"I'll Be There Next": A Multiplex Care Robot System that Conveys Service Order Using Gaze Gestures

KEIICHI YAMAZAKI, Saitama University
AKIKO YAMAZAKI, Tokyo University of Technology
KEIKO IKEDA, Kansai University
CHEN LIU, Saitama University
MIHOKO FUKUSHIMA, University of Miyazaki
YOSHINORI KOBAYASHI and YOSHINORI KUNO, Saitama University

21

In this article, we discuss our findings from an ethnographic study at an elderly care center where we observed the utilization of two different functions of human gaze to convey service order (i.e., "who is served first and who is served next"). In one case, when an elderly person requested assistance, the gaze of the care worker communicated that he/she would serve that client next in turn. In the other case, the gaze conveyed a request to the service seeker to wait until the care worker finished attending the current client. Each gaze function depended on the care worker's current engagement and other behaviors. We sought to integrate these findings into the development of a robot that might function more effectively in multiple human-robot party settings. We focused on the multiple functions of gaze and bodily actions, implementing those functions into our robot. We conducted three experiments to gauge a combination of gestures and gazes performed by our robot. This article demonstrates that the employment of gaze is an important consideration when developing robots that can interact effectively in multiple human-robot party settings.

Categories and Subject Descriptors: H.5.m [**Information Interfaces and Presentation (e.g., HCI)**]

General Terms: Human Robot Interaction

Additional Key Words and Phrases: Interaction analysis, gaze, mobile robot, multiparty interaction, embodied, human-robot interface

ACM Reference Format:
Keiichi Yamazaki, Akiko Yamazaki, Keiko Ikeda, Chen Liu, Mihoko Fukushima, Yoshinori Kobayashi, and Yoshinori Kuno. 2016. "I'll be there next": A multiplex care robot system that conveys service order using gaze gestures. ACM Trans. Interact. Intell. Syst. 5, 4, Article 21 (January 2016), 20 pages.
DOI: http://dx.doi.org/10.1145/2844542

This article belongs to Part 1 of the special issue on New Directions in Eye Gaze for Interactive Intelligent Systems. Its reviewing was managed by the special issue associate editors Yukiko Nakano, Roman Bednarik, Hung-Hsuan Huang, and Kristiina Jokinen. The editorial introduction to the special issue will appear in Part 2, which will be published in the next issue of *ACM TiiS*, Issue 1 of Volume 6.
This work was supported by JSPS Kakenhi Grant Numbers 23252001, 22243037, and 26240038.
Authors' addresses: K. Yamazaki and C. Liu, Graduate School of Humanities and Social Sciences, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama City, Saitama 338-8570, Japan; email: BYI06561@nifty.com, muguagirl@gmail.com; A. Yamazaki, School of Media Science, Tokyo University of Technology, 1404-1 Katakuramachi, Hachioji City, Tokyo 192-0982, Japan; email: akiko.a.yamazaki@gmail.com; K. Ikeda, Division of International Affairs, Kansai University, 3-3-35 Yamate-cho, Suita-shi, OSAKA 564-8680, Japan; email: osakakeiko@hotmail.com; M. Fukushima, Faculty of Engineering, University of Miyazaki, 1-1 Gakuen Kibanadai Nishi, Miyazaki, 889-2192, Japan; email: mihokofukushima@hotmail.com; Y. Kobayashi, Graduate School of Science and Engineering, Saitama University; email: yosinori@cv.ics.saitama-u.ac.jp; Y. Kuno, Graduate School of Science and Engineering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama City, Saitama 338-8570, Japan; email: kuno@cv.ics.saitama-u.ac.jp.

DOI: http://dx.doi.org/10.1145/2844542

1. INTRODUCTION

Two pressing issues await solutions in human-computer interactions: (1), to develop real-world–oriented interfaces that utilize existing computer systems to successfully enhance human activities in authentic contexts, and (2) such interfaces require systems that permit and enable the support of multiparty activities (i.e., more than one participant; the term “multiparty” can be used in the same sense as the conversation-analysis term, “multiuser”). Many attempts to facilitate multiparty activities in the virtual world, such as on Twitter or Facebook, have been made. However, this issue must be considered together with the first, that is, in real-world contexts. Our research group has been developing assisted-care robots for elderly persons in order to respond to a strong demand for them in rapidly aging societies like Japan. As robots should work in the real world among humans, consideration of the first issue is essential. Research on assisted-care robots offers an excellent opportunity to engage these two vital issues. Although our specific aim is to develop assisted-care robots, our findings from the study are also informative for human-computer interaction systems in a range of other settings and domains.

Several assisted-care robots that aid elderly care have been developed in the last 2 decades. For example, RI-MAN helps a person to get into bed [Mukai et al. 2008] and My Spoon assists the elderly with eating [Soyama et al. 2003]. These care robots are designed to assist a single user, one at a time, for a specific physical task. However, in order to develop care robots that can work in actual care facilities, we need to consider the issue of interaction in multiparty settings. In actual facilities, workers generally assist multiple people in multiparty situations. Needless to say, a care worker can perform only one physical task requested by one person at a time. Therefore, it is necessary for care workers to consider service order provision and convey that message to everyone. In addition to speech, humans convey necessary information to others through nonverbal communicative modes such as gaze and other gestures in these situations.

In order to assess the particulars of such interactions, we conducted an ethnographic study to observe in real-world settings how individuals coordinated their methods of responding and conveyed information to others. To this end, we carried out an experiment at an elderly care center and examined care workers responding and attending to the needs of large numbers of waiting clients via verbal and nonverbal communication.

Analysis of the observed patterns of interaction revealed that individuals not only employed various nonverbal behaviors to convey to multiple people information appropriate to each person but that certain nonverbal behaviors had multiple meanings. In other words, recipients received different messages appropriate for them depending on the situation and/or other behaviors communicated at the time. To take a particularly germane example crucial to our study, consider a case where an individual conveys a request while a care worker is already attending another client. By just turning his/her head toward the requester, the care worker indicates that he/she has recognized the request while conveying that the other client needs to wait. Moreover, that same gesture communicates to the client currently receiving service that the care worker will finish assisting him/her first.

Similar observations have been made with regard to “hierarchical display of orientation” [Kendon 1990] and “body torque” [Schegloff 1998] in the previous literature. Furthermore, in the experiment called TalkTorque, Kuzuoka et al. [2010] examined ways in which human body positions shift according to robot body positions and movements.

Based on our findings, we developed a care robot named MCR (Multiplex Care-Robot) that can simultaneously observe the actions of multiple people in a particular space while performing multiple actions. We performed experiments with this robot to examine human responses to the robot's nonverbal behaviors.

