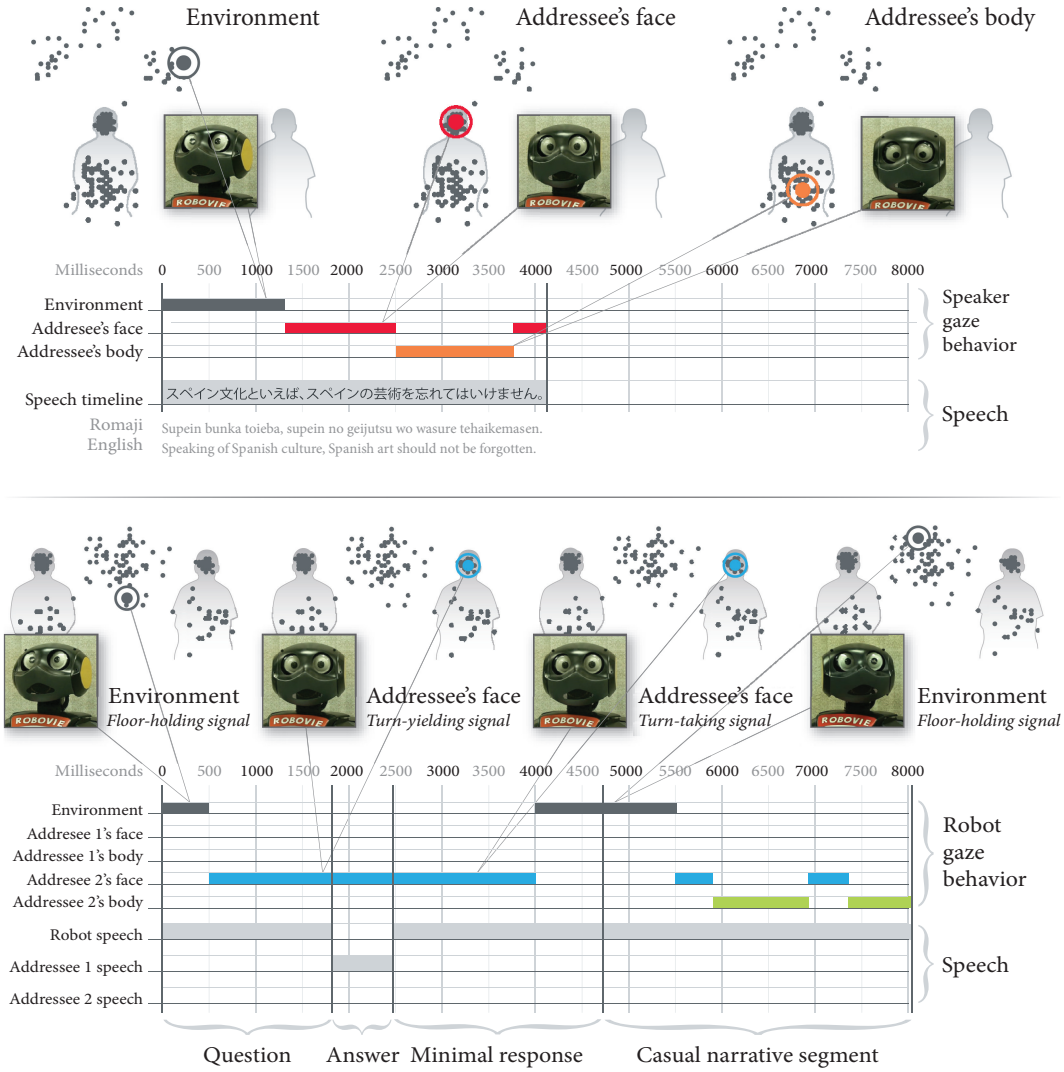

Fig. 8. The most frequent topic-signaling gaze pattern (top) and the turn-taking mechanism (bottom) that the robot simulated based on our implementation.

appropriate horizontal angle for convergence (e.g., 1.5-degrees while looking towards a conversational partner at a distance of two meters).

3.2. Topic-Signaling Mechanism

We designed the patterns that we identified as signaling information structure into the robot's gaze behavior and marked the robot's speech for thematic fields. For each new thematic field, the robot produced the appropriate series of gaze shifts based on the gaze patterns identified for each conversational scenario, probabilities of occurrence for gaze each pattern, and length distribution parameters for each gaze shift provided in Table II. The top illustration in Figure 8 shows the robot simulating one of the patterns identified in two-party conversations.

3.3. Turn-Taking Mechanism

We also marked the robot's speech for turn-relevant places. During turn-exchanges, the robot looked at its addressee at the end of a question, producing a *turn-yielding* gaze signal. Towards the end of its partner's response, it looked at its partner, producing a *turn-taking* signal. Finally, when it took the turn, it looked away from its partner, producing a *floor-holding* signal. During minimal responses (e.g., an acknowledgement, mirror response, or laughter), the robot looked at its addressee as our speaker did in question-answer pairs. The bottom illustration in Figure 8 shows the robot producing these turn-exchange signals in a three-party conversation.

3.4. Role-Signaling Mechanism

We designed the robot's gaze behavior to adapt to the three conversational scenarios that we studied in the modeling phase of our research. In the two-party conversation, the robot acknowledged its addressee during greeting and leave-taking, spent most of its time looking at the addressee's face or body following the patterns that we identified in our analysis of the two-party conversation, and producing turn-yielding signals for the addressee. In the two-party conversation with the bystander, in addition to the behaviors it produced in the previous conversational scenario, it greeted the bystander at the beginning and at the end of the interaction and reaffirmed the bystander's role with short glances directed at him at random intervals during the body of the conversation. Finally, in the three-party conversation, the robot greeted both addressees during greeting and leave-taking and looked at both of them during the conversation following the patterns that we identified in the three-party conversation, producing turn-yielding signals for both partners.

In the next section, we describe our experimental evaluation of how the designed gaze cues might shape participant roles in human-robot conversations.

4. PHASE 3: EXPERIMENTAL EVALUATION OF ROBOT GAZE MECHANISMS

The last phase of this investigation involved an experimental evaluation of the effectiveness of the designed conversational mechanisms in maintaining fluent human-robot conversations in different participation structures and the social and cognitive effects of participating in these conversations on human participants. More specifically, the evaluation sought to find answers to the following questions: Can a robot use humanlike gaze mechanisms to accurately manage turn-exchanges and to signal appropriate participant roles to human interlocutors? How do people respond to the robot's turn-yielding and role-signaling cues? What social and cognitive effects do different forms of conversational participation have on people? In this section, we describe our hypotheses, participants, the experimental setup, task, and procedure, measurements, and results.

