Museum Guide Robot with Communicative Head Motion

Yoshinori Kuno, Hiroyuki Sekiguchi, Toshio Tsubota, Shota Moriyama,
Keiichi Yamazaki, and Akiko Yamazaki

Abstract—**Face or head movement plays an important role in human communication. This paper presents a museum guide robot that moves its head to communicate smoothly with humans. We have analyzed the behavior of human guides when they explain exhibits to visitors. Then, we have developed a robot system that can recognize the human's face movement using vision. The robot turns its head depending on the human's face direction and the contents of utterances. We use the analysis results of human behavior to control the head movements. Experimental results show that it is effective for the guide robot to turn its head while explaining exhibits.**

I. INTRODUCTION

Face or head movement plays an important role in human communication [1]. Robots should also move their heads for smooth communication with humans. ROBITA [2] turns its head toward the person when it talks to him/her. The robot also moves its head toward the person when he/she starts talking to the robot. Robovie [3] and Cog [4] have the same function. However, humans move their heads in various other occasions. Sidner et al.[5] have investigated this further. They have examined the effect of tracking faces during an interaction. They have shown that people direct their attention to the robot more often in interactions when the robot makes head gestures.

We are developing a museum guide robot that can explain exhibits friendly and interestingly through the interaction with visitors. There were several museum guide robot projects [6]–[8]. They mainly focused on the autonomy of the robots and did not much emphasize the interaction part. Bennewitz et al. [9] have recently presented a humanoid guide robot that interacts with multiple persons. The robot can direct the attention of its communication partners towards objects of interest by pointing gestures with its eyes and arms. Although the research shows an important role of head motion for attention control, human guides may move their heads in various other occasions

so that they can keep interests of visitors and explain exhibits interestingly. For example, they may turn their heads towards the visitors to check if the visitors are following their explanation.

In this paper, we present a museum guide robot that moves its head communicatively. This is a joint work by researchers of robotics and sociology. We first investigate the behavior of human guides and visitors by the conversation analysis method used in ethnomethodology in sociology [1]. Then, we show a guide robot turning its head based on the analysis results. Finally, we report robot operation experiments performed in an exhibition at Science Museum, Tokyo.

II. HUMAN EXPERIMENTS

Before developing a guide robot, we observed behaviors of human guides in two occasions.

We performed the first experiment at Saitama University. A guide explained the exhibit showing the history of roof tiles in ancient Korea. The guide made an explanation of 15 minutes to a visitor four times and that of 30 minutes to a pair of visitors twice. The guide was a researcher on the exhibit and the visitors were students at Saitama University. We recorded the experiments with video cameras. Fig. 1 shows an experimental scene.

We performed the second experiment in an exhibition of photographs introducing Thailand at Future University-Hakodate. The guide was the photographer and he explained to a visitor three times and to a pair twice, each for about 30 minutes. The visitors were students at the university. We recorded the experiments with video cameras.

Fig. 1. Guide scene at Saitama University.

TABLE I
NUMBER OF CASES GUIDES TURNED THEIR HEADS IN THE TWO
EXPERIMENTS

	Number of occurrences
TRP (transition relevance place)	61
When saying keywords with emphasis	14
When saying unfamiliar words or citing figures	6
When using deictic words such as this	26
With hand gestures	41
When the visitors asked questions	12

Total 136 times. Counted multiple if multiple conditions are satisfied.

We extracted136 cases where the guides clearly turned their heads to the visitors from the video data. Table I summarizes the occasions of head movements of the guides.

Transition relevance places (TRPs) are the points where the speaker can allow others to take a turn [10] such as the point when finishing the explanation of a thing, and the point when asking a question. At these points, the guides looked at the visitors to know if the visitors understood them and to show that they could accept questions. Fig. 2 shows an example case. The left person is the guide and the right the visitor in the figures. The guide explained the process of making roof tiles and turned his face toward the visitor (Fig. 2 (a)). The guide confirmed that the visitor followed his explanation from the fact that the visitor moved his face toward the guide and then turned back to the exhibit. The guide proceeded to the next part of the explanation with his face turning back to the exhibit as in Fig. 2(b).

The guides turned their heads when they mentioned some keywords. They wanted to show the importance of the words by turning their heads.

The guides also turned their heads when they said unfamiliar words or cited figures. For example, in the second experiment, the guide did so when he used a word of Thai language. In these cases, they turned their heads to check if the visitors could understand the words.

The guides often turned their heads when they used deictic words and made hand gestures. These two actions often appeared together. The guides turned their heads to know if the visitors were looking at what they indicated. Fig. 3 shows an example case. The guide pointed his finger at a certain part in the exhibit (photograph) while saying "Food is placed like this." (Fig. 3 (a)) Then, he checked if the visitor was looking at the part where he pointed.

