

Museum Guide Robot Moving Its Head for Smooth Communication While Watching Visitors

Kazuhisa Sadazuka[†], Michie Kawashima[†], Yoshinori Kuno[†], and Keiichi Yamazaki[‡]

[†]: Graduate School of Science and Technology, Saitama University

{sadazuka, mkawashi, kuno}@cv.ics.saitama-u.ac.jp

[‡]: Graduate School of Cultural Science, Saitama University

yamakei@post.saitama-u.ac.jp

Abstract Face and head movement plays an important role in communication. We observe situations and human actions through vision to obtain information necessary for smooth communication. In order to develop robots that can coexist with humans, it is necessary to take into account vision. In this paper, we will discuss a museum guide robot that can move its head in a communicative way while explaining exhibits to visitors as well as detect human faces during the explanation of exhibits. We use the analytical results from human behavior to determine and program the robot head movements. We also developed a face detection system so that the guide robot can use visual information to conduct guiding. We developed the system with which the robot starts explaining after having detected a visitor's continuous gaze towards the robot based on the experiment result.

1. Introduction

Face and head movement play important roles in human communication [1]. Robots should also move their heads for smooth communication with humans. ROBITA [2] turns its head towards the person when it talks to him/her. This robot also moves its head towards the person when he/she starts talking to the robot. However, humans move their heads on various other occasions. Sidner et al. [3] 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 in a friendly and interesting way. There were several museum guide robot projects [4]–[6]. These mainly focused on the autonomy of the robots and did not much emphasize interaction. Bennowitz et al. [7] 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 through pointing and eye gaze. Although the research shows the important role of head motion for attention control, human guides may move their heads on various other occasions to keep the interest of visitors and to explain exhibits in an interesting way. 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 project between researchers in robotics and sociology. We first investigate the behavior of human guides and visitors through conversation analysis method used in ethnomethodology within sociology [1]. Then, we show a guide robot turning its head based on the analytical results. We also developed a robot that begins explaining an exhibit when it detects human gaze towards the robot, and the robot can ask a question when it receives continuous eye gaze from a visitor.

2. Observations from Guide-Visitor Interaction

Before attempting to develop a guide robot, we observed how human guides behave in two experimental situations. We performed the first experiment in our lab in which a guide explained an exhibit on the history of roof tiles in ancient Korea. The guide explained for fifteen minutes each to four different visitors. The guide was a researcher on the exhibit and the visitors were university students.

We performed the second experiment at Future University-Hakodate using an exhibition of photographs introducing Thailand. The guide was the photographer himself, and he explained to three different visitors. We recorded the experiments with video cameras.

Upon reviewing the video segments, we extracted 136 instances where the guides clearly turned their heads towards the visitor. Table I summarizes the instances of head movements.

Table 1. Number of cases guides turned their heads.(Total 136 times. Counted multiple if multiple conditions are satisfied.)

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

The guides made frequent head movements at transition relevance places (TRPs) –places in the talk where it is most appropriate for the listener to take a turn [8] such as at the completion of a sentential unit. Fig. 1 shows an example. The guide is on the left and visitor is on the right. Here the guide was explaining the process of making roof tiles. The talk and gaze direction are closely connected between the guide and the visitor. Here

the guide faces the visitor, which comes towards the end of a sentential unit.

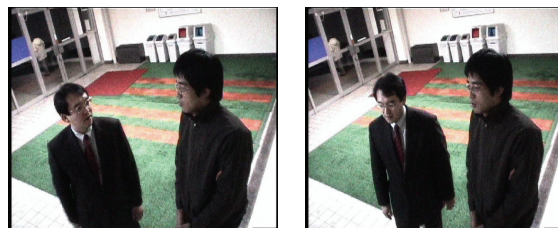


Fig. 1. Examples of the transition relevance place case.

In addition to TRPs, the guides also turned their heads towards the visitor when they said keywords. For example, in the experiment with Thailand pictures, the guide turned his head towards the visitor as he said the name of a ghost, which is a keyword in this explanation.

The guide is pointing and gazing towards the picture while explaining that the ghost lives inside the shrine. He marks the term “Pi” (name of the ghost) as new or unfamiliar information. While saying “Pi”, the guide turns his head towards the visitor. The guide’s gaze indicates an attempt to check the visitor’s understanding. At this point, the visitor also starts looking at the guide, and repeats the term “Pi” (with rising intonation). This repetition functions as a check for understanding. The visitor starts nodding as the guide confirms the term by repeating it. This exchange and mutual gaze during this interaction clearly display that the visitor registers the term as something new and significant in the guide’s explanation.

The guides often turned their heads and also made hand gestures when using deictic words. These two actions typically appeared simultaneously.

