Selective Function of Speaker Gaze Before and During Questions-Towards Developing Museum Guide Robots

Yoshinori Kobayashi
Saitama Univ.
255 Shimo-Okubo, Sakura-ku,
Saitama 338-8570 JAPAN
yosinori@cv.ics.saitama-u.ac.jp

Yoshinori Kuno
Saitama Univ.
255 Shimo-Okubo, Sakura-ku,
Saitama 338-8570 JAPAN
kuno@cv.ics.saitama-u.ac.jp

Takashi Shibata
Saitama Univ.
255 Shimo-Okubo, Sakura-ku,
Saitama 338-8570 JAPAN
shibata@cv.ics.saitama-u.ac.jp

Mai Okada
Saitama Univ.
255 Shimo-Okubo, Sakura-ku,
Saitama 338-8570 JAPAN

Yosuke Hoshi
Saitama Univ.
255 Shimo-Okubo, Sakura-ku,
Saitama 338-8570 JAPAN
hoshi@cv.ics.saitama-u.ac.jp

Keiichi Yamazaki
Saitama Univ.
255 Shimo-Okubo, Sakura-ku,
Saitama 338-8570 JAPAN

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
This paper presents a method of selecting the answerer from audiences for a museum guide robot. First, we observed and videotaped scenes when a human guide asks visitors questions in a gallery talk to engage visitors. Based on the interaction analysis, we have found that the human guide selects the appropriate answerer by distributing his/her gaze towards visitors and observing visitors' gaze responses during the pre-question phase. Then, we performed the experiments that a robot distributed its gaze towards visitors to select an answerer and analyzed visitors' responses. From the experiments, we have found that the visitors who are asked questions by the robot feel embarrassed when they have no prior knowledge about the questions and the visitor's gaze before and during the question play an important role to avoid being asked questions. Based on these findings we have developed a function for a guide robot to select the answerer by observing visitors' gaze responses.

Keywords
Human-robot interaction, ethnomethodology, service robot, non-verbal communication, computer vision

ACM Classification Keywords
H5.2. Information interfaces and presentation (e.g., HCI). User Interfaces – Interaction styles.

General Terms
Design, experimentation

Introduction
In recent years there has been an increasing interest in developing museum guide robots, which have the following two advantages over conventional PDA guide systems. First, robots can use visible actions such as gaze and point in addition to verbal actions. Second, robots can attend to multiple visitors simultaneously through such visible actions. Recent researche on this field focuses on how robots can combine visible actions with speech to effectively explain exhibits to multiple visitors ([1] [10] [12]).

We have also been working on developing museum guide robots, which has the following features. We videotape interactions between guides and visitors at actual museums and analyze them using interaction analysis developed in sociology. In particular, we focus on how visible actions such as gaze, head gesture and body turn are coordinated with speech. We then develop a guide robot that can coordinate visible actions and speech, and perform experiments to examine how visitors interact with the robot.

Through such interdisciplinary research combining ethnographic study and robot development, we have shown that the robot can increase the responses of a particular visitor by repeatedly turning its head towards the same visitor and asking him or her questions even when multiple visitors are present [11]. Kuzuoka et al. has shown that the robot can elicit the attention of visitors by deploying "restarts" and "pauses" at particular moments in its talk [6]. Although these studies have shown that robots can interact effectively with a person in multiparty settings, they have not tackled the essential problems in multiparty interaction, such as to whom it asks questions. In this paper we propose a museum guide robot that can attend to multiple visitors and ask questions of appropriate visitors. We draw out strategies of our robot's behavior by videotaping and analyzing naturally occurring gallery talk given at museums. To select the answerer, we develop vision techniques to observe people behaviors.

Analyzing gallery talk at an art museum
In recent years, many guides use questions for drawing out visitors' interest and promoting conversation and interaction. We focused on this trend and have found through ethnographic studies of gallery talk that guides pose questions to a proper visitor. When guides ask a question to multiple visitors, s/he takes a strategy of deploying gaze before s/he poses a question.