(a) The guide (left) turned toward the visitor (right).

(b) The guide turned back toward the exhibit to proceed to the next explanation.

Fig. 2. Example of the transition relevance place case.

In addition to the above cases initiated by the guides, there were cases initiated by the visitors. When the visitors asked the guides, the guides turned toward the visitors to show that they could accept questions.

III. MUSEUM GUIDE ROBOT

We have developed a museum guide robot that moves its head like the human guides while explaining exhibits. Fig.4 shows a photograph of the robot. The robot has two pan-tilt-zoom cameras (EVI-D100, Sony). We attach a plastic head on the upper camera. We use the pan-tilt mechanism of the camera to move the head. We do not use the images of the upper camera in the current implementation. The robot uses the images of the lower camera to make eye contact and to observe the visitor's face.

(a) The guide is pointing while saying, "This."

(b) The guide checks if the visitor is looking at the pointed part by turning his head toward the visitor.

Fig. 3. Example of the deictic word / gesture case.

Fig. 4. Guide robot. The robot consists of a mobile robot, Pioneer II by ActivMedia, a laptop PC, and two pan-tilt-zoom cameras.

Visitors can beckon the robot by making eye contact with it. When a visitor stands close to an exhibit and makes eye contact with the robot, the robot comes to a person, asking, "May I explain this exhibit?" Then, it starts explaining the exhibit. The actual eye contact process is as follows. The robot looks around with the lower camera to find a visitor who is looking at the robot. If it finds such a visitor, it turns its body toward him/her. If he/she is still looking at the robot, the robot considers that the visitor would like to make eye contact to ask for explaining the exhibit. This eye contact process is the same as the one used in our eye contact robot [11][12] except in that the current robot has a real head instead of a CG head.

We briefly describe the face image processing method used for eye contact. Our robot first searches for face candidates with the zoomed-out camera. When a candidate is detected, the camera zooms in on it. Then, the robot examines detailed face features.

Face candidate regions are detected in the images with a wide field of view. First, skin color regions are extracted.

Then, small regions and too elongated regions are removed. Inside the remaining regions, subtraction between consecutive frames is computed. The largest region among those where the sum of absolute values of the subtraction exceeds a given threshold is considered as a face candidate. Fig. 5 shows an example of a face candidate. Then, the pan, tilt, and zoom of the camera are adjusted so that the candidate region can be taken large enough to examine facial features. Experiments show that it can detect human faces indoors at a distance of 6 meters.

The system detects the eyes (pupils) and the nostrils in the zoomed-in image. We use the feature extraction module in the face recognition software library by Toshiba [13] for this process. Then, the system measures the horizontal distance between the left pupil and the left nostril dl and that for the right side dr as shown in Fig. 6. From these two values it determines the direction of the gaze (face). In actuality, the robot does not need to compute the accurate direction. It only needs to determine whether or not the person is looking at the robot. Since the camera has turned in the human's direction, the frontal face must be observed if the human is looking at the robot face. If the ratio between dl and dr is close to 1, the human face can be considered to be facing toward the robot.

This face direction computation process is also activated while the robot is explaining the exhibit.

Fig. 5. Face candidate.
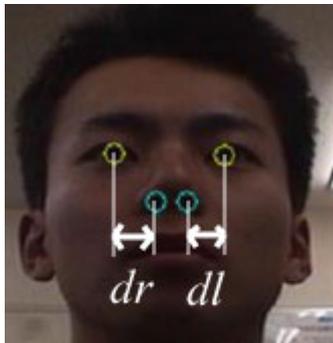
Fig. 6. Face direction computation.

The robot explains the exhibit by synthesized speech. During the speech, the robot turns its head toward the visitor at the points when the human guides often do so according to the human observation results. There are two cases of head turning: predetermined one and online one. The human experiments show that the human guides often turn their heads at certain points in explanation. We manually input the annotation marks for the robot to turn its head at such positions in the text of explanation. We call such cases as predetermined ones. In the current implementation, we choose the following points based on the human experiments.

1. TRP 1: At the end of a certain set of explanation.
2. TRP 2: When the robot asks a question.
3. When the robot says a keyword or unfamiliar word.
4. When the robot uses a deictic word to indicate something.

The online case is that the robot turns its head when it finds the visitor turning its head toward the robot. The robot can do this since it keeps observing the visitor's face direction all the time with the lower camera. The visitor may have a question or would like to say something in such cases. Thus, the robot turns its head toward the visitor, saying, "Any questions?"

The robot can obtain the movement of the visitor's face when it turns its head in predetermined cases. This information can indicate the visitor's response to the explanation by the robot. The robot should modify the explanation depending on the visitor's response. However, the current robot does not change anything and keeps its explanation. This is left for future work. In addition, the current robot cannot answer questions if the visitor asks them. Thus, we did not use the online head turning in the experiments described below.