We videotaped the experiments and analyzed the video data. We asked all participants to complete a questionnaire immediately following the experiment. However, we did not solely rely on the self-report responses that gauged how the participants "remembered" their experiences with the MCR. In the process of remembering a past event, people do not simply read off a description from a memory storage device (i.e., access stored information in a manner analogous to a computer's information retrieval system). Rather, recollection and recall occur in social interactional contexts where accounts and descriptions are proffered. Furthermore, these recollections may come packaged with justifications or interpretations of the event. Our approach is thus clearly ethnomethodological and employs conversation analytic methods to unpack how such social activities like "remembering" or "doing a remembering" are performed as and in interaction [ten Have 2004]. With this in mind, we focused on a fixed point to observe the participants objectively and categorized their responses according to our criteria.

The results demonstrated that the MCR's nonverbal behaviors could effectively convey the robot's intention to humans. We also performed experiments to examine the effects of combining multiple nonverbal gestures. The results revealed that gaze behavior may be interpreted differently depending on how it is combined with other nonverbal behaviors—an outcome having important implications for designing robot actions.

Human-robot interactions in multiparty settings are well studied for cases involving museum guide and storytelling robots. For instance, Bennewitz et al. [2005] developed a robot that chooses an individual among a group and then turns its gaze toward that person after considering positions and frequency of speech obtained through facial image processing. Shiomi et al. [2007] demonstrated the effectiveness of a robot's explicit utterances indicating visitors to move to an appropriate position with respect to the robot. Yonezawa et al. [2008] developed a vision system that can detect gaze directions of multiple visitors simultaneously; they then presented a robot incorporating a system that deploys appropriate speech and actions depending on the gaze direction of people looking at an advertisement. Kobayashi et al. [2010] developed a museum guide robot that observes nonverbal gestures and selects an appropriate respondent to questions based on visitors' gazes. In contrast, Mutlu et al. [2009] have shown that nonverbal robot behaviors can influence the participation framework of multiple individuals. However, while these studies are important contributions, they do not consider the multiplex aspects of human and robot actions addressed in this article.

2. GAZE IN INTERACTION

Studies on gaze can be found in various disciplines. While the majority has been conducted within social psychology or kinesics, more recently attention has been directed toward the function of gaze in talk-in-interaction. In conversation analysis, Sacks et al. [1974] first identified an important function of gaze: it can enable a speaker to select the next speaker in on-going talk. In addition to speaker's gaze, Goodwin [1981] introduced listener's gaze into the picture; he proposed that a speaker should obtain recipient's gaze during the course of a turn of talk and that the recipient should be gazing at the speaker when the speaker is gazing at the hearer. Lerner [2003] points out that gaze-directional addressing (i.e., selecting the next speaker by using gaze) is vulnerable to the looking or glancing practices of recipients. Due to its limitation as a routine method for explicit addressing, gaze is most effective when selectively deployed and more popularly, more-than-addressing action, such as verbalization of address terms (e.g., names) accompanies the gaze.

Fig. 1. Nurses checking vital signs at an elderly care center.

Gaze is a key resource for interactions and affords the construction of projectability in interaction, that is, the ability to provide for anticipation of possible aspects of talk or social action, which are yet to come about [Streeck 1995]. In natural human-human interaction, gaze is never used separately from other bodily orientations. Kendon [1990] described a hierarchical organization of various bodily conducts:

> "Changes in orientation of upper-body segments, such as the head and eyes, are characteristically done within the framework of limits set by the position of the lower body. Changes in the orientation of lower-body segments, therefore, have longer term implications for change in attention than changes in the orientation of upper-body segments..." [Kendon 1990, p. 249]

This hierarchical organization of bodily conducts allows us to interact with more than one person occupying the same space in multiple ways. Schegloff [1998] calls such coordination between gaze and multiple bodily conducts "body torque." It is a state of bodily configuration when two different body segments (e.g., head direction and torso positioning) are oriented in different directions by different participants. Gaze, in this case, is an integral part of the whole embodied conduct in interaction.

3. ETHNOGRAPHY IN AN ELDERLY CARE CENTER

To understand interaction in multiparty settings further, we chose to study an elderly day care center as our site of investigation. We conducted an ethnographic study at a facility located in the countryside of western Japan (Nara). Over the course of 4 years (2008–2012), we recorded over 200 hours of footage. We then analyzed the material to identify how care workers coped with multiple requests from multiple elderly people at once. Under Japanese law, elderly care centers must employ one health worker for every 15 elderly persons. The following sequence illustrates a typical scenario.

In Figure 1, a nurse is checking and assessing the vital signs of several clients in succession by measuring their temperature, pulse rate, and blood pressure. While there are many different care workers at this facility, under Japanese law, only nurses can carry out such medical activities. Clients are waiting for permission to take a Japanese-style bath, and this health check is required because the depth of the tub and bathwater temperature can have an impact on the circulatory dynamics of elderly persons. The order in which the visitors had their health check was not systematic (e.g., first come

Fig. 2. N1 gazes toward E1. (E1 is outside of the image).

Fig. 3. N1 turns her body and attends to E1.

first serve), rather, the nurse attended to a client who was ready to have his/her vital signs taken.

In this situation, then, it is critical that the nurse displays service order to multiple clients. In doing so, gaze plays an important role, as Sacks et al. [1974] and Lerner [2003] have shown in the use of gaze-directional addressing to select the sequence of speakers in a turn-taking system.

By observing actual practice many times, we realized that the nurse's gaze played two important functions with regard to displaying service order. However, given that gaze can serve other functions, we isolated the two relevant types used in gaze-directional addressing and categorized them as Type A and Type B:

Type A: The gaze signals that the nurse will attend to a particular client next. The addressed elderly person then understands the nurse's projection that he/she will be attended to next in sequence.

Type B: The gaze signals that the nurse is asking a client to *wait* until she completes her current activity. The client understands the nurse's projection that he/she will be attended to next, once the nurse finishes her activities.

We see the nurse's initial stage of Type A gaze in Figure 1. The nurse (N1) is taking the temperatures of several clients. Figures 2 through 4 show how the next person in line gets served by the nurse. In Figure 2, N1 gazes at E1, then turns her body (Figure 3), and finally moves toward E1 (Figure 4).

Following this pattern, the nurse continues to turn her gaze and head toward one client after another. The clients can thus comprehend the timing of their respective turns by observing the nurse's bodily conduct.

Now let us discuss Type B, where the nurse employs her gaze to ask a client to wait until she finishes her current activities. Figure 5 shows the sequence of the nurse taking a client's blood pressure.

In Figure 5(b),the nurse finishes checking E3. She then turns her gaze toward the next client, E2 (Figure 5(c)). She then walks over to put a client's bag on the table (Figure 5(d)), before approaching E3 to check her condition (Figure 5(e)). Having received the projected request to wait, the second client (E3) quietly waits for the nurse to come to her. In Figure 5(f), we see that E2 is now attended by N2.