4.1. Hypotheses

Four hypotheses were developed from existing human communication theory on conversational participation, person perception, and group formation.

Hypothesis 1. Subjects will correctly interpret the footing signals that the robot communicates to them and conform to these roles in their participation to the conversation. Therefore, addressees will take more speaking turns and speak longer than bystanders and overhearers will. The support for this prediction is the suggestion that, in conversations that involve more than two participants, the gaze of the speaker toward another participant indicates that the speaker is addressing that participant [Bales et al. 1951; Sacks et al. 1974; Goodwin 1981] and breakdowns might occur when the speaker's gaze is not directed toward the intended participant [Schegloff 1968].

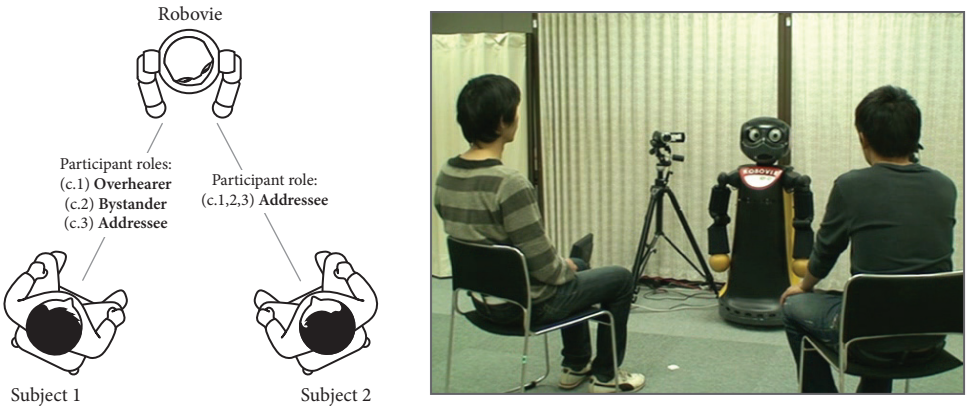

Fig. 9. The spatial setup of the experiment (left) and participants conversing with the robot in an experimental trial (right).

Hypothesis 2. Addressees will have better recall of the details of the information presented by the robot than bystanders and overhearers will, as the robot will look toward the addressees significantly more. This prediction is supported by findings from classroom research, which suggest that increased teacher gaze toward students improves information recall across different student populations from primary school to college [Otteson and Otteson 1980; Sherwood 1987]. Our previous research has also shown that increased gaze of a storytelling robot toward subjects improves their recall of the information presented in the story [Mutlu et al. 2006].

Hypothesis 3. Addressee or bystanders will evaluate the robot more positively than overhearers will. The basis for this hypothesis is that gaze cues shape people's perceptions of individuals; people who maintain eye contact and overall use more gaze toward conversational partners are evaluated more positively by their partners than those who avert gaze or overall use less gaze toward their partners are [Kleck and Nuessle 1968; Cook and Smith 1975; Mason et al. 2005].

Hypothesis 4. Addressees will express stronger feelings of groupness (with the robot and the other subject) than bystanders and overhearers will. This prediction is supported by the suggestion that gaze cues serve as a form of evaluation of the speaker's relationship with others and avoiding gaze might indicate social exclusion [Wirth et al. 2010].

4.2. Participation

A total of 72 subjects participated in the experiment in 36 trials. All subjects were native-Japanese-speaking university students recruited from the Osaka area. The ages of the subjects varied between 18 and 24 with an average of 20.8 years. Subjects represented a variety of university majors including management sciences, social sciences, humanities, engineering, and natural sciences. 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 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.

4.3. Setup, Task, and Procedure

To contextualize our design of robot gaze cues for an experimental task, we choreographed a conversational scenario in which Robovie played the role of a travel agent. Figure 9 illustrates the spatial configuration of the robot and subjects and

participants in an experimental trial. The robot provided participants with options of travel packages ("value" and "premium") and destinations (Spain and Turkey—two equally popular travel destinations in Japan) and adapted its information to their choices. It also assessed participants' knowledge of the travel destinations that they chose 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?" We used Wizard-of-Oz techniques to evaluate participants' responses. Following is the English translation of an example segment of the robot's conversation with subjects from an experimental trial.

> *Robovie:* Antalya, which is located on the South coast of Turkey, is famous for its festivals, long beaches and natural beauty. You can also visit close-by architectural sites from the Roman and Greek periods. Summer is the best time to be in Antalya. [Pause] Do you plan to travel with your family or with your friends?
>
> *Subject:* With my friends.
>
> *Robovie:* That's good. This package is particularly preferred by people like you. [Pause] There are many summer activities that can be done in groups in Turkey. In Antalya, you can do outdoor sports such as rafting, canoeing, and water or hill gliding. In Istanbul, you can take a tour of the Bosporus on the boat or see a Formula 1 race. Races in the Turkish track are held in the summer. [Pause] Did you know that Turkey has the newest Formula 1 track in the World?
>
> *Subject:* I didn't know that.
>
> *Robovie:* Oh, yes. They say that this is one of the most difficult tracks as well.

The robot followed the common interaction rituals of a conversation. During the greeting, it introduced itself to its conversational partners, asked them for their names, and told them that it was happy to meet them. Before leave-taking, it told its partners that it had to talk to another customer, but it was nice meeting them, and thanked them for their interest.

Robovie's speech was identical across conditions except for changes due to the adaptive dialog. A prerecorded nongendered voice was used for Robovie's speech. We did not use speech recognition during the experiment. Instead, the experimenter initiated the robot's turns in the conversation, selecting from among a preset sequence of utterances from a library. Following a between-participants design, we manipulated the robot's gaze behavior in three conditions:

Condition 1. The robot produced gaze cues for an addressee and an overhearer (ignoring the individual in the latter role), following the norms of a *two-party conversation*.

Condition 2. Gaze cues were produced for an addressee and a bystander, signaling the participation structure of a *two-party conversation with bystander*.

Condition 3. The robot produced gaze cues for two addressees, following the participant roles of a *three-party conversation*.

In the experiment, 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 preexperiment questionnaire. Both subjects were told that researchers were developing a travel agent robot and would like their help in evaluating their design.

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 would be asked to answer a questionnaire on their experience and to recall the material presented by the robot. After completing the task, subjects answered a postexperiment questionnaire that measured their recall of the information, their affective state, their perceptions of the robot, the group, and the task, and basic demographic information.

