Footing In Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues

Bilge Mutlu[1], Toshiyuki Shiwa[2], Takayuki Kanda[2], Hiroshi Ishiguro[2,3], Norihiro Hagita[2]

(1) Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Ave,
Pittsburgh, PA 15213, USA
bilge@cs.cmu.edu

(2) ATR Intelligent Robotics and Communication Laboratory
2-2-2 Hikaridai, Keihanna,
Kyoto, Japan
{yamaoka, kanda, hagita}@atr.jp

(3) Faculty of Engineering
Osaka University
2-1 Yamadaoka, Suita City,
Osaka, Japan
ishiguro@ams.eng.osaka-u.ac.jp

ABSTRACT

During conversations, speakers establish their and others' participant roles (who participates in the conversation and in what capacity)—or "footing" as termed by Goffman—using gaze cues. In this paper, we study how a robot can establish the participant roles of its conversational partners using these cues. We designed a set of gaze behaviors for Robovie to signal three kinds of participant roles: *addressee*, *bystander*, and *overhearer*. We evaluated our design in a controlled laboratory experiment with 72 subjects in 36 trials. In three conditions, the robot signaled to two subjects, only by means of gaze, the roles of (1) two addressees, (2) an addressee and a bystander, or (3) an addressee and an overhearer. Behavioral measures showed that subjects' participation behavior conformed to the roles that the robot communicated to them. In subjective evaluations, significant differences were observed in feelings of groupness between addressees and others and liking between overhearers and others. Participation in the conversation did not affect task performance—measured by recall of information presented by the robot—but affected subjects' ratings of how much they attended 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, Human Factors

Keywords: Conversational participation, Participant roles, Participation structure, Footing, Gaze, Robovie

1. INTRODUCTION

In the future, robots might serve a variety of informational tasks as information booth attendants, museum guides, shopkeepers, security guards, and so on. In this capacity, such robots will have to communicate using human verbal and nonverbal language and carry on *conversations* with people. Consider the following three scenarios that involve our robot Robovie (Figure 1):

> Aiko is a shopper at a shopping mall in Osaka, where Robovie serves as an information booth attendant. Aiko is trying to find the closest Muji store and wants to know whether the store also sells furniture. She approaches Robovie's booth to inquire about the shop.

This conversational situation is a two-party conversation in which Robovie and Aiko will take turns to play the roles of *speaker* and *addressee* [11]. There might also be *overhearers* of this conversation without the knowledge of neither the speaker nor the addressee [19].

HRI'09, March 11–13, 2009, La Jolla, California, USA.

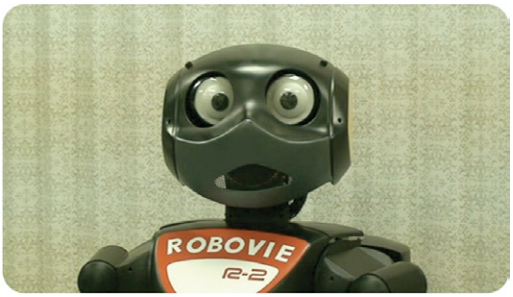

Figure 1. Robovie R-2, the humanlike robot we used in our study.

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

This scenario differs with the addition of a non-participant [11] into the social situation who is playing the role of a *bystander* [19].

> After Robovie's conversation with Yukio is over, a couple, Katsu and Mari, approach the booth, inquiring about Korean restaurants. Robovie asks the couple a few questions on their dining preferences and leads them to a suitable restaurant.

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

These situations illustrate different forms of "participation structures" [20], "participant roles" [24], or "footing" [19]—that is, the "position or status assigned to a person, group, etc., in estimation or treatment" [12]. Considerable evidence suggests that, during conversations, people use gaze cues to perform a social-regulative process of establishing their and others' footing [4,24,37,38]. Research in human-computer interaction has shown that gaze cues can be effective in shaping participant roles when used by virtual agents [3,36]. While a robot's use of these cues is

shown to perform other conversational functions such as managing turn-taking behavior [31,46] and showing appropriate listening behavior [41], whether and how they might shape different forms of participation are unexplored. Furthermore, whether social cues that affect social phenomena in human communication, such as person perception and group formation, lead to similar social outcomes in human-robot communication is unknown.

What cues might robots use to shape participant roles in conversations? Would the use of these cues lead to significant social outcomes such as stronger feelings of groupness or more liking? In this paper, we try to answer these questions by gaining a deeper understanding of the concept of footing from human communication theory and observations of human conversations, exploring how these cues might be designed for robots to shape participant roles in human-robot conversations, and examining the social outcome led by different forms of participation.