It is helpful to discuss how people distinguish between two types of gaze. In Figures 6 through 9, two clients (E5, E6) await their turns while the nurse (N2) attends to another client (E4).

Fig. 4. N1 attends E1.

As shown in Figure 6, N2 first hands a thermometer to E5 and tells her to check her temperature. Our focus of this example is E6. By this point, E6 has already been given a thermometer and has been checking his temperature with it. E6, observing E4 and E5, then waits for N2's display of service availability. Figure 7 shows N2 pausing with her hand in midair, motioning her readiness to assist E5 at any time. At the same time, N2 gazes toward E5 and begins to slightly lean forward toward her.

At this point, E6 says to N2, "Please check my temperature." Observing that N2's activity is on hold while E5 is engaged with an individual action (using the thermometer), E6 inserts his request to be served next.

Figure 8 shows N2 raising her upper body and gazing toward E6. This bodily conduct communicates that she has ended the previous action, which was to deal with E5. Upon this, E6 gazes toward N2, communicating his recipiency to her. Figure 9 shows E6 now nodding toward N2. When E6 talks to N2, mutual eye gaze was not established; thus, he did not obtain a certain recipiency from N2. But this time, E6 certainly acknowledges that N2 is addressing him. N2 says to E6, "Okay, the temperature has appeared on the display?" and walks toward him.

It is important to note here that N2 simply walks toward E6, leaving E5 without saying anything to her before leaving. This implies that all participants including E5 understand one activity has been completed (i.e., dealing with E5) and another activity has begun (i.e., dealing with E6). We can say that E6's request was in fact made spontaneously.

Once E6 is taken care of, E7, who is in a wheelchair left of E6, now initiates his participation into the scene.

In Figure 11, E7 communicates his anticipation for N2's attention by gazing at her and rocking his wheelchair. Immediately after this, N2 gazes at him to signal him to wait, also verbalizing the directive, "Please wait a moment" (Figure 12). The posture of N2 at this moment reinforces her projection. She places her hand on the desk in front of her and positions her torso toward E6 while gazing at E7.

The nurse's gaze and its direction are not markedly different in Figures 8 and 11. In Figure 8, the nurse communicates completion of her current activity and detachment from the current service recipient. However, in Figure 11, she communicates that she is still currently engaged in providing service to client E6 by using other nonverbal signals such as upper-body posturing and hand positioning to reinforce gaze.

Fig. 5. A sequence of nurse’s use of Type B gaze and bodily conduct.

We can observe the same phenomenon occurring in Figures 2 through 4. The nurse physically removes herself from the space near the recipient as she completes a previous service. At the same time, she gazes at the next recipient to communicate service provision (Figure 2). In contrast, the nurse’s gaze toward E2 in Figure 5(d) communicates that she is currently engaged in another activity (i.e., carrying E3’s bag to the table).

As we have shown in this section, the nurses incorporate gaze and other body gestures to manage multiple clients in one locale. Their methods function to make the provision of service order clear to those present so that those waiting would not be left puzzled or frustrated for not being attended. Based on these observations, it is clear that the difference between Type A gaze and Type B gaze is closely related to body motion and positioning, rather than the ways each gaze shift is made. Type A gaze is not only

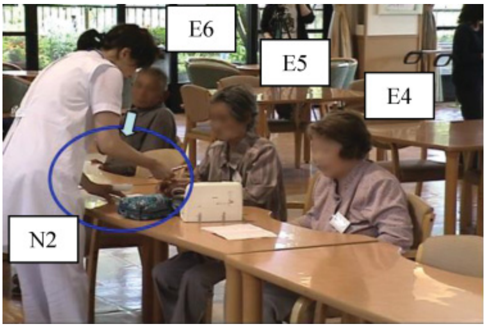

Fig. 6. N2 is looking for a thermometer. E6 is carefully observing her conduct.

Fig. 7. N2 hands a thermometer to E5 and then waits for her own action. E6 then says to N2, “Please check my temperature.”

Fig. 8. As E6 talks to N2, N2 raises her upper body and gazes at E6.

Fig. 9. N2: “Okay, the temperature has appeared on the display?” E6 nods as N2 walks toward E6.

Fig. 10. N2 collects the thermometer and writes down the temperature. E7 is carefully observing her conduct. N2 says “yes that’s good” while jotting down E6’s temperature.

about gaze movement to a particular person, it also involves simultaneous movement of the lower body turning in the same direction. This motion signals that the next movement will be an action toward the person in that direction. On the other hand, Type B gaze is made only with gaze. In other words, the other parts of the body indicate that the current activity is still going. Therefore, Type B gaze communicates to the gaze recipient that he or she needs to wait until the current activity is completed.

Fig. 11. N2 notices that E7 is shifting his wheelchair back and forth. She looks at E7 while her hands are still in the same position on the table.

Fig. 12. N2 stops writing, and says to E7, "Please wait a little while," without changing the position of her hands on the desk. N2 says to E7, "Please wait a little while."

Based on these observations, we designed our robot to perform these two types of gaze and a multiplex of bodily orientations and movements in tandem so as to effectively attend to multiparty settings.

4. DEVELOPMENT OF OUR ROBOT SYSTEM

In this section, we introduce our assisted-care robot design based on the analysis of human-human interactions in care facilities. We developed a mobile robot system to provide assisted care in nursing homes that is capable of circulating among multiple clients and handling multiple requests such as serving tea. The robot is designed to perform a specific action—bringing tea to a person who calls the robot by raising his/her hand. We call our robot system MCR (Multiplex Care-Robot). The development of our robot system and its autonomous functions has already been reported [Kobayashi et al. 2011a, 2011b]. Thus, in this section, we will limit our discussion to those areas related to the experiments.

Figure 13 details an overview of our robot system. We adapted a Segway mobile robot (RMP50, Segway Japan) as the mobile platform so that the robot could move smoothly. Eight joints were installed in the upper body to enable the robot to move its arms in ways similar to humans. The robot is equipped with a rotational joint inside its waist section so that it can change its upper-body orientation separately from its lower-body orientation. Controlling the gaze direction is important to perform the following experiment. We installed a pan-tilt mount (PTU-D46-17, Directed Perception) inside the neck for head motion. The pan-tilt direction can be controlled with 300-degrees/s speed and 0.05-degree angular resolution. Therefore, the robot head can behave like a human head. We employed a PC with Uninterruptible Power Supply (UPS), situated in the lower body of the robot, for processing sensor data and controlling robot motion.