The task and the entire experiment procedure in total 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 ¥1,500 for their participation including their travel expenses.

4.4. Measurement and Analysis

The manipulation in the robot's gaze behavior was the only independent variable. The dependent variables involved three kinds of measurements: behavioral, objective, and subjective.

Behavioral. We captured subjects' behavior using high-definition cameras at 1080i resolution and stereo speakers. From the video and audio data, we measured whether subjects took turns in responding to the robot and how long they spoke.

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

Subjective. We 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. 1996], feelings of closeness to the robot [Aron et al. 1992], feelings of groupness and ostracism [Williams et al. 2000], perceptions of the task (how much they enjoyed and attended to the task), and demographic information.

The subjective evaluation also included a question for manipulation check: we asked subjects how much they thought the robot looked towards them and towards the other subject. We also used single-item measures to measure how much subjects thought the robot ignored them and considered their preferences in providing travel information. Seven-point rating scales were used in all questionnaire items.

All 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 might 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, the number of speaking turns taken by one of the subjects is affected by the number of turns taken by the other subject in the same trial given that the robot yielded a fixed number of turns. In this situation, the analysis of covariance compared the number of turns taken by subjects with different participant roles while accounting for the number of turns taken by other subject in the same trial. From the statistical modeling point of view, for each dependent variable, data from subjects with different participant roles (overhearers, bystanders, and addressees) were entered into the model as response variables and data from the other subject (addressees) were entered in the model as covariates. In the third condition, because both subjects were addressees, data was randomly sampled into response variables and covariates in equal size. Our analysis also included calculating item reliabilities for scales and correlations among dependent measures.

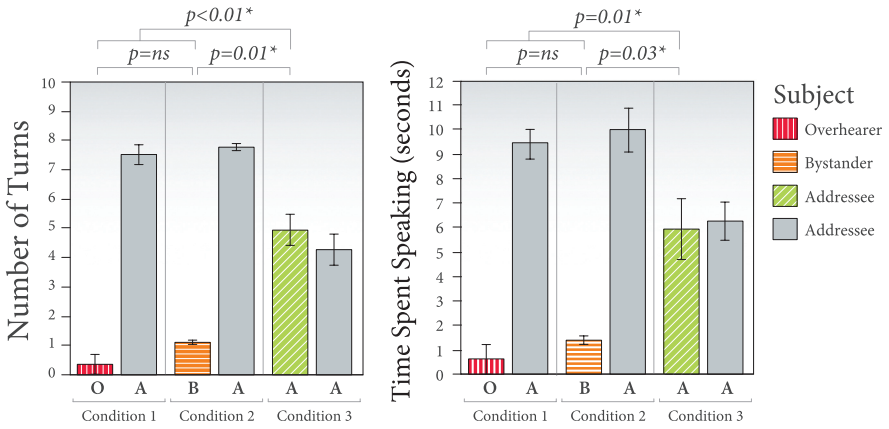

Fig. 10. The number of speaking turns that subjects took and the total time they spent speaking across the three conditions. The response variables are indicated with vertical stripes, horizontal stripes, and diagonal stripes for overhearers, bystanders, and addressees respectively. Covariates are indicated with light gray color and no texture.

4.5. Results

The paragraphs below provide a breakdown of the results from our analysis for behavioral, objective, and subjective measures.

4.5.1. Behavioral Measurements. In analyzing the behavioral data, we first looked at whether subjects to whom the robot yielded speaking turns took these turns. The analysis showed that subjects correctly interpreted these signals 98.71% of the time (307 of 311 turn-yielding signals) and conformed to them by taking speaking turns 97.11% of the time (302 of 311 turns). Of the nine turn-yielding signals to which they did not conform, six were passed between subjects (some addressees passed their turns to overhearers because they felt awkward talking to the robot while other subject was being ignored), three were not taken by the subjects due to ambiguities in robot's speech (in three trials, subjects did not perceive one of the questions as a question), and two were taken by both subjects as surprised responses to information presented by the robot (e.g., "Oh, I didn't know that")—these responses did not seem to be attempts to take the floor. The nonzero values for the overhearers in both measures are due to the six turns that addressees passed to them. Bystanders took an average of one turn as they responded to the robot during greetings.

Next, we conducted an analysis of covariance on the number of speaking turns that subjects took and the total time they spent speaking across the three conditions. Pairwise comparisons fully supported our first hypothesis that subjects would correctly interpret the booting signals that the robot communicated to them and conform to these roles in their participation to the conversation. Addressees took significantly more speaking turns than bystanders and overhearers did, $F(1, 30) = 17.58$, $p < .01$, and spoke significantly longer than bystanders and overhearers did, $F(1, 30) = 7.41$, $p = 0.01$. They also took significantly more speaking turns than bystanders alone did, $F(1, 30) = 6.75$, $p = .01$. They also spoke significantly longer than bystanders alone did, $F(1, 30) = 5.11$, $p = .03$. No significant differences were found between bystanders and overhearers. Figure 10 illustrates these comparisons.

4.5.2. Objective Measurements. Our second hypothesis predicted that addressees would have better recall of the information presented by the robot than bystanders and overhearers. This prediction was not supported by our analysis. There were no significant