2. RELATED WORK

In conversations, people work together as participants [11]. The roles of the participants, a phenomenon described by Goffman as "footing" [19], and how these roles might shift during social interaction are particularly important for understanding spoken discourse [25,27]. At the core of these roles are those of the *speaker* and the *addressee* [11]. While these roles might be fixed in some social settings (e.g. lectures), most conversational settings allow for shifting of roles. At any "moment" [19] in a two-party conversation, one of the participants plays the role of the speaker and the other plays the addressee. Conversations with more than two participants also involve "side participants" who are the "unaddressed recipients" of the speech at that moment [11,19,43].

In addition to these "ratified participants" [19], conversations might involve "non-participants" [11]. For instance, there might be *bystanders* whose presence the participants acknowledge and who observe the conversation without being participants in it [10,11,19]. 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 [19]. Figure 2 provides an abstract illustration of these different levels of participation.

The direction of gaze plays an important role in establishing and maintaining conversational participant roles. In conversations that involve more than two people, the gaze of a speaker towards another participant can signal that the speaker is addressing that participant [24,37]. In this situation, the speaker indicates a "communication target" [4]. When there is no intended target (i.e., when a speaker is addressing a group), gazing at a participant long enough might create the belief that the speaker is addressing primarily that participant [5]. 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 [38].

Gaze direction also serves as an important cue in shifting roles during turn-exchanges [13,23,24,35,38] and overlapping talk [39]. 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 [35]. 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 [44, as described in 30]. Shifting of roles might be delayed when remarks do not end with gazing at another participant [30,42]. When gaze levels are particularly low, such as

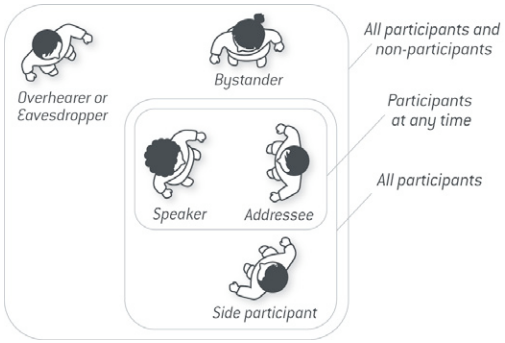

Figure 2. An illustration of different forms of conversational participation (adapted from [11]).

in a conversation between strangers, gaze plays an especially important role in cueing role exchanges [6].

Gaze Cues and Conversations with Embodied Virtual Agents

In human-computer interaction research, the use of gaze cues in conversations has been extensively studied in the context of designing embodied conversational agents [8,21,26,36,40]. Cassell and her colleagues developed a number of systems that use verbal and nonverbal behaviors to support communicative mechanisms such as turn-taking, feedback, repair, synchronized speech, and intonation [8,9]. While these systems combined nonverbal cues such as gaze, facial expressions, hand gestures, and postural shifts in the design of the agent, gaze cues were considered as the most salient signal to establish conversational roles and regulate turn-taking [9,43]. Furthermore, research in this area has shown that signals that are designed to resemble human gaze behavior (as opposed to randomly generated signals) lead to more efficient conversations, better task performance, and more positive evaluations of the agent [21,22,26].

Two studies on the use of gaze cues in conversational agents focused on understanding how these cues might shape participant roles and how different forms of participation might affect the social outcome of human-agent conversations [3,36]. Bailenson and his colleagues [3] studied how speaker gaze cues might be "augmented" to create the impression in two listeners that they are being addressed simultaneously. They compared participants' evaluations of the speaker across augmented and normal gaze conditions and found that people agreed with the speaker's message more in the augmented gaze condition than in the normal gaze condition. Rehm and Andre [36] asked two participants to play a game with a virtual character in which each player took turns to play the roles of speaker and addressee and evaluated people's involvement in the conversation. Their results suggest that, when appropriate cues are present, people conform to the participant roles that an agent communicates to them.