The robot can recognize its position and orientation by receiving information from the base station, which observes the ellipse marker. For the localization system, we attached an ellipse marker on the robot's body atop a pole placed in its back. A Laser Range Finder (LRF; UTM-30LX, Hokuyo) was set at the marker's height in a fixed locale where the robot would work—the corner of the room in this case. The LRF was connected to another PC for observing the ellipse marker. The ellipse marker was in turn connected to the PC inside the robot through a wireless LAN. The localization system tracks the position and orientation of the ellipse marker using the particle filter on subsequent images, which are generated by plotting range data fed from the LRF. Thus, the system can obtain the position and direction of the robot and send

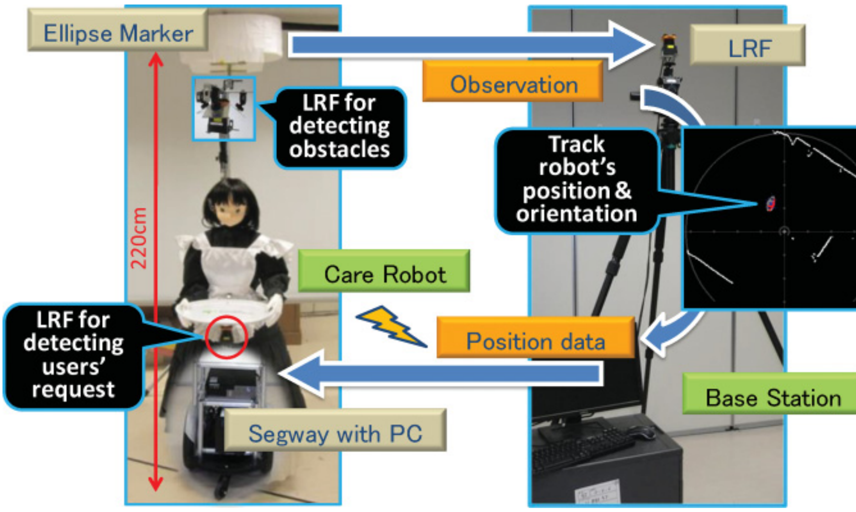

Fig. 13. Overview of our MCR robot system.

this information back to the robot. The robot then compares location information sent from the localization system with the position the robot is headed and determines the number of wheel rotations needed to complete the task.

In addition to localization provided by the LRF fixed in the room, the robot employs two other LRFs on its body to avoid obstacles and detect users' requests. The combination of these two sensors completes the autonomous function of our robot system. For more details, see Kobayashi et al. [2011a]. As for the robot's appearance, we used a skirt to hide the robot's mechanical parts. We chose a human face because it was critical to see where the face was looking.

5. ROBOT BEHAVIOR

It is essential for service robots to accept user requests, respond to them appropriately, and complete interactions successfully. In multiparty settings, in particular, it is especially important to communicate who will be served next. As discussed in the previous section detailing our ethnographic case study, people can indicate who will be served next through gaze. In addition, gaze can also simultaneously communicate service order to the next person waiting. This enables the next person to prepare for service provision. We expect that this interaction system would also work well in a case where a robot provided such services in a multiparty setting. Machines capable of demonstrating these abilities may encourage users to seek services from robots.

However, it is not enough for a robot working in multiparty settings to deal with requests from multiple users one by one. For instance, when the service robot concentrates on taking care of a specific person, other people who want to initiate interaction may feel frustrated with the robot. Similarly, when the robot finishes providing service to one person and moves to a second, a third may feel frustrated if given no indication of service order.

Therefore, we decided to adapt the Type B gaze function to our robot in order to make all users feel satisfied. In other words, for our experiment, we designed a robot that can maintain the direction of its torso and lower body toward the person to be served next and direct its face and gaze to the person to be serviced thereafter. Figure 14 serves as a good example of this tea-serving task. On the left, the robot concentrates on serving tea and snacks to a specific person. The next person who raises his hand to request

Fig. 14. The robot concentrates on serving tea and snacks to a specific person (left). The robot displays an acknowledgement to the next person (right).

service may feel that the robot is ignoring him, even if the robot recognizes his request through a sensor system. On the right, the robot turns its gaze toward the next person to display acknowledgement of his request. The next person will then sense that the robot will bring tea and snacks to him next, and that person is more likely to accept the service order provision. In addition, this kind of nonverbal behavior enables the third person to understand service order by observing the sequence of interaction between them. We, therefore, designed our robot to behave in this manner.

Some restaurants in Japan employ a waiter calling system to enable customers to call for service by pushing a button on the table. However, waiters complain that the system actually lowers the quality of service because they are constantly being called and, therefore, hindered in their ability to take care of individual customers. Meanwhile, at care facilities like the one we observed, care workers provide services while considering the feelings of elderly clients who may be hesitant to ask for assistance. Therefore, it is important to develop a robot system that can deal with various situations through employing nonverbal behaviors.

For our project, we specified hand-raising as the sign for users to request service. It is a commonly used gesture to request service (at least in Japan) and appropriate for all those participating in the experiment. Eye contact is also a very common method to request service in natural settings; however, systematically it is very difficult for a robot to recognize eye contact in complex multiparty settings. While Miyauchi et al. [2005] designed a robot that can look toward and then approach a person who makes eye contact with it in a laboratory setting, a robot that can respond to a person's request through eye contact in naturally occurring multiparty settings has not been developed.

With that in mind, we focused on serving tea as the robot's task. We designed a robot exhibiting behaviors that would let users feel more satisfied even when it could not immediately address their needs. Specifically, the behaviors our robot performs are:

(1) When not servicing an individual, the robot circulates among multiple persons conveying availability (the robot looks around).
(2) When the robot detects a request, triggered by hand-raising, the robot moves toward that individual to serve tea.
(3) If the robot detects another person's request while approaching the first person, the robot conveys acknowledgement by momentarily stopping and turning its gaze toward the second person.
(4) After performing this action, the robot continues to move toward the first person. After serving the tea, the robot then moves to the second person.

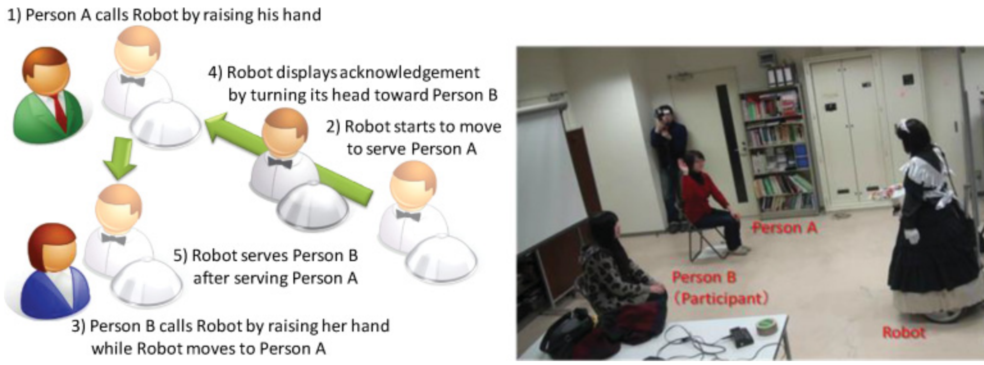

Fig. 15. The robot movement in Experiment 1 (left). An example scene from Experiment 1 (right).