IV. EXPERIMENTS AT A MUSEUM

We organized an interactive art exhibition using magnetic fluid by Sachiko Kodama and Minako Takeno at Science Museum, Tokyo from December 3 through 17, 2005. We demonstrated our robot on December 12 and performed experiments.

Sixteen visitors agreed to participate in our experiments (14 females, 2males, ages from 20 to 28, students and office clerks). When a visitor stood near the artwork named Morphotower, and made eye contact with the robot, the robot came to the visitor, explaining the work. The robot explained the work in two modes: the proposed mode where the robot turned its head to the visitor at predetermined points and the fixed mode where the robot watched the work all the time and did not turn its head. In the former mode, however, the robot did not use online head-turning, because the robot could not answer to the visitor in the current implementation if the visitor asked a question.

Eight participants first tried the fixed mode, then the proposed mode (Group A). The rest eight participants did in the reverse order (Group B). We placed about half an hour interval between the two trials. The participants were asked to look around the museum during the interval and not to see the experiments by the other participants. We did not tell the difference between the two trials to the participants. We videotaped the experiments. Fig. 7 shows an experimental scene.

After the experiments, we asked the participants which presentation mode you would prefer if the robot explained to you next time. For the participants of Group A, six preferred the proposed mode and two the fixed mode. These numbers, 6 and 2, were the same for the participants of Group B. The results suggest that humans prefer the head movements of the robot although we cannot conclude decisively, since the number of participants was small, and the genders and the ages of the participants did not vary much.

Fig. 7. Robot experiments at Science Museum, Tokyo.

To make quantitative evaluation, we examined when and how often the participants turned their heads. In the proposed mode, the robot moved its head 7 times for each trial at the predetermined points as follows.

1. When the robot comes to the visitor, the robot looks at the visitor. Then, the robot turns its head toward the work at this point while saying that it will explain the exhibit. The turning direction is different at this time alone from the other six points where the robot turns its head from the exhibit to the visitor.

2. The robot emphasizes the word "magnetic fluid".

3. The robot uses a deictic word "this".

4,5,7 TRPs: The robot finishes explaining a thing.

6. TRP: The robot asks a question.

Figs. 8 and 9 show the percentages of the participants moving their heads around each predetermined point for Group A and for Group B, respectively. In the figures, the horizontal axes indicate the time scale with the seven predetermined points. At time 1, the figures show the percentages of the participants who turned their heads from the robot to the exhibit. In other occasions, the figures show the percentages of participants who turned their heads from the exhibit to the robot. Fig. 10 shows the number of head movements of each participant in both mode cases. Although it may be a natural response for humans to turn their heads toward the robot when the robot turns its head to them, the experimental results confirm this. The larger number of head movements of the participants may not necessarily mean that the robot in the proposed mode is more user-friendly. However, considering the result that twelve participants preferred the proposed mode while four did the fixed mode, it can be an effective way for guide robots to turn their heads at the points when humans do so in explaining exhibits.
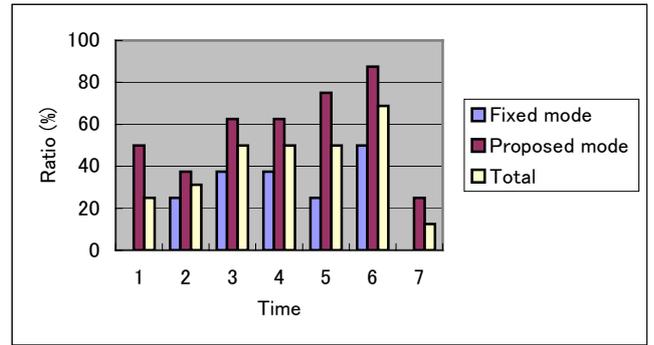
Fig. 8. Percentages of the participants moving their heads when the robot turned its head. (For the participants who tried the fixed mode first (Group A).)
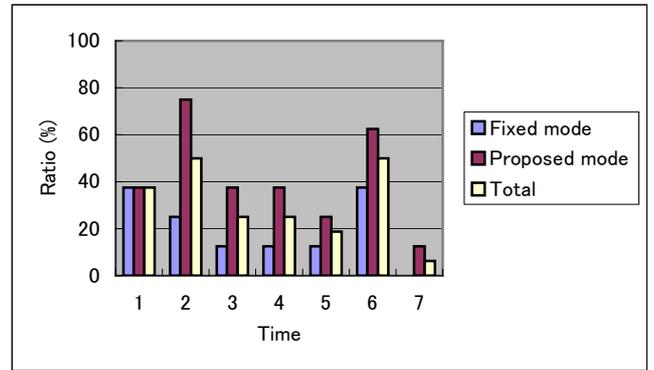
Fig. 9. Percentages of the participants moving their heads when the robot turned its head. (For the participants who tried the proposed mode first (Group B).)
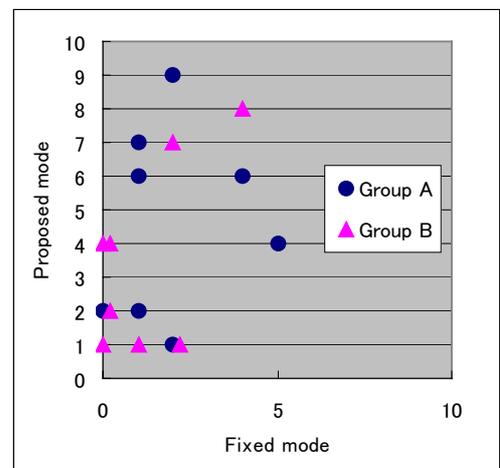
Fig. 10. Number of head movements of each participant in both fixed and proposed modes.