Gaze Cues and Conversations with Humanoid Robots

In research in human-robot interaction, a more recent but growing body of literature looks at social gaze behavior [28,29,31,34,40,41,46,47]. Among these, a few promising studies have examined the conversational effectiveness of robot gaze, particularly in regulating turn-taking in two-party [31,46] and multi-party conversations [7,33,41]. Kuno and others [32] developed gaze behaviors for a museum guide robot that looked at its addressee at "turn-relevant places" [37] (points in the conversation when turns exchange is expected) to regulate turn-taking. Yamazaki and others [46] showed that the robot evoked more backchannel responses when it looked at participants at turn-relevant places than when it looked at random places. Matsusaka and his colleagues [33] and Bennewitz and his colleagues [7]

developed robots that participated in multi-party conversations following the turn-taking model suggested by Sacks and his colleagues [37] for human conversations. Trafton and his colleagues [41] developed appropriate listening behaviors for a robot as a bystander and experimentally showed that interlocutors rated the robot's gaze behavior to be more natural when the robot looked at the speaker only during turns as opposed to during turns and backchannel responses.

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

3. METHODOLOGY

To gain a deeper understanding of how gaze cues might shape footing in human-robot conversations, we conducted a laboratory experiment in which we asked participants to converse with a humanlike robot, ATR's Robovie R2 (see Figure 1). We designed the robot's gaze behavior to signal three kinds of participant roles:

Addressees are participants who take speaking turns and contribute to the conversation, and whom the robot addresses while speaking.

Bystanders are acknowledged non-participants who do not take speaking turns (except during greetings and leave-taking) and whom the robot does not address while speaking, but whose presence it acknowledges during the conversation, particularly during greetings and leave-taking.

Overhearers are unacknowledged non-participants who do not take speaking turns, whom the robot does not address while speaking, and whose presence it does not acknowledge 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 as interchangeable with eavesdropper or ignored.

In the following paragraphs, we describe our interaction design of the robot's gaze signals, experimental design, hypotheses, the procedure we followed in the experiment, and, finally, our evaluation measures and subject profile.

3.1 Interaction Design of the Gaze Cues

Designing the robot's gaze cues to signal the three participant roles described above, we followed a *theoretically and empirically grounded design methodology* in which design decisions were informed by theories of human social communication and formal observations of how human speakers signal participant roles using gaze cues. These observations involved placing naïve participants in conversational situations with different role structures and studying speakers' gaze behavior. We hired four all-male triads and placed them in three conversational structures (Figure 3):

1. A two-party conversation with a speaker, an addressee, and an overhearer.
2. A two-party conversation with a speaker, an addressee, and a bystander.
3. A three-party conversation with a speaker and two addressees.

We used data from the triad that exhibited the most fluent interaction—evidenced by a qualitative evaluation of participants' involvement in the conversation and a quantitative assessment of the total time spent speaking without substantial pauses—for a detailed analysis of how gaze cues might signal footing. Only a brief account of our findings from this analysis will be provided here due to space constraints. Further discussion of the subject profile is provided in section the Participation subsection.

Figure 3. Participants in different conversational structures: two-party, two-party-with-bystander, and three-party conversations.

Greetings and *summonses* – An important point in conversations 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 [17] 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." Bales [5] suggests that speakers rely primarily on gaze cues to signal these roles. Schegloff [38] 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 structure, the two-party conversation with a bystander, at the point of the 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 participant roles.

Based on our findings from existing theory and our observations, we designed the robot's gaze behavior to acknowledge the presence of addressees and bystanders, but divert gaze towards addressees and away from bystanders at the point of transition from greetings to the body of the conversation.

The body of the conversation – In our observation, the speaker spent the majority of his speaking time looking at addressees. In the first conversational scenario, he looked towards his addressee 74% of the time and the environment 26% of the time. In the second scenario, the speaker allocated some of his gaze for the bystander (8%), mostly in short acknowledging glances averaging nearly half the average length of the gazes towards his addressee (in seconds, M=0.77, SD=0.58 vs. M=1.40, SD=1.30). The speaker looked towards the addressee, the bystander, and the environment 76%, 8%, and 16% of the time respectively. Finally, in the last scenario, the speaker looked towards his addressees 71% of the time and the environment 29% of the time.

We used these figures directly to design the gaze behavior of the robot during the body of the conversation. The addressees received the majority of gaze, between 71% and 76% of the time, and bystanders received 8% of the robot's gaze, mostly in very short, acknowledging glances.

Turn-exchanges – Another important point in conversations where participant roles are re-negotiated is turn-exchanges. Kendon [30] found that speakers mostly looked toward their addressees at the end of a turn, yielding the turn to the next speaker. Weisbrod [44 as described in 30] observed in seven-party conversations that the person towards 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 is also an important footing signal. We also observed that, 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

Table 1. Footing signals designed into the robot's gaze behavior that cue different participant roles.