As previously discussed, we found that gaze during a task can convey the message that people need to wait their turn, that is, Type B gaze. Robot behavior #3 (moving toward the first person to serve tea) can be interpreted by others as the physical task undertaken by the robot. Therefore, the robot's gaze while performing behavior #3 should be recognized as Type B gaze. In this way, people will likely feel more satisfied with the robot even if they need to wait because service order is conveyed to everyone in the locale via nonverbal behavior.

Although we confirmed that our robot worked autonomously in such situations, we applied Wizard of Oz experiment method (Woz) in some parts of the experiments. The movement of the robot is autonomously controlled based on the positions of the user predetermined in advance. For recognizing a user's request, we applied WoZ to avoid recognition errors. The operator manually input the command to the system when the user raises his or her hand. Turning the direction of the robot's head is autonomously computed and controlled based on the position of the user with respect to the robot. Other robot behaviors such as hand motions are performed by playing prerecorded motions.

We analyzed the ways in which participants reacted to the robot from our video recordings. We also conducted a survey that gathered participants' accounts of the activity. In this article, we focus on detailed video analysis rather than the survey results. As pointed out in the study on video analysis by Heath et al. [2010], it is necessary to analyze activities that occurred during and at the site of the experiment.

Experiment 1. As discussed in the previous section, gaze movement of the nurse (Type B gaze) functioned as a signal by the nurse to ask a client to wait until she finished her current task. Moreover, the client could immediately understand this nonverbal message and know that the nurse would attend to him or her when the current task was completed.

We contend that adapting this gaze behavior is also important for a robot system in multiparty settings. Therefore, we conducted experiments to confirm whether our robot's gaze movement could convey service order and politely indicate that participants needed to wait for service. In the experiments, we observed and compared the responses of two groups of participants. Each group experienced gazing and nongazing robot behaviors, respectively.

Setting for Experiment 1. We devised a situation where the robot (MCR) delivered snacks to two people in a room. As detailed in Figure 15(left), the robot performed the following behaviors:

(1) Person A calls the standing robot by raising his/her hand.
(2) MCR starts to move to deliver snacks to Person A.

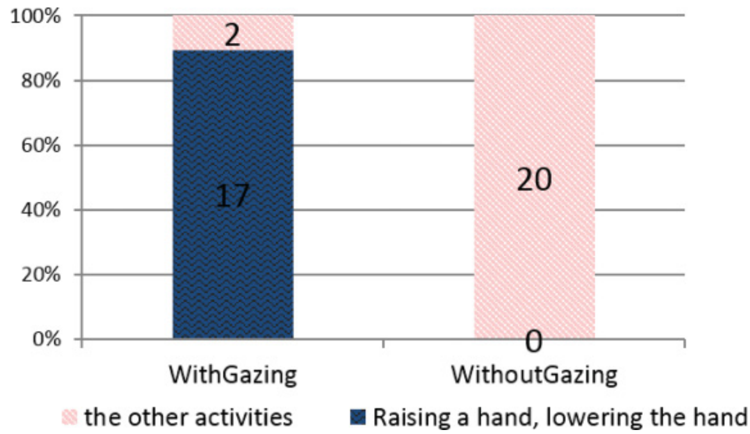

Fig. 16. Classification of participants' behavior when they raised their hands.

(3) Person B calls MCR by raising his/her hand while the robot is on its way to Person A.
(4) MCR displays acknowledgement by turning its gaze toward Person B who raised his/hand, whereupon it continues toward Person A.
(5) After serving Person A, MCR serves Person B.

Here, Person A is a virtual user acting as a control within the experiment. Person B was our target subject who only knew that the robot would deliver snacks when called. Person B was instructed to call the robot by hand-raising once the robot started to move.

In order to confirm the effectiveness of gaze movements, the robot demonstrated two types of motion sets. For the first set, the robot stopped momentarily and turned its gaze toward Person B when Person B raised his/her hand to seek service. For the second set, the robot did not stop and did not turn its gaze when Person B raised his/her hand to seek service. To confirm the effectiveness of the Type B gaze function, robot behavior #4 was left out of the latter set. The experiment consisted of 40 participants (10 male, 30 female; age average 20.6 (SD = 1.0)). All of the participants were undergraduate students at Saitama University in Japan. Each participant performed the role of Person B once. We divided the participants into two groups, with each group experiencing one type of robot behavior explained above. Because one participant raised his hand at the same time as Person A raised her hand, we had 39 sets of successful data. As shown in Figure 16(right), two people sat in chairs at the same distance from the robot (about 5m). Person A sat 3m apart from Person B.

Results of Experiment 1. We analyzed the experiments from scenes videotaped by two cameras located at different viewpoints. We found that there was a difference in behavior between the groups of participants concerning how they called the robot. For the group that experienced the robot without gazing, we observed some participants making emphatic gestures such as raising their hand several times or waving their hands in a sweeping motion.

Furthermore, we classified the behavior of participants into two groups. One consisted of participants who raised their hands once and soon lowered them while the robot was moving toward Person A. The other group consisted of participants who were observed performing other behaviors, such as raising their hands several times or not lowering their hands for a while. The results, shown in Figure 16, reveal a clear difference. A Fisher's exact test shows that there are significant differences between the two cases (two-sided test p = 0.0000).

In the group that experienced robot gazing, 17 of the 19 participants raised their hands once and then lowered them shortly afterward. This result shows that most participants were content to wait without raising their hands again until the robot finished providing service to Person A. This implies that several successful recognitions were performed as detailed below:

(1) Participants recognized that the robot received their requests.
(2) Participants recognized that Person A would be served first.
(3) Participants recognized that the robot would serve them next.

However, nearly half of the participants of the group that experienced robot-gazing did not lower their hands for a while; in fact they seemed hesitant to do so. For Experiment 1, participants' behaviors were very difficult to clearly sort out. Consequently, we could not fully confirm if the recognition sequence was successfully performed by all participants.

The group that experienced the second set of robot motions (no gazing) stands in sharp contrast to the group that experienced robot-gazing. In fact, none of its 20 participants responded by raising one hand and lowering it shortly afterward. Seventeen participants raised their hands more than one time, while three participants continuously raised their hands. This clearly demonstrates that the three kinds of recognitions observed for the first group were not performed by the second group because the MCR did not convey meaning to warrant the intended response.