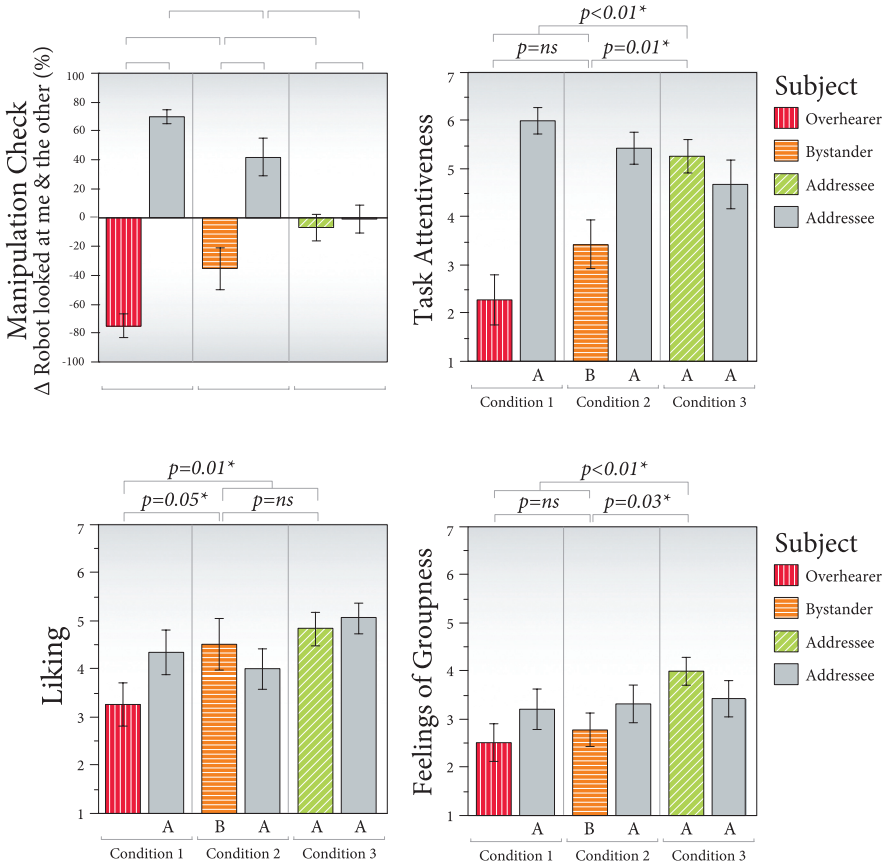

Fig. 11. The objective and subjective measurements.

differences across conditions in how well subjects recalled the information presented by the robot. The number of correct answers out of eight questions on average were 2.75 ($SD = 1.66$), 3.83 ($SD = 1.59$), and 3.17 ($SD = 1.47$) for overhearers, bystanders, and addressees respectively. The results from the subjective measure of attention might shed some light on this result, which will be provided in the next section.

4.5.3. Subjective Measurements. In analyzing the data from subjective measures, we first tested whether the gaze manipulation was successful. We did a manipulation check by taking the difference between subjects' ratings of how much the robot looked at them and their ratings of how much it looked at the other subject. We ran pairwise tests between pairs of different participant roles across and within conditions. We expected to see no difference between the ratings of the two addressees in the third condition and significant differences in all other pairwise comparisons. The results of the analysis supported our predictions. No differences were observed between the addressees in the third condition and all other comparisons were statistically significant with a marginal difference between ratings of bystanders and overhearers. Figure 11 provides results for all pairwise tests.

Next, we calculated item reliabilities for the two main measures that we used to test our third and fourth hypotheses. Item reliabilities for the three-item scale that

measured how much subjects liked the robot (*Cronbach's* $\alpha = .76$) and the six-item scale for measuring feelings of groupness (*Cronbach's* $\alpha = .92$) were sufficiently high.

While participant role did not affect subjects' recall of information, it affected their subjective ratings of how much they attended to the task. Addressees rated themselves as attending to the conversation significantly more than bystanders and overhearers did, $F(1, 29) = 12.90$, $p < .01$.

The third hypothesis predicted that subjects whose presence the robot acknowledges (addressees and bystanders) would like the robot more than those whose presence it does not acknowledge (overhearers). An analysis of covariance on subjects' liking of the robot supported our prediction. Addressees and bystanders liked the robot significantly more than overhearers, $F(1, 30) = 7.35$, $p = .01$. Bystanders alone also liked the robot significantly more than overhearers did, $F(1, 30) = 4.05$, $p = .05$, suggesting that the simple acknowledging gaze led subjects to like the robot more. There were no significant differences in addressees' and bystanders' liking of the robot. Figure 11 illustrates the results from these comparisons.

Our fourth hypothesis was also supported by our analysis. As predicted, those who were communicated the role of addressee by the robot and who contributed in the conversation as active participants rated their feelings of groupness significantly higher than those who did not contribute to the conversation as bystanders (except during greetings and leave-taking) and as overhearers, $F(1, 30) = 8.95$, $p < .01$. Addressees also rated their feelings of groupness as higher than bystanders alone, $F(1, 30) = 5.36$, $p = .03$, and overhearers alone, $F(1, 30) = 8.25$, $p < .01$. These comparisons are illustrated in Figure 11.

Our analysis of the data from single-item scales (on how much subjects thought the robot ignored them and considered their preferences in providing travel information) provides further explanation why overhearers liked the robot less than others did and why addressees felt more feelings of groupness than others did. Subjects whom the robot ignored did, in fact, feel significantly more ignored than both bystanders did, $F(1, 30) = 4.41$, $p = .04$, and addressees did, $F(1, 30) = 14.14$, $p < .01$, which perhaps led to their liking the robot less. Similarly, addressees, who contributed to the conversation more than others did, thought that the robot considered their preferences significantly more than bystanders did, $F(1, 30) = 4.05$, $p = .05$, and overhearers did, $F(1, 30) = 6.98$, $p = .01$. This mutual exchange conceivably led to more cohesion in the group as reflected in subjects' feelings of groupness. Figure 11 also highlights these comparisons.

Finally, Pearson product-moment correlations were calculated to understand how dependent variables related to each other. These analyses showed statistically significant correlations between familiarity with robots and liking, $r(34) = .26$, $p = .03$, task attentiveness $r(34) = .25$, $p = .04$, and feelings of groupness, $r(34) = .37$, $p < .01$.

4.5.4. Qualitative Observations. We also made a set of qualitative observations of how subjects interacted with the robot and performed the participant roles that the robot communicated to them. In our observations, subjects did not speak unless they were granted a turn to speak, with the exception that, in three trials, addressees showed in their nonverbal behavior hesitation and discomfort that the robot ignored the other conversational partner; therefore, they passed some of their speaking turns to overhearers. While this behavior is a breakdown in the participant structure established by the robot, it also illustrates how well people conformed to the signals that the robot communicated to them. Those to whom the robot did not yield speaking turns still did not take turns unless passed by the other subject, while those to whom the robot yielded turns knew that they had the floor, but took the liberty to pass their turns to the other subject.

Table III. Summary of Our Hypotheses and Whether They Were Supported by the Results

Hypothesis	Results
Subjects' compliance with the robot's turn-yielding and role-signaling cues **Hypothesis 1:** Subjects will correctly interpret the footing signals that the robot communicates to them and conform to these roles in their participation to the conversation. Therefore, addressees will take more speaking turns and speak longer than those bystanders and overhearers will.	*Supported*
The effect of participation on information recall **Hypothesis 2:** Addressees will have better recall of the details of the information presented by the robot than bystanders and overhearers will, as the robot will look toward the addressees significantly more.	*Not supported*
The effect of participant role on liking of the robot **Hypothesis 3:** Addressee or bystanders will evaluate the robot more positively than overhearers will.	*Supported*
The effect of participant role on feelings of groupness **Hypothesis 4:** Addressees will express stronger feelings of groupness (with the robot and the other subject) than bystanders and overhearers will.	*Supported*

In a number of trials, subjects hesitated to take the speaking turn after they received the first turn-yielding signal from the robot. We believe that this behavior was partly because they were not sure that the robot could understand them and partly because they felt uncomfortable talking to a robot in front of the experimenter and the other participant. We did not observe this behavior after the first turn exchanges.