	Addressees	Bystanders	Overhearers
Greetings	*Gaze*	*Gaze*	*No gaze*
Conversation	*Gaze*	*Short glances*	*No gaze*
Turn-exchanges	*Gaze*	*No gaze*	*No gaze*

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.

Based on the findings presented by Kendon [30] and Duncan [13] and our observations, we designed the robot's gaze behavior to produce turn-yielding signals only for addressees. Table 1 provides a summary of the footing cues designed for the robot's gaze behavior for the three participant roles that are considered in this study.

3.2 Experimental Design

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

We conducted a between-subjects experiment in which Robovie played the role of a travel agent and gave subjects information on travel packages. The robot first greeted subjects and introduced itself. It asked subjects for their names and told them that there are promotions to two travel destinations (Spain and Turkey) that they could choose between. After subjects chose a destination, the robot provided them with details of the travel package and general information on the travel destination. Throughout the interaction, the robot asked subjects questions regarding their travel preferences and their knowledge of the travel destination. Below is a typical question-answer pair from the experiment:

> Robovie: [Looking towards one of the participants] Did you know that the world's first coffee shop opened in Istanbul in the 15th century?
>
> Participant: Oh, I didn't know that.

Robovie's speech was identical across conditions except for changes due to the adaptive dialog. 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. We manipulated its gaze behavior in three conditions (Figure 4):

In **condition 1**, the robot regarded one of the participants as an addressee and the other as an overhearer, ignoring the individual in the latter role.

In **condition 2**, the robot regarded one of the participants as an addressee and the other as a bystander.

In **condition 3**, the robot regarded both participants as addressees.

3.3 Hypotheses

We developed four hypotheses from existing human communication theory on conversational participation, person perception, and group formation:

Hypothesis 1. – Subjects will correctly interpret the footing signals that robot communicates to them and conform to these roles in their participation to the conversation. Therefore, we predict that those who are granted speaking turns (addressees) by the robot will take more speaking turns and speak longer than those who are not granted speaking turns (bystanders and overhearers).

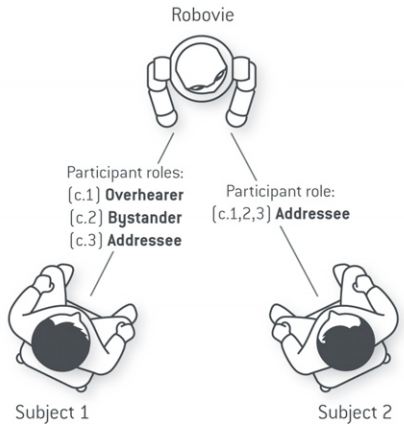

Figure 4. The spatial configuration of the robot and subjects and the participant roles that the robot communicated to the members of each pair of subjects in each experimental condition.

Hypothesis 2. – Subjects who contribute to the conversation by taking speaking turns (addressees) will recall the details of the information presented by the robot better than those who do not contribute to the conversation (bystanders and overhearers).

Hypothesis 3. – Subjects whose presences the robot acknowledges and to whom it communicates a participant role (either as addressee or bystander) will evaluate the robot more positively than those whose presences the robot does not acknowledge and to whom it does not communicate a participant role (overhearers).

Hypothesis 4. – Subjects to whom the robot communicates the role of addressee and who contribute to the conversation as active participants (addressees*)* will express stronger feelings of groupness (with the robot and the other subject) than those who are not active participants of the conversation (bystanders and overhearers).

3.4 Experiment Procedure

Subjects were first given a brief description of the purpose and the procedure of the experiment. After the introduction, they were asked to review and sign a consent form. Subjects were then provided with more detail on the task and asked to answer a pre-experiment questionnaire. Both subjects were told that researchers were developing a travel agent robot and would like 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 their recall of the material presented by the robot. After completing the task, subjects answered a post-experiment questionnaire that measured their information recall, affective state, perceptions of the robot, the group, and the task, and demographic information.

The task and the experiment procedure in total took an average of 7.5 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 ¥ (roughly $14 or €9) for their participation including their travel expenses. Figure 4 shows participants in the experiment.

3.5 Measurement

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.

Figure 5. The robot and subjects in the experiment.

Behavioral – We captured subjects' behavior using high-definition cameras and stereo speakers. From the video and audio data, we measured the number of turns that subjects took to respond to the robot and the total time that they spent speaking.