We performed further qualitative analysis to identify characteristic behaviors for this group. We observed some participants were making emphatic gestures, not only raising their hands several times or waving their hands about frequently, but also issuing utterances such as "Excuse me." We suggest that such behaviors indicate that participants were suspicious about the robot's competence resulting in their attempt to find other ways to get the robot's attention. This implies that participants were looking for some kind of salient response from the robot in order to confirm that their requests were being acknowledged.

Taking all these factors into consideration, we can infer that the robot incorporating gaze behavior was able to effectively convey awareness of user requests and successfully communicate service order while participants waited. This behavior clearly has an impact on users' perceptions by enabling users to generally feel more satisfied with the robot even when the robot cannot immediately address user needs.

Parts of this experiment and the analysis described thus far were briefly reported in Kobayashi et al. [2011a]. In the following section, we will discuss additional experiments we conducted which explored the effectiveness of robot behaviors in more detail.

Experiment 2. For participants in Experiment 1, a combination of motions (head turning and gazing toward them) was far more effective than making no motion to acknowledge their requests. In Experiment 2, in order to identify the most effective types of robot behaviors to recognize service order, we set the following conditions: three participants were included in a setting to test various responses to the robot. In addition to (i) turning its head (gaze-utilization), the MCR also employed several gestures, such as (ii) raising a hand toward participants or (iii) nodding. We compared participants' recognition of both individual and compound sets of robot gestures.

Setting for Experiment 2. In this experiment, the robot moved to the first person to initiate a request. When another participant raised his/her hand, the robot's torso stopped in its current position and the robot performed a target gesture or gestures using other parts of its body to acknowledge the request (Figure 17). There are five

Fig. 17. An example scene from Experiment 2.

Table I. The Functions of MCR's Behavior

Function 1 (Continuous Service)	Conveys continuous service provision toward the current user.
Function 2 (Request Acknowledgement)	Conveys acknowledgement of a user request.
Function 3 (User Identification)	Conveys identification of the user who requested services.
Function 4 (Next Person)	Conveys the person who is going to be served next.
Function 5 (Please Wait)	Conveys to the next user, "Please wait until service provision to the current user is finished."

categories of function which the robot carried out during the experiment (Table I). Table II shows how the MCR's gestures corresponded to these functions. A single asterisk indicates that the robot gesture in question performed this function, and the number of asterisks shows the number of functions a given gesture performed in total.

In the ethnographic observation, we found that it was not gaze (head movement) on its own but also simultaneous body movement concurrent with gaze that conveyed the critical information between nurse and clients. Therefore, we implemented seven patterns of gaze into the robot's system, which included a combination of head movements, nods, and other gestures. The seven patterns were (short titles for these patterns are

Table II. Combinations of Functions of MCR's Behavior

		Function 1	Function 2	Function 3	Function 4	Function 5
1	Moving toward 1st Person	*				*
2	Stopping		*			
3	Raising Hand		*	*		*
4	Nodding		*			
5	Gaze(Turning Head)		*	*	*	
6(case1)	1+2+3	*	**	*		**
7(case2)	1+2+4	*	**			*
8(case3)	1+2+5	*	**	*	*	*
9(case4)	1+2+3+4	*	* * *	*		**
10(case5)	1+2+4+5	*	* * *	*	*	*
11(case6)	1+2+3+5	*	* * *	**	*	**
12(case7)	1+2+3+4+5	*	* * * *	**	*	**

Note: A single asterisk indicates that the robot gesture in question can perform the function. Multiple asterisks denote the number of functions a given gesture can perform in total.

used in Table I below in parentheses): (1) robot raises its hand (hand), (2) robot nods (nodding), (3) robot turns its head (gaze), (4) robot raises its hand and nods (hand + nodding), (5) robot turns its head and raises its hand (gaze + hand), (6) robot turns its head and nods (gaze + nodding), and (7) robot turns its head, raises its hand, and nods (gaze + hand + nodding).

Like the aforementioned experiments, all of the participants were from Saitama University in Japan (68 male, 50 female; average age: 20.4 (SD = 2.8)).

As indicated in Table I, MCR performed five primary functions. And as indicated in Table II, by combining different gestures, which we categorized into patterns, MCR could perform additional functions. In this experiment, we tested all the patterns in order to examine the ways in which different combinations of gestures might affect participants' responses. We also wanted to see if the result of combining different movement patterns would lead to the creation of additional functions.

Results for Experiment 2. Each experienced just one pattern exclusively. We excluded some failed cases, such as participants incorrectly timing their hand-raising, or cases when a minor malfunction of the robot occurred (all such minor malfunctions were immediately repaired immediately). The significant number of cases was as follows: 18 cases of pattern 1, 18 cases of pattern 2, 14 cases of pattern 3, 18 cases of pattern 4, 17 cases of pattern 5, 16 cases of pattern 6, and 17 cases of pattern 7, for a total number of 118 significant cases.

Because Experiment 2 required programming several actions into the MCR, the duration of the robot's actions was longer than in Experiment 1. We categorized participants' responses, as shown in Table III. Figure 18 shows the summary of results of the experiment according to each category.

We analyzed how participants raised and lowered their hands in detail. One commonly observed response was to lower the raised hand when the robot stopped and made a gesture. We categorized this as "raising a hand, lowering the hand" (Figure 18). Sometimes participants lowered their hands gradually, pausing as the robot moved toward the first person. We called this "lowering the hand gradually with a pause." We observed participants keeping their hands raised until the robot reached the first person, which we identified as "raising a hand and keep it holding." Finally, there were two categories of participants raising their hands more than twice: "raising

Table III. Categories of Raising a Hand

Raising a hand once		
	⟶	Raising a hand, lowering the hand
	⟶	Raising a hand, lowering the hand gradually with a pause
	⟶	Raising a hand and keep it holding
Raising a hand more than twice		
	⟶	Raising a hand more than twice, lowering the hand
	⟶	Raising a hand more than twice, not lowering the hand