When responding to the robot, subjects produced gaze signals similar to those observed in human communication. For instance, human communication research has found that "breaking mutual gaze" (looking away from the speaker) when answering questions is a common behavior [Libby 1970]. In our human-robot conversation, subjects broke mutual gaze with the robot when replying to 35.37% of all the questions and 47.12% of the questions that required them to make an evaluation (e.g., choosing of the travel destination) before answering. This behavior provides some evidence that the subjects perceived the turn-yielding gaze cues from the robot as valid social stimuli and responded to these signals by creating the appropriate communicative behavior.

5. DISCUSSION

Drawing on discourse theory and formal observations of human conversations, we identified three gaze mechanisms that signal three kinds of structural conversation information: *role-signaling* (communicates participation structure), *turn-taking* (signals conversation structure), and *topic-signaling* (expresses information structure). We reconstructed these mechanisms as a conversational system, implemented this system in a humanlike robot, and contextualized it in a human-robot conversation scenario. The experimental evaluation of the system supported three of four hypotheses that we posited based on theory on conversations. Our hypotheses and whether or not they were supported by our evaluation are summarized in Table III. Using only gaze cues, the robot manipulated who participated in and attended to a conversation, subjects' feelings of groupness, and their liking of the robot. Subjects accurately read the robot's turn-yielding gaze signals 99% of the time and conformed to these signals by taking 97% of the speaking turns. People also conformed to the participant roles that the robot communicated to them. Those whom the robot treated as addressees took

more speaking turns and spoke longer than those who were treated as bystanders or as overhearers. Addressees also attended to the task more and felt stronger feelings of groupness than others. Those whose presences were acknowledged (either as addressees or as bystanders) liked the robot more than those who were ignored by the robot as overhearers. Contrary to our prediction, participant role did not affect people's recall of the information by the robot.

A number of alternative explanations exist for why our prediction on information recall was not confirmed. One explanation is that providing subjects with the ability to choose a travel destination in order to create a more interactive experience might have compromised random assignment and introduced preferences and/or prior knowledge as confounding factors. In fact, we found that Condition 1 had an unbalanced number of travel destinations; 9 for Spain and 3 for Turkey. Furthermore, we found an effect of the topic of conversation (the travel destination) on recall of information, $F(1, 33) = 10.67, p < .01$, providing support for the potential effect of preferences and/or prior knowledge of the topic on information recall. These factors have likely increased the variance in our objective data in ways that our analysis cannot isolate. A second explanation is that limiting the subject population to males might have affected the outcome. Our previous work found that increased robot gaze improved subjects' recall of the information that the robot provided, while this effect was only present for females and males were not affected by the gaze manipulation [Mutlu et al. 2006]. This finding is consistent with the results presented here. Our prediction might hold for female participants, although the cultural differences between the U.S. American (the cultural context of the prior study) and Japanese subjects limit our ability to generalize across these studies. A final explanation is that, while we have confirmed the reliability of our measures in a pretest, the questionnaire that we developed to capture the subjects' recall of information failed to do so.

In this study, we also demonstrated an approach to understanding human-robot interaction that draws on an integrated process of carefully designing robot behaviors using theory and empirical findings, building these behaviors on a robotic platform, and experimentally evaluating how the designed behaviors shape people's interactions with the robot and the social and cognitive effects of these interactions on them. We argue that this approach can be applied to designing and evaluating a wide variety of technologies that embody and reciprocate with human communicative mechanisms such as augmented and virtual reality, computer-medicated communication and computer-supported collaborative work applications, and virtual characters. Examples of this approach exist in research in embodied conversational agents [Cassell et al. 1999a].

The findings of our modeling phase also point at interactions among the three gaze mechanisms that have not been discussed by human communication research. How speakers employ the topic-signaling and turn-taking mechanisms depend to a large extent on the participation structure of a conversation. For instance, a two-party conversation in which the participants' footings are equal, the interlocutors would not employ role-signaling, while conversations with unequal footings or those with multiple participants playing different roles would require speakers to rely significantly on role-signaling cues to achieve fluent conversations.

5.1. Limitations

While this work makes a number of theoretical, methodological, and practical contributions to the study of social behavior, our understanding of gaze behavior in conversations, and the design of communicative mechanisms for humanlike robots, these contributions are limited by a number of design decisions and shortcomings of current technology. Below, we describe such limitations for each of the three phases of this research.

Phase 1: Computational Modeling. A key limitation of the modeling phase is the use of data from a single speaker. While this choice allowed us to conduct an in-depth analysis of the data, it also imposed a limitation on the generalizability of the results of our observation to a larger population. This limitation can be addressed by collecting data from a larger sample and facilitating the data analysis through the use of eye-trackers for data collection and statistical methods such as contingency analysis to identify patterns in behavior.

The modeling of the speaker gaze behavior did not take the behaviors of the addressees into consideration. While this decision significantly reduced the complexity of the modeling problem, it also imposes an important limitation on the designed behavior and the results of the evaluation. Similarly, addressee gaze behavior was not considered in the evaluation of the conversational mechanisms. A major obstacle in modeling joint gaze behavior among conversational participants and in achieving interactive systems that take human gaze into account is the lack of robust, nonintrusive real-time tracking of gaze direction. These challenges might be alleviated by the development of such technologies.

A key characteristic of the formal design process that we followed in this work is to ground design decisions in validated theories of communication and findings from formal analyses of empirical data. Nevertheless, these decisions are influenced significantly by qualitative judgements that we made during in the modeling and design phases. For instance, our choice of unitizing the speech data at points of thematic transition—creating a rather large unit of analysis—forced us to seek patterns (initiated by the onset of a "thematic field") in our speaker's gaze behavior. A smaller unit of analysis (such as intonation units that represent the prosodic structure of speech) could have led to closer coupling between information structure and gaze shifts in the designed gaze behavior. Furthermore, our analysis ignored the link between the speaker's gaze shifts and many other linguistic properties of his speech, while an analysis at various levels of linguistic structures would have provided a more complete picture of this link. This sensitivity to differences in designer's choices in studying empirical data can be addressed by verifying the outcome of these design decisions through intermediary user studies.