V. CONCLUSION

Humans move their heads effectively to achieve smooth communication. We have observed human guides to examine when to move their heads. Then, we have developed a museum guide robot that moves its head as the human guides. Experiments at a museum have shown a promising result.

However, the research is in the initial stage. We have used the mechanical part of a pan-tilt camera to move the robot's head. Thus, the robot cannot move its head so fast and subtly as humans. We have recently obtained Robovie-R ver.2 [14]. We are now working on implementing the head-turning method on the robot. We will perform more experiments with the robot to confirm the usefulness of head gestures. We will also examine the effects of other actions such as hand gestures and body movements. Furthermore, we would like to investigate how to modify the way of explanation in accordance with the observation results of the visitors.

REFERENCES

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

[2] Y. Matsusaka, S. Kubota, T. Tojo, K. Furukawa, and T. Kobayashi, Multi-person conversation robot using multi-modal interface, *Proc. SCI/ISAS 1999, vol.7*, pp. 450-455, 1999.

[3] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, Development and evaluation of an interactive humanoid robot "Robovie", *Proc. IEEE ICRA 2002*, pp.1848-1855, 2002.

[4] R.A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M.M. Williamson, The cog project: Building a humanoid robot, Nehaniv, C. (ed.), *Computation for Metaphors, Analogy and Agents, Lecture Notes in Artificial Intelligence 1562*, Springer-Verlag, pp.52-87, 1998.

[5] C.L. Sidner, C. Lee, C.D. Kidd, N. Lesh, and C. Rich, Explorations in engagement for humans and robots, *Artificial Intelligence*, *vol.166*, pp.140-164, 2005.

[6] I. Nourbakhsh, C. Kunz, and T. Willeke, The Mobot museum robot installations: A five year experiment, *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2003)*, pp. 3636-3641, 2003.

[7] R. Sigwart, K.O. Arras, S. Bouabdallah, D. Burnier, G. Froidevaux, X. Greppin, B. Jensen, A. Lorotte, L. Mayor, M. Meisser, R. Philippsen, R. Piguel, G. Rmael, G. Terrien, and N. Tomatis, Robox at Expo. 02: A large- scale installation of personal robots, *Robotics and Automation System*, *vol.42*, pp.203-222, 2003.

[8] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, J. Schulte, and D. Schulz, Probabilistic algorithms and the interactive museum tour-guide robot Minerva, *Int. Journal of Robotics Research*, *vol.19*, pp.972-999, 2000.

[9] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke, Towards a humanoid museum guide robot that interacts with multiple persons, *Proc. 2005 5th IEEE-RAS Int. Conf. on Humanoid Robots*, pp.418-423, 2005.

[10] H. Sacks, E. Schegolff, and G. Jefferson, A simplest systematics for the organization of turn-taking in conversation, *Language*, *vol.50*, pp.696-735, 1974.

[11] D. Miyauchi, A. Sakurai, A.Nakamura, and Y. Kuno, Active eye contact for human-robot communication, *Extended Abstracts CHI2004*, ACM Press, pp.1099-1102, 2004.

[12] D. Miyauchi, A. Nakamura, and Y. Kuno, Bidirectional eye contact for human-robot communication, *IEICE Trans. Inf. &. Syst. E88-D*, no.11, pp. 2509-2516, 2005.

[13] K. Fukui and O. Yamaguchi, Facial feature point extraction method based on combination of shape extraction and pattern matching, *Systems and Computers in Japan, vol. 29*, no. 6, pp.49-58, 1998.

[14] ATR Intelligent Robotics and Communication Laboratories, http://www.irc.atr.jp/productRobovie/robovie-r2-e.html.