Conversation analysis has an interest in the relationship between gaze and question. Sacks et al. pointed out that one technique for selecting the next speaker is to pose a question to the intended next speaker [9]. When the current speaker selects a next speaker among multiple recipients, one of the embodied ways of selecting the next speaker is to gaze towards him/her. However, Lerner pointed out there is some limitation in assigning gaze to selecting next speaker. He observes that the function of gaze as selecting a next speaker works when the recipient is aware of the gaze of the speaker [8]. In other words, the speaker's gaze works only when a particular recipient notices that gaze. Thus, the gaze of the speaker does not select a particular hearer in a context-free way. Because of this, speakers may attempt to draw the attention of an intended recipient before posing a question. Based on these

concerns, we observed guide-visitor multiparty interaction in gallery talk at museums, focusing on guides' actions and gaze before posing a question, and how guides select the next speaker by using gaze.

In an example case (Figure 1), a guide is talking to three visitors about a Picasso and tries to extract their prior knowledge about Picasso. The guide inserts phrases such as, "I think everyone is familiar with, has heard the name Picasso," before uttering a complete question form, "What kind of Picasso work do you know? What comes to mind?" During these utterances, the guide distributes her gaze to each of three visitors. This distribution of gaze promotes every visitor's participation as a hearer, and allows the guide to monitor each visitor ([3] [4]). These inserting phrases and sentences before posing a question can be considered as preparative actions. By these preparative actions, a visitor can predict that a question will come, and can either prepare to answer the question or try to avoid an answer. In fact, when the guide finally asks a question as a complete form "What kind of Picasso work do you know?" M1 lowers his gaze, which does not allow direct eye contact with the guide. In contrast, F2 keeps gazing towards the guide. The guide gazes towards F2, after she gazes towards M1 and completes the question. In response, F2 answers "Guernica". Through thes preparatory actions, the guide get the visitors to project that a question is forthcoming. In addition, through these preparatory actions, the guide draws the visitors' gaze onher.
The guide then begins constructing a question, and then distributes her gaze towards multiple visitors. By distributing her gaze towards multiple visitors, she is able to observe the orientation of the visitors' gaze. The guide is able to select a visitor who is gazing towards her, rather than one who is looking away. Consequently, the guide's gaze becomes a resource for projecting a question for each visitor and also a resource for preparing to answer a forthcoming question.
In addition, we have found that guides use this type of gaze particularly in the case of questions involving prior knowledge. This may be because a question about prior knowledge becomes a test, and when the visitor cannot answer the question, s/he "fails" the test. In this case, issues of 'face' are coming up [2] [7]. Therefore guides do not fix their gaze on a single visitor, but ascertains who might be able to answer.

figure 1. The guide 'GE1' distributes her gaze towards multiple visitors while saying a question phrase.

In contrast, when the guide asks a question that does not require prior knowledge such as, "What kind of things are painted on this painting?", the guide may ask a particular visitor without distributing his/her gaze because every visitor can answer from their perspective, and there is ostensibly no wrong answer. Therefore issues of 'face' do not come up immediately.

Experiments and analysis

We have designed the robot to act with our strategy based on the aforementioned analysis and conducted experiments with it. We have programmed the robot that the robot first attracts visitors' attention, and then

the robot distributes its gaze towards all visitors one by one during the question. While completing a question, the robot finally fixes its gaze towards a particular visitor to ask a question. In the experiments, the robot explained the painting "Still life with a skull" by Pablo Picasso and asked visitors two questions. One is "What kind of Picasso's work do you know?" and the other is "What war is related to Guernica?" with inserting phrases. Participants need prior knowledge to answer both of questions. We used 27 participants, who were all students at Saitama University and were divided into 9 groups of three people. To avoid the case that no participant in a group has the knowledge, we asked a participant in each group to read a brochure about the painting before the experiments. We observed the participants using three video cameras to analyze interaction between the participants and the robot later. After the experiments, we asked all participants to fill in a questionnaire.