While we drew on theory and data to develop models of communicative mechanisms, other equally valid approaches such as using animation principles (e.g., Van Breemen [2004]) or theatrical scripts of behavior (e.g., Lu et al. [2010]) exist. We argue that different approaches might be better suited for different applications or used together to achieve greater communicative complexity in human-robot conversations.

Phase 2: Design & Implementation. The generalizability of our results also suffers from a number of design decisions we have made in integrating the gaze mechanisms into a conversational system. For instance, this work focused on designing conversational gaze mechanisms and therefore necessarily limited the robot's behavior to speech and gaze. However, all aspects of nonverbal cues work together to create rich, human-like behavior. The elimination of gestures and body movement might have affected how people perceived the robot's gaze signals.

Another limitation is brought up by the limited interactivity of the robot, which forced us to design a conversational scenario where the robot held the floor for most of the conversation and yielded turns only at predetermined points in the conversation. The results of this study might have been different with a more fluent conversational scenario where participants took more turns and held the floor for longer periods. Robust speech recognition and adaptive speech generation would allow for the exploration of unscripted, fluent conversational scenarios.

Phase 3: Evaluation. The findings of the evaluation phase also have a number of limitations. First, because we only recruited male subjects, our results have limited generalization to conversational situations with female subjects or mixed-gender groups. Ideally, a gender-balanced, full-factorial-design study is required to understand how gender affects participation structure in human-robot conversations. Secondly, these results might not generalize beyond the cultural context of the study. Factors such as Japanese subjects closer familiarity with robots and the frequent use of interfaces that use speech in Japan might have affected our results. In fact, contrary to the results of this study, previous work that we conducted with a U.S. participant population [Mutlu et al. 2006] showed that people's liking of the robot was significantly correlated with video gaming experience and not with familiarity with robots, suggesting fundamental differences in how people perceive and interact with robots across cultures. Furthermore, differences in conversational conventions—particularly those brought about by age, social status, organizational rank, and so on—across cultures might affect our results. Our understanding of these cultural differences would greatly benefit from cross-cultural studies of human communication and human-robot interaction.

Because we did not tell subjects that they might be communicating with the robot in different participant roles, subjects might have felt the need to further regulate the roles that the robot communicated to them by, for instance, passing speaking turns among themselves. We argue that these subjects expected to be treated as equals by the robot—subjects' equal body orientations relative to the robot further supported this expectation—and that the robot's ignoring one of the subjects caused some discomfort. They might have tried to alleviate this discomfort through passing some of their speaker turns to the ignored subject. While this behavior shows the effectiveness of the robot's gaze behavior in signaling who is granted the next turn, it also highlights the ever-changing nature of participant roles in conversations as also emphasized by Goffman [1979]. This behavior also shows the importance of context in adapting participant roles. It was important for our study that subjects were given minimal information on the nature of the study—we wanted to test how well the robot could communicate to subjects their participant roles. The dynamic nature of participant roles and the role of context pose fruitful areas for future research on human-robot conversations.

Finally, gaze cues are sensitive to the cultural context and language of the conversation and, therefore, the models and results presented here might have limited generalizability beyond Japanese conversations. Cross-cultural studies of nonverbal behavior have classified 30 countries to fall within either "contact" or "noncontact" cultures and found that participants of conversations in contact cultures maintained more eye contact with their partners than those in noncontact cultures did [Watson 1970]. Therefore, conversations in noncontact cultures such as those in Japanese are expected to involve significantly lower levels of gaze among participants and patterns identified and outcomes observed in these conversations might not generalize those in contact cultures.

6. CONCLUSIONS

During conversations, speakers employ a number of gaze cues or *gaze mechanisms* to signal information on the structure of their speech, seamlessly exchange turns, and clarify the roles of the other participants of the conversation. The current work drew on discourse theory and formal observations of human conversations to develop models of three conversational gaze mechanisms, *role-signaling*, *turn-taking*, and *topic-signaling*, integrated these models into a conversational system, and implemented this system into a humanlike robot. The evaluation of the designed system assessed the

effectiveness of these models in signaling three participant roles, *addressee*, *bystander*, and *overhearer*, and the social and cognitive effects of these participant roles on subjects.

In a controlled laboratory experiment conducted with 72 subjects in 36 trials, we showed that manipulations in models of conversational gaze mechanisms affected subjects' participation in a conversation with the robot, how much they attended to the conversation, how much they liked the robot, and how strongly they felt a part of the group with the robot and their conversational partners. We found that subjects correctly interpreted 99% of the turn-yielding signals and took 97% of the turns that the robot yielded to them. Those who took turns as active participants of the conversation rated their attentiveness to the conversation higher than those who did not take speaking turns did. They also felt more acknowledged, welcomed, and valued by their group and that they belonged more to the group than those who remained as nonparticipant bystanders and as overhearers. Bystanders, whose presence the robot acknowledged with simple non-turn-yielding gaze signals, evaluated the robot more positively than overhearers, for whom the robot did not produce these signals.

While the results presented here are limited to the participant gender and culture we studied and conversational context we created, they provide evidence on how robots might use conversational gaze mechanisms to achieve fluent conversations, successfully managing conversational turns and effectively communicating participant roles.

ACKNOWLEDGMENTS

We would like to thank Sara Kiesler for her guidance in experimental design, Toshiyuki Shiwa for his help with data coding and the execution of the experimental evaluation, Laurel Bancroft for her help with reliability coding, Fumitaka Yamaoka and Kazuhiko Shinozawa for their support with the implementation of the experiment, and the coeditors of this special issue, Joyce Chai and Elizabeth André, and the three anonymous reviewers for their feedback on the earlier versions of this article.

REFERENCES

ARGYLE, M. AND DEAN, J. 1965. Eye-contact, distance and affiliation. *Sociometry 28,* 3, 289–304.

ARGYLE, M. AND INGHAM, R. 1972. Gaze, mutual gaze, and proximity. *Semiotica 6,* 1, 32–49.

ARON, A., ARON, E., AND SMOLLAN, D. 1992. Inclusion of other in the self scale and the structure of interpersonal closeness. *J. Person. Soc. Psych. 63,* 4, 596–612.

BAILENSON, J., BEALL, A., LOOMIS, J., BLASCOVICH, J., AND TURK, M. 2005. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Hum. Comm. Resear. 31,* 4, 511–537.

BALES, R. 1970. *Personality and Interpersonal Behavior*. Holt, Rinehart, and Winston, New York.

BALES, R., STRODTBECK, F., MILLS, T., AND ROSEBOROUGH, M. 1951. Channels of communication in small groups. *Amer. Soc. Rev. 16,* 4, 461–468.