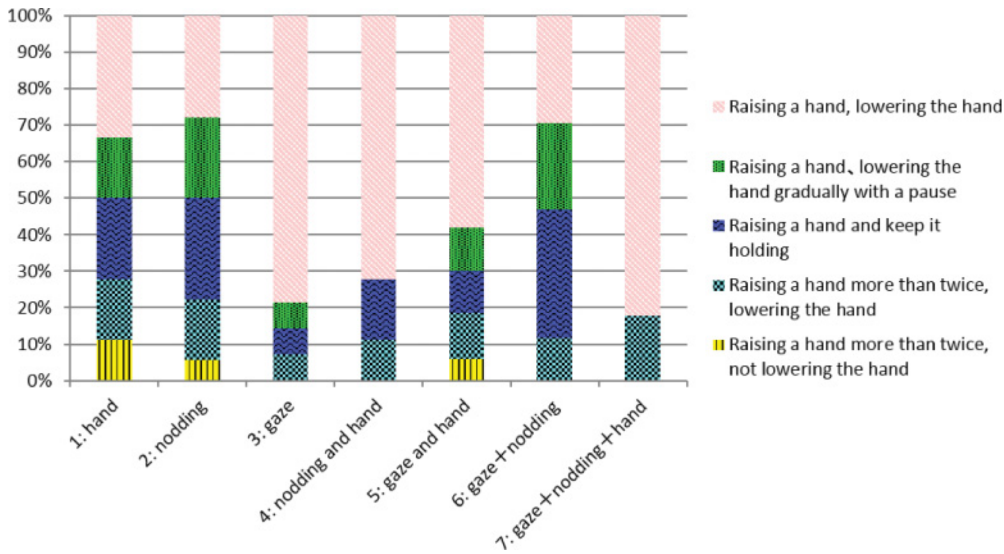

Fig. 18. Participants' responses to MCR's gesture patterns.

a hand more than twice, lowering the hand" and "raising a hand more than twice, not lowering the hand."

From these results, it is clear that participants lowered their hands when the robot turned its gaze toward them. These responses are similar to the results from Experiment 1.

We next analyzed participants' responses to the robot's gesture(s) by categorizing them into lowering their hands or using other responses. When the robot turned its gaze toward them (pattern 3), 80% of the participants lowered their hands. In this pattern, the robot kept its torso orientated toward the first person requesting its service, while it turned its head toward the second person. This performs the same function as the Type B gaze in multiparty settings discussed earlier. From this result, we can imply that the robot moving its head to direct its gaze effectively serves multiple functions, such as (1) displays that the participant's request is recognized, (2) displays that the first person to make a request will be served first, and (3) displays that the next person to receive service will be the participant in question.

In contrast, the robot raising its hand and nodding was relatively less effective than the robot turning its head (Figure 18). The reason for this is that these gestures do not serve all the functions that would be performed by the Type B gaze in human interaction. For our robot, turning its head was intended to perform the same practice. Hand-raising conveys that the robot recognizes the participant's request and communicates that the participant should wait (see Table II). However, it fails to carry out

Function 4 (see Table I) because it does not convey that the next person will necessarily be that participant. Nodding proved even less effective because it fails both Functions 3 and 4 because it neglects to specify an addressee of the performed gesture.

Combining gestures was not necessarily more effective either. When we included head-turning with other gestures, there were some interesting outcomes. We already observed that when the robot turned its head, 80% of participants lowered their hands and did not raise them again after that. However, when the robot turned its head and nodded, this number dropped to 30%. A Fisher's exact test shows that there are significant differences between the two cases (two-sided test $p = 0.0136$).

One of the reasons for this is that combined gestures overemphasize and therefore become identified with only Function 2 (the robot recognizes the request) and not Function 5 (Please Wait). Another reason is rooted in the robot's action sequence. Nodding after gazing at the participant is interpreted as the robot confirming the participant's request (conveyed by gaze and accepted by nodding) in sequential order. This sequence—request recognition and the display of an addressee followed by acceptance through gazing—constitutes a form of "adjacency pair" in conversation analysis, specifically a "request-acceptance" sequence. The sequence projects that the participant's request will be accepted and that they will be served immediately at that point, regardless of the presence of the first person. In other words, we can imply that participants may have interpreted the robot's gaze as Type A and not Type B. Examining the results for pattern 6 (gaze + nodding) in Figure 18, there were responses such as gradual lowering of hands or keeping hands up, and this suggests that participants misinterpreted the robot's behavior. Participants may have kept their hands up because the robot projected that it would come to them but then proceeded to the first person.

To compare the participants' responses, a one-way ANOVA was used to test for differences among three single gestures of the robot. The participants' responses differed significantly across the three gestures, $F(2,49) = 5.54$, $p = 0.007$. Tukey-Kramer post-hoc comparisons of the three gestures indicated that "gaze" was raising a hand and lowering a hand significantly more than "hand" ($p = 0.022$) and "nodding" ($p = 0.009$). Furthermore, we compared gaze independently with three types of combined cases (nodding + gaze, hand + gaze, and hand + nodding + gaze) by conducting a one-way ANOVA test. The result shows that the differences were significant, $F(3, 63) = 4.23$, $p = 0.009$. Tukey-Kramer post-hoc comparison shows that gaze itself and the hand + nodding + gaze combination was significantly different from the nodding + gaze combination ($p = 0.011$). From this result, we can see that the nodding + gaze combination inhibit the participants' hand-raising and lowering reaction more strongly than the other two (gaze independent and hand + nodding + gaze). We can say that gaze has a stronger function to order the action sequence than the other gestures. We can also say that the gaze + nodding combination (and this order of behavior, not the other way around) brings out a rather unwanted response from the participants (i.e., wrong interpretation about who is to be served next).

We would like to note here that the effect of gestures and combined gestures may also differ in relation to culture, gender, and age, which must be further investigated.

Experiment 3. Another reason is rooted in the robot's action sequence. We conducted a third experiment in order to clarify the matter of action sequence. This trial also took place at Saitama University with 63 students as participants (39 male, 24 female; average age 20.8 (SD = 2.1)). In this experiment, in order to find out if the different order of the robot's behavior solicited different reactions from human participants, we implemented three patterns of movements to the robot: (i) turn&gaze→stop→nod, (ii) turn&gaze→nod (without stopping), and (iii) nod→turn & gaze (Table III).

Figure 19 shows the result. Pattern iii (nod →turn&gaze) was the most effective sequence to trigger participants' reaction (lowering the hand from raising a hand). On

Table IV. Summary of Experiment 3

(i) turn&gaze→stop→nod	13 participants
(ii) turn&gaze→nod (without stopping)	13 participants
(iii) nod→turn&gaze	13 participants