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

Subjective – We evaluated subjects' affective state, perceptions of the robot's physical, social, and intellectual characteristics, feelings of closeness to the robot, feelings of groupness and ostracism, 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 rating scales to measure how much subjects thought the robot ignored them and considered their preferences in providing travel information. Seven-point Likert scales were used in all questionnaire items.

3.6 Participation

Research in nonverbal behavior reports strong effects of group composition on both the production and the perception of gaze, particularly of gender [1,2,16] and age [14,32]. Our previous work also found gender effects on how the robot's gaze affected people's performances and their perceptions of the robot [34]. 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. Ideally, a full factorial, gender-balanced-design study is required to account for and have a better understanding of these gender-based differences. However, the number of subjects to conduct a full-factorial design goes beyond the resources of a single study. Therefore, as a start, we decided to control for these group composition effects, test our hypotheses in a smaller population, and plan for re-runs of the same design with other populations. Accordingly, in this study, we limited our subject profile for the observation to an all-male triad and experiment to all-male pairs with little variance in age. Similarly, a male experimenter administered the study.

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 were chosen to represent a variety of university majors. Of all the subjects, 26 studied management sciences, 23 studied social sciences & humanities, 16 studied engineering, 5 studied natural sciences, and 2 did not report their majors. Subjects were randomly assigned to the experimental conditions. The computer use among subjects was very high (M=6.27, SD=0.98) on a scale from 1 to 7. Their familiarity with robots was relatively low (M=2.97, SD=1.67), so was their video gaming experience (M=2.92, SD=1.91). Five (out of 72) subjects had toy robots and 23 owned pets. Figure 5 shows subjects in the experiment.

4. RESULTS

We analyzed behavioral, objective, and subjective measures 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 constant 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 the 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. In the figures hereafter, the response variables are indicated with vertical, horizontal, and diagonal stripes for overhearers, bystanders, and addressees respectively. Covariates are indicated with no texture. An ID number for each pairs of subjects was also included in the model as a random effect. We also calculated item reliabilities for scales and correlations across dependent measures. Below, results of the analyses of each set of measures are provided.

Behavioral – 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) and 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). Table 2 summarizes the mean and standard deviation values for the number of speaking turns that subjects took and the total time they spent speaking for each participant role in each condition. The non-zero 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

Table 2. Mean and standard deviation values for the number of speaking turns taken and total time spent speaking by subjects for each participant role.

	Condition 1 Overhearers Addressees Mean (StDev)	**Condition 2** Bystanders Addressees Mean (StDev)	**Condition 3** Addressees Mean (StDev)
Number of speaker turns (counts)	**0.33** (1.15) **7.50** (1.17)	**1.08** (0.29) **7.75** (0.45)	**4.54** (1.82)
Total time spent speaking (seconds)	**0.60** (2.09) **9.43** (2.17)	**1.38** (0.66) **10.00** (3.19)	**6.09** (3.48)

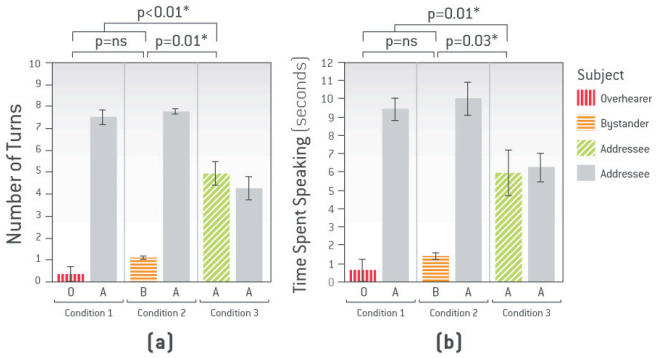

Figure 6. (a) The number of speaking turns taken, and (b) total time spent speaking by subjects. Textured bars represent response variables and plain bars represent covariates. (*) indicates statistically significant probabilities below .05.

speaking across the three conditions. Pairwise comparisons fully supported our first hypothesis. Addressees took significantly more speaking turns (F[1,30]=17.58, p<.01) and spoke significantly longer (F[1,30]=7.41, p=.01) than bystanders and overhearers. They also took significantly more speaking turns (F[1,30]=6.75, p=.01) and spoke significantly longer (F[1,30]=5.11, p=.03) than bystanders alone. No significant differences were found between bystanders and overhearers. These results are illustrated in Figures 6.a and 6.b.