These experiments show that head movements occur at fairly predictable places within the talk of exhibit guides. In employing robots to do the work of guides at a museum, it may be important for a robot to conduct non-verbal behavior at interactionally appropriate points to create a more naturalistic interaction in general and a more personable robot in particular.

3. Prototype Museum Guide Robot

Based on the above findings from our guide experiment described above, we developed a prototype museum guide robot that moves its head while explaining exhibits similar to the human guides. Fig. 2 shows a photograph of the robot. The robot has two pan-tilt-zoom cameras. (EVI-D100, Sony). We attached a plastic head on the upper camera and used the pan-tilt mechanism of the camera to move the head. We did 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.



Fig. 2. Guide robot.

Visitor eye contact towards the robot may function as a request to be assisted. When a visitor stands close to an exhibit and makes eye contact, the robot approaches the person and starts the explanation. The actual eye contact process is as follows. The robot pans around with its lower camera to find a visitor who is gazing towards the robot. If it finds such a visitor, it turns its body towards him/her. If he/she is still gazing towards the robot, the robot assumes that the visitor might like an explanation of the exhibit. This eye contact process is the same as our eye contact robot [9, 10] except that the current robot has a head shaped figure instead of a Computer Graphic head.

Now let us briefly describe the face image processing method used for eye contact. Our robot first searches for a face candidate with the zoomed-out camera. When a candidate is detected, the camera zooms in. The robot then examines detailed facial features.

The candidate face regions can be detected in the images with a wide field of view. First, skin color

regions are extracted. Then, small regions and greater 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 a face candidate. Fig. 3 illustrates an example of a face candidate. Then, the pan, tilt, and zoom of the camera adjust 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 approximately 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 [11] 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. 3. From these two values it determines the direction of the gaze (face). The robot does not actually 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's face. If the ratio between dl and dr is close to 1, the human can be considered to be facing towards the robot. This same computation process is used while the robot is explaining the exhibit.

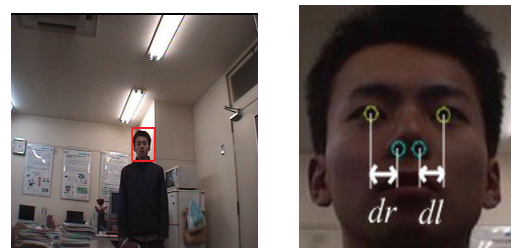


Fig. 3. Face image processing.

The robot explains the exhibit using synthesized speech. While speaking, the robot turns its head towards the visitor at similar points identified in the human guide experiment described above. There are two types of head motions: *predetermined* and *online*. The observations from the guide-robot

interactions showed that human guides often turn their heads at interactially significant points during the explanation. We manually inputted annotation marks for the robot to turn its head at such positions in the text of the explanation. We call such cases predetermined. In the current implementation, we chose the following points based on our earlier guide experiments.

- 1) TRP 1: At the end of a certain 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 refer to something.

In the online cases, the robot reacts to the visitor. In particular, the robot turns its head when it sees the visitor turning its head towards the robot. The robot is able to do this since it continuously monitors the visitor's face direction with the lower camera. In response to the visitor's head movement towards the robot, the robot responds by turning its head towards the visitor, saying, "Do you have 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 robot's explanation. The robot should be able to modify the explanation depending on the visitor's response. However, the current robot is not yet able to do this. This is a task for our future work. In addition, the current robot cannot answer questions if the visitor asks them. As a result, we have not implemented online head turning in the experiments described below.

4. 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 and 2 males, ages 20-28, students and office clerks). When a visitor stands near the artwork named Morphotower, and makes

eye contact with the robot, the robot comes close to the visitor and explains the work. The robot explains the work in two modes: the *proposed mode* in which the robot turns its head to the visitor at interactially significant points and the *fixed mode* in which the robot continuously gazes towards the exhibit without turning its head. In the former mode, however, the robot does not use online head turning, because the robot cannot answer the visitor when the visitor asks a question in the current implementation.

Eight participants participated in the fixed mode followed by the proposed mode (Group A). The other eight participants did so in the reverse order (Group B). We allowed about a half an hour interval between the two modes. The participants were asked to look around the museum during the interval and not to observe the experiments by the other participants. We did not tell the participants the differences between the two modes. We videotaped the experiments. Fig. 4 shows the experimental scene.

After the experiments, we asked the participants which presentation mode they would prefer if the robot were to provide an explanation again. For the participants of Group A, six preferred the proposed mode and two the fixed mode. These numbers, 6 and 2, are the same for the participants of Group B. The results suggest that while viewing the museum exhibits with a robot guide, visitors prefer robot head movements to no head movements, although the evidence is not decisive since the number of participants was small, and gender and ages of the participants did not vary a great deal.