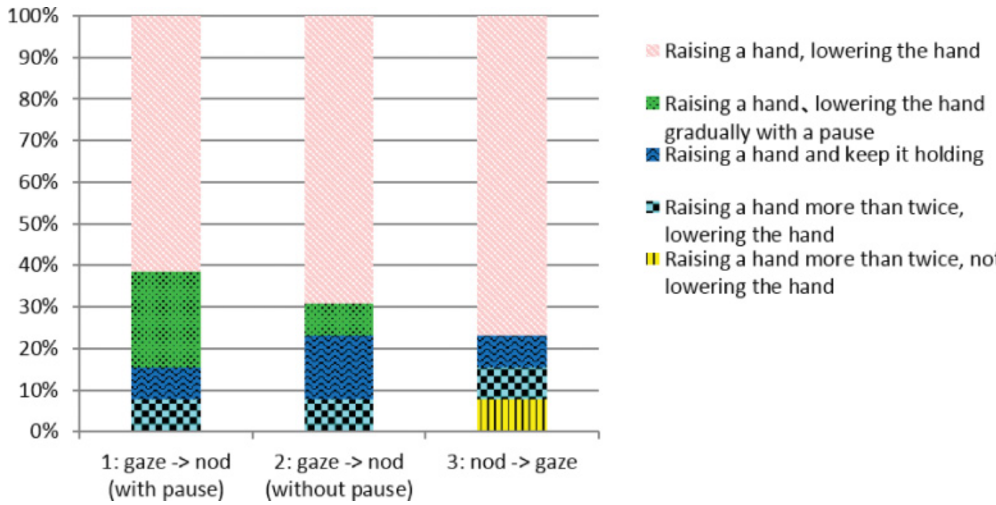

Fig. 19. Comparing three responses.

the other hand, with the pattern (i) turn&gaze→stop→nod and (ii) turn&gaze→nod (without stopping), the participants tended to lower their hands gradually with a pause from the hand-raising position. Such a sequence of "first gaze then nod" might have increased the possibility of misunderstanding that the robot will serve immediately toward the person who raised a hand secondly. It has become clear that the sequence of activities should be considered as a key factor in the experiment.

6. CONCLUSION

In this article, we examined multiple functions and features of human gaze in multiparty settings through an ethnographic study. We then outlined the potential benefits and challenges of implementing these functions and features in a robot. Our findings are significant with regard to the design of computer systems and computer interfaces at large. More specifically, however, we identified limitations with certain aspects of multiplex human activities and recommend the need to design systems to effectively support them. Our findings also suggest that combined individual activities may create entirely different functions from when they were performed individually. Therefore, when designing systems that can be implemented to suit the multiflexibility of human activities, we must simultaneously identify all the functions of these activities through experimentations.

This is important to keep in mind when designing embodied computer-human interfaces like robots. In designing robots to serve multiparty settings, simply mimicking human behavior is insufficient. Instead, research that utilizes detailed ethnographic studies of multiparty settings must take place first. Then we can move on to the implementation of required functions and test these in experimental settings.

Thorough experimentation with robots can identify the effects of and differences between various combined gestures. Such investigation holds ample possibility to reveal complex aspects of human behavior. Our project, which is a joint venture among sociologists and engineers, addresses issues that are important to both fields. Such intellectual adventures can bear fruit in designing more effective systems and gaining a better understanding human nature.

REFERENCES

Maren Bennewitz, Felix Faber, Dominik Joho, Michael Schreiber, and Sven Behnke. 2005. Towards a humanoid museum guide robot that interacts with multiple persons. In *Proceedings of the 5th International Conference on Humanoid Robots*. IEEE, 418–423.

Samuel R. Delany. 2002. *Babel-17/Empire Star* (1st Vintage Books Ed Edition). Vintage Books, New York.

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

Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research: Analyzing Social Interaction in Everyday Life*. Sage Publications, London.

Adam Kendon. 1990. Behavioral foundations for the process of frame-attunement in face-to-face interaction. In *Conducting Interaction: Patterns of Behavior in Focused Encounters*, Adam Kendon (Ed). Cambridge University Press, Cambridge, 239–262.

Yoshinori Kobayashi, Takashi Shibata, Yosuke Hoshi, Yoshinori Kuno, Mai Okada, and Keiichi Yamazaki. 2010. "I will ask you": Choosing answerers by observing gaze responses using integrated sensors for museum guide robots. In *Proceedings of the 19th International Symposium on Robot and Human Interactive Communication (RO-MAN'10)*. IEEE, 687–692.

Yoshinori Kobayashi, Masahiko Gyoda, Tomoya Tabata, Yoshinori Kuno, Keiichi Yamazaki, Momoyo Shibuya, Yukiko Seki, and Akiko Yamazaki. 2011a. A considerate care robot able to serve in multi-party settings. In *Proceedings of the 20th International Symposium on Robot and Human Interactive Communications (RO-MAN'11)*. IEEE, 27–32.

Yoshinori Kobayashi, Masahiko Gyoda, Tomoya Tabata, Yoshinori Kuno, Keiichi Yamazaki, Momoyo Shibuya, Yukiko Seki, and Akiko Yamazaki. 2011b. Considerate care robot which supports projectability of users in multi-party settings. *IPSJ J ournal* 52, 12 (Dec. 2011), 3316–3327 (in Japanese).

Hideaki Kuzuoka, Yuya Suzuki, Jun Yamashita, and Keiichi Yamazaki. 2010. Reconfiguring spatial formation arrangement by robot's body. In *Proceedings of the 5th International Conference on Human-Robot Interaction (HRI'10)*. ACM/IEEE, 285–292.

Gene H. Lerner. 2003. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society* 32, 2 (April 2003), 177–201.

Dai Miyauchi, Akio Nakamura, and Yoshinori Kuno. 2005. Bidirectional eye contact for human-robot communication. *Transactions on Information and Systems (The Institute of Electronics, Information and Communication Engineers)* E88-D, 11 (Nov. 2005), 2509–2516.

Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In *Proceedings of the 4th International Conference on Human Robot Interaction (HRI'09)*. ACM/IEEE, 61–68.

Toshiharu Mukai, Masaki Onishi, Tadashi Odashima, and Shiya Hirano. 2008. Development of the tactile sensor system of a human-interactive robot "RI-MAN." *Transactions on Robotics* 24, 2 (April 2008), 505–512.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 4 (Dec. 1974), 696–735.

Emanuel A. Schegloff. 1998. Body torque. *Social Research* 65, 3(Fall 1988), 535–596.

Masahiro Shiomi, Takayuki Kanda, Satoshi Koizumi, Hiroshi Ishiguro, and Norihiro Hagita. 2007. Group attention control for communication robots with wizard of Oz approach. In *Proceedings of the 2nd International Conference on Human Robot Interaction*. ACM/IEEE, 121–128.

Ryoji Soyama, Sumio Ishii, and Azuma Fukase. 2003. The development of meal-assistance robot "My Spoon." In *Proceedings of the 13th International Conference on Rehabilitation Robotics (ICORR'13)*. IEEE, 88–91.

Jürgen Streeck. 1995. On projection. *Social Intelligence and Interaction*, Esther N. Goody (Ed.), Cambridge University Press, Cambridge, MA, 87–110.

Tomoko Yonezawa, Hirotaka Yamazoe, Akira Utsumi, and Shinji Abe. 2008. GazeRoboard: Gaze-communicative guide system in daily life on stuffed-toy robot with interactive display board. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS'08)*. IEEE, 1204–1209.

Received December 2014; revised July 2015; accepted September 2015