Objective – Our second hypothesis predicted that addressees would have better recall of the information presented by the robot than bystanders and overhearers. Unfortunately, this prediction was not supported by our analysis. There were no significant differences across conditions in how well subjects recalled the information presented by the robot. The numbers 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. While participant role did not affect subjects' recall of information, it affected their ratings of how much they attended to the task. Addressees rated themselves as attending to the conversation significantly more (F[1,29]=12.90, p<.01) than bystanders and overhearers did (Figure 7.b). Furthermore, we found an effect of the topic of conversation (the travel destination) on recall of information (F[1,33]= 10.67, p<.01). The effect of participant role on attentiveness to the task and the effect of travel destination on information recall provide some insight into why our prediction was not supported by the results, which is further considered in the Discussion section.

Subjective – 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 participant. We conducted an analysis of variance (ANOVA) and 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 7.a. 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 α=.76) and the six-item scale for measuring feelings of groupness (Cronbach's α=.92) were sufficiently high.

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 (Figure 7.c). 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.

Our fourth hypothesis was also supported by our analysis (Figure 7.d). 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 (F[1,30]=8.95, p<.01) than those who did not contribute to the conversation as bystanders (except during greetings and leave-taking) and as overhearers. 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).

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 of 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, felt significantly more ignored than both bystanders (F[1,30]=4.41, p=.04) and addressees (F[1,30]=14.14, p<.01) did, which perhaps led to their liking the robot less. Similarly, addressees, who contributed to the

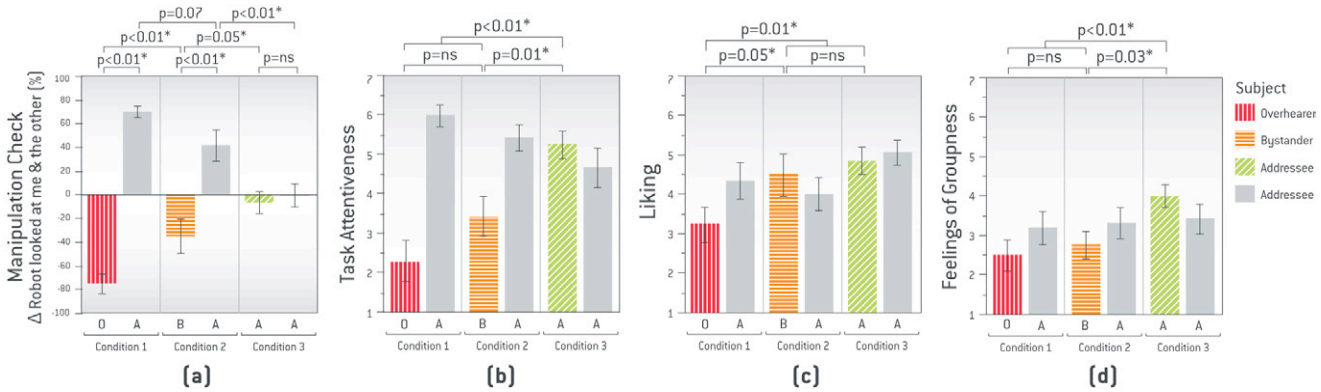

Figure 7. Subjects' ratings of (a) how much they thought the robot looked at the other subject subtracted from how much they thought the robot looked at them (i.e. manipulation check), (b) their task attentiveness, (c) their liking of the robot, and (d) their feelings of groupness. (*) indicates statistically significant probabilities below .05.

conversation more than others did, thought that the robot considered their preferences significantly more than bystanders ($F[1,30]=4.05$, $p=.05$) and overhearers ($F[1,30]=6.98$, $p=.01$) did. This mutual exchange conceivably led to more cohesion in the group as reflected in subjects' feelings of groupness.

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

Qualitative – 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, 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. Similarly, those to whom the robot yielded turns knew that they had the floor and felt the liberty to pass their turns to the other subject.

When responding to the robot, people often used articulate language—full sentences instead of phrases. They also 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 [32]. 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 producing the appropriate communicative behavior.

5. DISCUSSION

The results provided strong support for three of our four hypotheses. Only using 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 as addressees or as bystanders liked the robot more than those who were ignored as overhearers. Contrary to our prediction, participant role did not affect information recall.