BENNEWITZ, M., FABER, F., JOHO, D., SCHREIBER, M., AND BEHNKE, S. 2006. Towards a humanoid museum guide robot that interacts with multiple persons. In *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 418–423.

BOUMAN, C. 1997. Cluster: An unsupervised algorithm for modeling Gaussian mixtures. http://www.ece.purdue.edu/ bouman.

BROWN, G., CURRIE, K., AND KENWORTHY, J. 1980. *Questions of Intonation*. Routledge.

BROWN, P. AND LEVINSON, S. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press.

CASSELL, J., BICKMORE, T., BILLINGHURST, M., CAMPBELL, L., CHANG, K., VILHJÁLMSSON, H., AND YAN, H. 1999a. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 520–527.

CASSELL, J., NAKANO, Y., BICKMORE, T., SIDNER, C., AND RICH, C. 2001. Non-verbal cues for discourse structure. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 114–123.

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 *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 413–420.

CASSELL, J., TORRES, O., AND PREVOST, S. 1999b. *Turn taking vs. discourse structure: How best to model multimodal conversation*. In Machine Conversations, Kluwer, 143–154.

CLARK, H. 1992. *Arenas of Language Use*. University of Chicago Press.

CLARK, H. 1996. *Using Language*. Cambridge University Press.

CLARK, H. AND CARLSON, T. 1982. Hearers and speech acts. *Language 58,* 2, 332–373.

COLBURN, A., COHEN, M., AND DRUCKER, S. 2000. The role of eye gaze in avatar mediated conversational interfaces. Tech. rep. MSR-TR-2000-81, Microsoft Research.

COOK, M. AND SMITH, J. 1975. The role of gaze in impression formation. *Brit. J. Soc. Clin. Psych. 14*, 19–25.

DUNCAN, S. 1972. Some signals and rules for taking speaking turns in conversations. *J. Person. Soc. Psych. 23,* 2, 283–292.

EDELSKY, C. 1981. Who's got the floor? *Lang. Soc. 10,* 03, 383–421.

EFRAN, J. 1968. Looking for approval: effects on visual behavior of approbation from persons differing in importance. *J. Person. Soc. Psych. 10,* 1, 21–25.

EXLINE, R. 1963. Explorations in the process of person perception: visual interaction in relation to competition, sex, and need for affiliation1. *J. Person. 31,* 1, 1–20.

GARAU, M., SLATER, M., BEE, S., AND SASSE, M. 2001. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 309–316.

GARAU, M., SLATER, M., VINAYAGAMOORTHY, V., BROGNI, A., STEED, A., AND SASSE, M. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 529–536.

GOFFMAN, E. 1955. On face-work; an analysis of ritual elements in social interaction. *Psych. Interper. Biol. Proc. 18,* 3, 213–231.

GOFFMAN, E. 1971. *Relations in Public: Microstudies of the Public Order*. Harper & Row.

GOFFMAN, E. 1979. Footing. *Semiotica 25,* 1-2, 1–30.

GOLDBERG, L., JOHNSON, J., EBER, H., HOGAN, R., ASHTON, M., CLONINGER, C., AND GOUGH, H. 2006. The international personality item pool and the future of public-domain personality measures. *J. Resear. Person. 40,* 1, 84–96.

GOODWIN, C. 1980. Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning. *Soc. Inq. 50,* 3-4, 272–302.

GOODWIN, C. 1981. *Conversational Organization: Interaction between Speakers and Hearers*. Academic Press New York.

GROSZ, B. AND SIDNER, C. 1986. Attention, intentions, and the structure of discourse. *Computat. Linguist. 12,* 3, 175–204.

HALLIDAY, M. 1967. *Intonation and Grammar in British English*. Mouton.

HANKS, W. 1996. *Language & Communicative Practices*. Westview Press.

HANNA, J. AND BRENNAN, S. 2007. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *J. Mem. Lang. 57,* 4, 596–615.

HAYASHI, R. 1988. Simultaneous talkÑfrom the perspective of floor management of English and Japanese speakers. *World Englishes 7,* 3, 269–288.

HEYLEN, D., ES, I., NIJHOLT, A., AND DIJK, B. 2005. Controlling the gaze of conversational agents. In *Advances in Natural Multimodal Dialogue Systems*, N. Ide, J. Véronis, H. Baayen, K. Church, J. Klavans, D. Barnard, D. Tufis, J. Llisterri, S. Johansson, J. Mariani, J. Kuppevelt, L. Dybkjær, and N. Bernsen, Eds. Text, Speech and Language Technology Series, vol. 30, Springer, Berlin, 245–262.

HINDS, J. 1976. *Aspects of Japanese Discourse Structure*. Kaitakusha.

HIRSCHBERG, J. AND GROSZ, B. 1992. Intonational features of local and global discourse structure. In *Proceedings of the Workshop on Speech and Natural Language*. Association for Computational Linguistics, 441–446.

HIRSCHBERG, J. AND PIERREHUMBERT, J. 1986. The intonational structuring of discourse. In *Proceedings of the 24th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 136–144.

HYMES, D. 1972. *Models of the Interaction of Language and Social Life*. Holt, Rinehalt & Winston, 35–77.

Ishiguro, H., Ono, T., Imai, M., Maeda, T., Kanda, T., and Nakatsu, R. 2001. Robovie: An interactive humanoid robot. *Indust. Robot. A Int. J. 28,* 6, 498–504.

Kendon, A. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica 26,* 1, 22.

Kirchner, N., Alempijevic, A., and Dissanayake, G. 2011. Nonverbal robot-group interaction using an imitated gaze cue. In *Proceedings of the 6th International Conference on Human-Robot Interaction*. ACM, 497–504.

Kleck, R. and Nuessle, W. 1968. Congruence between the indicative and communicative functions of eye contact in interpersonal relations. *Brit. J. Soc. Clin. Psych. 7*, 241–246.

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*. ACM, 1191–1194.

Laurel, B. 1991. *Computers as Theatre*. Addison-Wesley.

Lee, S., Badler, J., and Badler, N. 2002. Eyes alive. *ACM Trans. Graph.* 21. ACM, 637–644.

Levinson, S. 1988. *Putting Linguistics on a Proper Footing: Explorations in Goffman's Concepts of Participation.* 161–227, Oxford, UK, Polity Press.

Libby, W. 1970. Eye contact and direction of looking as stable individual differences. *J. Exper. Resear. Person. 4*, 303–312.

Lu, D., Pileggi, A., Wilson, C., and Smart, W. 2010. What Can Actors Teach Robots About Interaction? In *Proceedings of AAAI Spring Symposium Series*.