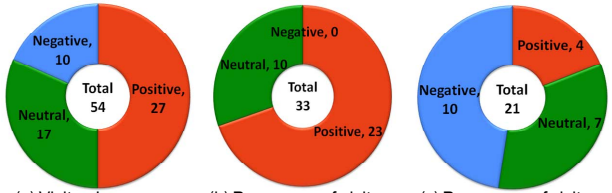
First, we analyzed the results of questionnaire. In the questionnaire, we asked the participants if they had prior knowledge about the questions and if they felt embarrassed when the robot asked them. We have found that the number of people who felt embarrassed is significantly larger for the participants without prior knowledge (41% of participants) than those with knowledge (9%). Many of participants who did not have prior knowledge mentioned "I felt embarrassed when the robot asked me" in the questionnaire. This result shows that participants may be embarrassed when asked questions if they do not know the answers. This confirms the importance for guide robots in asking questions to choose visitors who may be able to answer the questions.

Then, we examined the relationships between the visitors' gaze responses towards the robot's questions and their possession of prior knowledge. We categorized visitors' gaze responses toward robot's gaze distribution into the following three types.

- Positive responses: Mutual gaze/nod
- Neutral responses: Keeping looking at the painting
- Negative responses: Avoiding gaze from the robot

We categorized the gaze actions of 54 cases (2 questions for each of 27 visitors in 9 groups). We categorize the person who looks at the robot and does not avoid gaze as "mutual gaze." We categorize the person who avoids his or her gaze from the robot as "avoiding gaze from the robot." Figure 1(a) illustrates the number of visitors' positive, neutral, and negative actions. We examined differences in responses of visitors with or without prior knowledge. Figures 1(b) and (c) illustrate the differences of visitors' responses with or without prior knowledge. As shown in Figures 1, 70% (23 out 33) of the visitors with prior knowledge display positive responses, and none display negative responses. In contrast, in the case of visitors without prior knowledge, 19% (4 out of 21) display positive responses and 48% (10 out of 21) display negative responses. There is a clear difference in responses depending on the possession of knowledge.

These results suggest that a robot may choose an appropriate answerer who might be able to answer the questions by 1) selecting the visitor who displays positive response, 2) by not selecting the visitor who displays negative response. In the following section, we will describe our robot system that detects positive and negative responses.



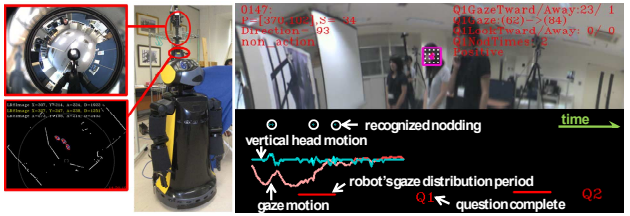
(a) Visitors' responses for the robot's question (b) Responses of visitors with prior knowledge (c) Responses of visitors without prior knowledge

figure 2. Visitors' responses with or without prior knowledge.

Guide robot system

Base on the findings so far, we have developed our guide robot system using a humanoid robot Robovie-R Ver.2 (ATR) which is developed as a research platform for human-robot communication. To capture multiple visitors' head movement in detail, we employ an omni-directional camera and a laser range sensor. Figure 3(a) shows an overview of our robot system. By using an omni-directional camera and a laser range sensor, visitors' heads are tracked by our proposed method [5].

To choose an answerer we recognize visitors' responses, as displayed in head nodding and gazing. To recognize visitors' nodding, we observe the vertical head motion during the robot's gaze distribution period. To recognize visitors' gazing, we observe the head direction which is estimated by our tracking framework [5]. When the direction of the head tends to move towards robot's direction during the robot's gaze distribution period, the system recognizes it as mutual gaze. On the other hand, when the direction of the head tends to move away from the robot's direction, the system recognizes that the visitor is avoiding gaze. The system only needs to observe the visitors' head motion only during the particular period in the robot gazing, which is drawn out from our experiments and analysis.



(a) Overview of our robot (b) Example of gaze response recognition

figure 3. Overview of our guide robot system and recognition of gaze response.