Further analysis provides some insight into why our prediction on information recall was not confirmed. We found that addressees rated their attentiveness to the task higher than others did. While it is conceivable that attentiveness should lead to better recall of information, that the topic of the conversation significantly affected information recall suggests that subjects' prior knowledge of the topic might have been too well established to be affected by the information presented by the robot. Administering a pre-experiment questionnaire to measure prior knowledge of the topic would have helped in identifying how much new information was learned during the experiment. Alternatively, choosing a conversation topic, such as a fictional story, on which subjects would have sparser pre-existing knowledge could have provided support for our predictions.

Limitations – The results presented here have a number of limitations. First, that we only recruited male subjects limits how much our results generalize to conversational situations with female subjects or mixed-gender groups. Ideally, a gender-balanced, full-factorial-design study is required to understand how gender might affect participation structure in human-robot conversations. In the future, we plan to extend this work to make comparisons across different gender populations. Secondly, these results might not generalize beyond the cultural context of the study. That Japanese participants are culturally more familiar with robots and other interfaces that use speech might have affected our results. In fact, contrary to the results of this study, previous work that we conducted with an U.S. American population [34] 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-robot interaction (e.g., [15]).

The generalizability of our results also suffers from 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 scripted points. 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 exploration of unscripted, fluent conversational scenarios.

In this study, we focused on understanding how gaze cues might lead to different forms of conversational participation, and, therefore, necessarily limited the robot's behavior to speech and gaze. However, all aspects of nonverbal cues work together to create rich, humanlike behavior. Therefore, that we eliminated gestures and body movement might have affected how people perceived the robot's gaze signals. We plan to conduct future studies that compare how gaze cues with and without gestures might affect human-robot conversations.

That we did not tell subjects that they might be assigned different participant roles caused some subjects to further regulate the roles that the robot communicated to them. In three trials, addressees passed some of their turns to overhearers. 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 ignored 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 [19]. 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. We argue that the dynamic nature of participant roles and the role of context pose fruitful areas for future research on human-robot conversations.

CONCLUSIONS

During conversations, people use gaze cues to establish and maintain their and their conversational partners' participant roles, or "footing." In this paper, we study how these cues can be used by a robot to regulate footing in human-robot conversations. We designed gaze behaviors for a robot to cue three kinds of participant roles: addressee, bystander, and overhearer. In a controlled laboratory

experiment conducted with 72 subjects in 36 trials, we showed that these cues 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 a group with their conversational partners.

We found that subjects correctly interpreted 99% of the turn-yielding signals and took 97% of these turns. 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 non-participant 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 results presented in this paper are limited to the participant gender and culture we studied and conversational context we created, they provide evidence on how robots might use gaze cues for shaping participant roles in conversations. Further work is required to generalize our results and extend our understanding of how gaze cues relate to conversational organization in human-robot interaction.

ACKNOWLEDGEMENTS

This research was supported by the Ministry of Internal Affairs and Communications of Japan. The first author was supported in part by NSF grant IIS-0624275. We would like to thank Sara Kiesler, Jodi Forlizzi, and Jessica Hodgins for their guidance in experimental design, Fumitaka Yamaoka and Kazuhiko Shinozawa for their support with the implementation of the experiment, and Justine Cassell, Leila Takayama, and Lindsay Jacobs for their valuable comments.

REFERENCES

[1] Argyle, M. and Dean, J. Eye-contact, distance and affiliation. *Sociometry*, 28 (3), 289-304, 1965

[2] Argyle, M. and Ingham, R. Gaze, mutual gaze and proximity. *Semiotica*, 6, 32-49, 1972.

[3] Bailenson, J. N. et al. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human Communication Research*, 31 (4), 511-537, 2005.

[4] Bales, R. F. et al. Channels of communication in small groups. *American Sociological Review*, 16 (4), 461-68, 1951.

[5] Bales, R. F. *Personality and Interpersonal Behavior*. Holt, Rinehart, and Winston, 1970.

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

[7] Bennewitz, M. et al. Towards a humanoid museum guide robot that interacts with multiple persons. *In Proc. HUMANOIDS'05*, 2005.

[8] Cassell, J. et al. Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. *In Computer Graphics Proc.,* 413-420, 1994.

[9] Cassell, J. et al. Embodiment in conversational interfaces: Rea. In *Proc. CHI'99*, 520-527, 1999.

[10] Clark, H. H. and Carlson, T. B. Hearers and Speech Acts. *Language*, 58 (2), 332-373, 1982.

[11] Clark, H. H. *Using Language*. Cambridge University Press, 1996.