Mason, M., Tatkow, E., and Macrae, C. 2005. The look of love. *Psych. Sci. 16,* 3, 236–239.

Maynard, S. 1986. On back-channel behavior in Japanese and English casual conversation. *Linguistics 24,* 6, 1079–1108.

Maynard, S. 1989. *Japanese Conversation: Self-Contextualization through Structure and Interactional Management*. Ablex Publishing, Norwood, NJ.

McLaughlin, M. and Cody, M. 1982. Awkward silences: Behavioral antecedents and consequences of the conversational lapse. *Hum. Comm. Resear. 8,* 4, 299–316.

Murray, N., Roberts, D., Steed, A., Sharkey, P., Dickerson, P., and Rae, J. 2007. An assessment of eye-gaze potential within immersive virtual environments. *ACM Trans. Multimed. Comput. Comm. Appl. 3,* 4, 8.

Mutlu, B., Forlizzi, J., and Hodgins, J. 2006. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 518–523.

Nielsen, G. 1962. *Studies in Self Confrontation*. Munksgaard, Copenhagen.

Otteson, J. and Otteson, C. 1980. Effect of teacher's gaze on children's story recall. *Percept. Motor Skills 50*, 35–42.

Parise, S., Kiesler, S., Sproull, L., and Waters, K. 1996. My partner is a real dog: cooperation with social agents. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM, 399–408.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X., Kirbas, C., McCullough, K., and Ansari, R. 2002a. Multimodal human discourse: gesture and speech. *ACM Trans. Comput.-Hum. Interac. 9,* 3, 171–193.

Quek, F., McNeill, D., Bryll, R., Kirbas, C., Arslan, H., McCullough, K., Furuyama, N., and Ansari, R. 2000. Gesture, speech, and gaze cues for discourse segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitionn*. Vol. 2, 247–254.

Quek, F., McNeill, D., Bryll, R., Kirbas, C., Arslan, H., McCullough, K., Furuyama, N., and Ansari, R. 2002b. Gesture, speech, and gaze cues for discourse segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 2, IEEE, 247–254.

Rehm, M. and André, E. 2005. Where do they look? Gaze behaviors of multiple users interacting with an embodied conversational agent. In *Intelligent Virtual Agents*, Springer, 241–252.

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

Schegloff, E. 1968. Sequencing in Conversational Openings. *Amer. Anthro. 70,* 6, 1075–1095.

Schegloff, E. 2000. Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc. 29,* 1, 1–63.

Schegloff, E. and Sacks, H. 1973. Opening up closings. *Semiotica 8,* 4, 289–327.

Schiffrin, D. 1988. *Discourse Markers*. Cambridge University Press.

Sherwood, J. 1987. Facilitative effects of gaze upon learning. *Percept. Motor Skills 64*, 1275–1278.

Sidner, C., Kidd, C., 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*. ACM, 78–84.

Staudte, M. and Crocker, M. 2011. Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition 120*, 268–291.

Steptoe, W., Wolff, R., Murgia, A., Guimaraes, E., Rae, J., Sharkey, P., Roberts, D., and Steed, A. 2008. Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM, 197–200.

Tanaka, H. 1999. *Turn-Taking in Japanese Conversation: A Study in Grammar and Interaction*. Vol. 3, John Benjamins Publishing Company.

Tannen, D. 2005. *Conversational Style: Analyzing Talk Among Friends*. Oxford University Press.

Thomas, F. and Johnston, O. 1995. *The Illusion of Life: Disney Animation*. Hyperion New York.

Thórisson, K. 2002. Natural turn-taking needs no manual: Computational theory and model, from perception to action. *In Multimodality in Language and Speech Systems,* Kluwer, 173–207.

Trafton, J., Bugajska, M., Fransen, B., and Ratwani, R. 2008. Integrating vision and audition within a cognitive architecture to track conversations. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*. ACM, 201–208.

Van Breemen, A. 2004. Bringing robots to life: Applying principles of animation to robots. In *Proceedings of Shaping Human-Robot Interaction Workshop at the 22nd ACM/SigCHI Conference on Human Factors in Computing*.

Vertegaal, R., Slagter, R., 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*. ACM, 301–308.

Vertegaal, R., van der Veer, G., and Vons, H. 2000. Effects of gaze on multiparty mediated communication. In *Proceedings of Graphics Interface*. 95–102.

Vilhjálmsson, H. and Cassell, J. 1998. Bodychat: Autonomous communicative behaviors in avatars. In *Proceedings of the 2nd International Conference on Autonomous Agents*. ACM, 269–276.

Wang, N. and Johnson, W. 2008. The Politeness Effect in an intelligent foreign language tutoring system. In *Intelligent Tutoring Systems*, Springer, 270–280.

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

Ward, N. and Tsukahara, W. 2000. Prosodic features which cue back-channel responses in English and Japanese* 1. *J. Pragmatics 32,* 8, 1177–1207.

Watson, D., Clark, L., and Tellegen, A. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Person. Soc. Psych. 54,* 6, 1063–1070.

Watson, O. 1970. *Proxemic behavior: A cross-cultural study*. Mouton, The Hague.

Weisbrod, R. 1965. Looking behavior in a discussion group. Unpublished manuscript. Cornell University.

Whittaker, S. and Stenton, P. 1988. Cues and control in expert-client dialogues. In *Proceedings of the 26th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 123–130.

Wilkes-Gibbs, D. and Clark, H. 1992. Coordinating beliefs in conversation. *J. Mem. Lang. 31,* 2, 183–194.

Williams, K., Cheung, C., and Choi, W. 2000. Cyberostracism: Effects of being ignored over the Internet. *J. Person. Soc. Psych. 79,* 5, 748–762.

Wirth, J., Sacco, D., Hugenberg, K., and Williams, K. 2010. Eye gaze as relational evaluation: Averted eye gaze leads to feelings of ostracism and relational devaluation. *Personal. Soc. Psych. Bull. 36,* 7, 869–882.

Yamazaki, A., Yamazaki, K., Burdelski, M., Kuno, Y., and Fukushima, M. 2010. Coordination of verbal and non-verbal actions in human-robot interaction at museums and exhibitions. *J. Pragmatics 42,* 9, 2398–2414.

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 26th SIGCHI Conference on Human Factors in Computing Systems*. ACM, 131–140.

Yngve, V. 1970. On getting a word in edgewise. In *Proceedings of the 6th Regional Meeting of the Chicago Linguistic Society*. 657–677.

Received December 2010; revised August 2011; accepted October 2011