The example of the estimation of visitor's response is shown in Figure 3(b). The line graph of gaze motion represents the gaze direction of the visitor indicated by the rectangle in the upper image. Here the lower direction indicates that the visitor moves her gaze towards the robot and the upper direction indicates that she moves her gaze away from the robot. The line of vertical head motion shows the visitor's vertical head motion, which is used to detect her nodding. The small white circles in the graph represent the timing of nodding recognized. In Figure 3(b), the period of robot's gaze distribution and the timing of question completion are also indicated by the horizontal bars and characters such as "Q1" respectively. As shown in this figure, visitor's gaze motion and nodding are precisely measured and visitor's response is estimated before completing the question.

We examined the performance of recognition by applying our method to the stored data recorded in the experiments described in the previous section. Two specialists of conversation analysis categorized the responses into positive, neutral, and negative cases as the ground truth. Our system correctly recognized 70% (23 out 27) in the positive cases, 88% (15 out 17) in the neutral cases and 80% (8 out10) in the negative

cases. This is a promising result although we need further experiments and modification.

Conclusion

In this paper we proposed a method of choosing the answerer from audiences for a museum guide robot. We analyzed scenes when a human guide asks visitors questions in a gallery talk. We found that the human guide selects the appropriate answerer by distributing his/her gaze towards visitors and observing visitors' gaze responses before and during the question. We performed the experiments using a robot and found that the visitors' gaze before and during the question plays an important role to be asked questions. Based on these findings we developed functions for a guide robot to select the answerer by observing behaviors of multiple visitors. We are now planning to perform experiments in an actual museum to confirm the effectiveness of our strategy and robot system.

Acknowledgements

The authors gratefully acknowledge the members of Mitsubishi Electric Corporation for providing their expertise. This work was partly supported by JST, CREST and KAKENHI (21013009, 20700152).

Reference

[1] Bennewitz, M., Faber, F., Joho, D., Schreiber, M. and Behnke, S. Towards a humanoid museum guide robot that interacts with multiple persons. In *Proc. Humanoids '05* (2005), 418-423.

[2] Brown, P. and Levinson, S. Politeness: Some universals in language usage. Cambridge University Press, Cambridge, 1987.

[3] Goodwin, C. Conversational organization: Interaction between speakers and hearers. Academic Press, New York, 1981.

[4] Heath, C. Talk and recipiency: Sequential organization in speech and body movement. In *J. M. Atkinson and J. Heritage (eds.): Structures of Social Action: Studies in Conversation Analysis*, Cambridge University Press (1984), Cambridge, 247-265.

[5] Kobayashi, Y., Kinpara, Y., Shibusawa, T. and Kuno, Y. Robotic wheelchair based on observations of people using integrated sensors. In *Proc. IROS2009* (2009), 2013-2018.

[6] Kuzuoka, H., Pitsch, K., Suzuki, Y., Kawaguchi, I., Yamazaki, K., Kuno, Y., Yamazaki, A., Heath, C. and Luff, P. Effect of restarts and pauses on achieving a state of mutual gaze between a human and a robot. In *Proc. CSCW2008* (2008), 201-204.

[7] Lerner, G.H. Finding "face" in the preference structures of talk-in-interaction. *Social Psychology Quarterly 59*, 4 (1996), 303-321.

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

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

[10] Shiomi, M., Kanda, T., Koizumi, S., Ishiguro, H. and Norihiro, H. Group attention control for communication robots with Wizard of OZ approach. In *Proc. HRI2007* (2007), 121-128.

[11] Yamazaki, K., Yamazaki, A., Okada, M., Kuno, Y., Kobayashi, Y., Hoshi, Y., Pitsch, K., Luff, P., Lehn, D.V. and Heath, C. Revealing Gauguin: Engaging visitors in robot guide's explanation in an art museum. In *Proc. CHI2009* (2009), 1437-1446.

[12] Yonezawa, T., Yamazoe, H., Utsumi, A. and Abe, S. GazeRoboard: Gaze-communicative guide system in daily life on stuffed-toy robot with interactive display board. In *Proc. IROS2008* (2008), 1204-1209.