[12] Dictionary.com. Unabridged (v 1.1). Retrieved July 21, 2008, from http://dictionary.reference.com/browse/footing, 2008.

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

[14] Efran, J. S. Looking for approval: effects on visual behavior of approbation from persons differing in importance. *J. Personality and Social Psychology*, 10 (1), 21-25, 1968.

[15] Evers, V. et al. Relational vs. group self-construal: untangling the role of national culture in HRI. *In Proc. HRI'08,* 255-262, 2008

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

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

[18] Goffman, E. *Behavior in Public Places*. The Free Press, 1963.

[19] Goffman, E. Footing. *Semiotica*, 25 (1/2), 1-29, 1979.

[20] Levinson, S. C. Putting linguistics on a proper footing: Explorations in Goffman's concepts of participation. In P. Drew and A. Wootton (Eds.) *Erving Goffman: Exploring the Interaction Order,* 161-227. Polity, 1982.

[21] Garau, M. et al. The impact of eye gaze on communication using humanoid avatars. *In Proc. CHI'01*, 2001.

[22] Colburn, A. et al. *The role of eye gaze in avatar mediated conversational interfaces.* Microsoft Research Report, 81, 2000.

[23] Goodwin, C. Restarts, pauses, and the achievement of mutual gaze at turn-beginning. *Sociological Inquiry*, 50, 272-302, 1980.

[24] Goodwin, C. *Conversational Organization: Interaction between Speakers and Hearers.* Academic Press, 1981.

[25] Hanks, W. *Language and Communicative Practices*. Westview, 1996.

[26] Heylen, D. et al. Controlling the Gaze of Conversational Agents. In J. van Kuppevelt, L. et al (Eds.), *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems.* Kluwer, 2005.

[27] Hymes, D. Models of the Interaction of Language and Social Life. In J. J. Gumperz and D. Hymes (Eds.), *Directions in Sociolinguistics: The Ethnography of Comm.*, 35-77. Holt, Rinehalt & Winston, 1972.

[28] Imai, M. et al. Robot mediated round table: Analysis of the effect of robot's gaze. *In Proc. ROMAN'02*, 411-416, 2002

[29] Kanda, T. et al. Body Movement Analysis of Human-Robot Interaction. *In Proc. IJCAI'03*, 177-182, 2003.

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

[31] Kuno, Y. et al. Museum guide robot based on sociological interaction analysis. *In Proc. CHI'07*, 2007.

[32] Libby, W. L. Eye contact and direction of looking as stable individual differences. *J. Exp. Research in Personality*, 4, 303-312, 1970.

[33] Matsusaka, Y. et al. Modeling of Conversational Strategy for the Robot Participating in the Group Conversation. *In Proc. EUROSPEECH'01*, 2001.

[34] Mutlu, B. et al. A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. *In Proc. of HUMANOIDS'06*, 2006.

[35] Nielsen, G. *Studies in Self-Confrontation*. Monksgaard, 1962.

[36] Rehm, M. and Andre, E. Where do they look? Gaze behaviors of multiple users interacting with an embodied conversational agent. *In Proc. of IVA'05*, 2005

[37] Sacks, H. et al. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50 (4), 696-735, 1974.

[38] Schegloff, E. A. Sequencing in Conversational Openings. *American Anthropologist*, 70 (6), 1075-1095, 1968.

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

[40] Sidner, C. L., et al. Where to look: A study of human-robot engagement. *In Proc. IUI'04*, 2004.

[41] Trafton, J. G., et al. Integrating vision and audition withing a cognitive architecture to track conversations. *In Proc. HRI'08*, 2008.

[42] Vertegaal, R., et al. Effects of Gaze on Multiparty Mediated Communication. *In Proc. Graphics Interface'00*, 92-102, 2000.

[43] Vertegaal, R. et al. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. *In Proc. CHI'01*, 2001.

[44] Weisbrod, R. M. Looking behavior in a discussion group. *Unpublished paper*, Department of Psychology, Cornell University, 1965.

[45] Wilkes-Gibbs, D. and Clark, H. H. Coordinating beliefs in conversation. *J. Memory and Language,* 31 (2), 183-194, 1992.

[46] Yamazaki, A. et al. Precision timing in human-robot interaction: Coordination of head movement and utterance. *In Proc. CHI'08,* 2008.

[47] Yoshikawa, Y. et al. Responsive robot gaze to interaction partner. In Proc. Robotics: Science and Systems, 2006.