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

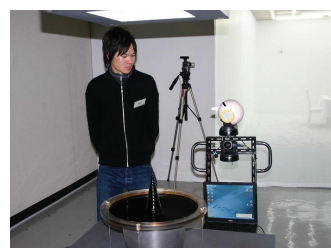


Fig. 4. Robot experiments at Science Museum.

1. When the robot approaches the visitor, the robot gazes towards the visitor. The robot then turns its head *from the visitor towards the exhibit* while saying that it will now explain the exhibit. At this time, head turning direction is different 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 the deictic word "this".
- 4,5,7. TRPs: The robot finishes explaining a point.
6. TRP: The robot asks a question.

Figs. 5 and 6 show the percentages of participants moving their heads in relation to each of the predetermined points for Group A and for Group B, respectively. In these figures, the horizontal axes indicate the time scale with the seven predetermined points. At the first predetermined point, both figures show the percentages of participants who turned their heads from the robot to the exhibit in response to the head movement of the robot. At other predetermined points, the figures show the percentages of participants who turned their heads from the exhibit to the robot.

Both figures show that the percentages of participant heads movements increase significantly in the proposed mode ($p < 0.01$, paired t-test). The significance still appears in both groups A and B, which suggests that the increase in head movement is not dependent upon the order of the two trials. In the fixed mode, the participants move their heads towards the robot at the point where the content of the explanation solicits the participants' attention toward the robot such as the time when the robot asks a question.

Fig.6 shows that the participants who experienced the proposed mode first gradually decrease their number of head movements, even though they are moving their heads in the beginning of the proposed mode. The participants also turn their heads towards the robot at the sixth point where the robot asks a question.

We recognize there is a possibility that it may be a natural response for humans to turn their heads

towards the robot when the robot turns its head towards them, and the larger number of head movements of the participants may not necessarily mean that the robot in the proposed mode is more user-friendly or personable. However, considering the finding that twelve participants out of sixteen conveyed a preference for the proposed mode, we suggest that it may be effective for guide robots to turn their heads towards the visitor(s) at interactionally significant points while explaining exhibits.

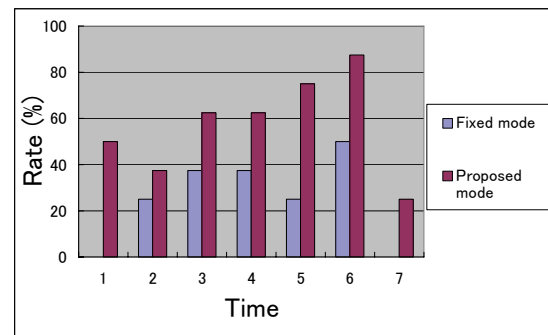


Fig. 5. Rate of participants moving their heads when robot turned its head. (Group A)

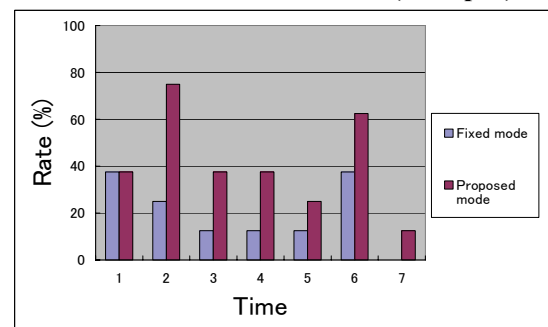


Fig. 6. Rate of participants moving their heads when the robot turned its head. (Group B)

5. Demonstration

We developed a guide robot with a monitoring camera on its chest, which detects human faces by integrating the results from our human experiments and prototype robot experiment. When starting its explanation and also during the explanation, the robot detects the human's face. When it detects gaze from a visitor, it asks whether or not the visitor has a question.



Fig. 7. Robot demonstration.

We conducted a demonstration at a workshop in the engineering department at Saitama University (Fig. 7.) When detecting the visitor's gaze for two seconds, the robot asks whether the robot starts the explanation after detecting continuous gaze from a visitor. While talking, the robot continues monitoring the visitor's face and calls for human assistance if it detects continuous gaze for three seconds towards a robot. On the other hand, if the robot recognizes that the visitor starts looking at the exhibition, the robot starts explaining again.

At this point of development, we are starting to work on a system that can be controlled remotely by human assistance so that the robot can answer a visitor's questions.

6. Conclusion

We have presented a robot system that we have developed upon considering the importance of vision and action within human-robot communication. Face and head movements play an important role in human communication. We have presented a museum guide robot that moves its head to enhance smooth communication with humans.

Even though our current robot can recognize only some mundane actions of humans such as gaze in simple environments, the robot should be able to detect human facial expressions in complex environments in the future. We need to improve the capability of computer vision. For example, if the robot can detect human facial expression and respond appropriate to a 'perplexed' face, the robot may be able to further assist visitor appreciation of exhibits in museum and other related venues.

This work was supported in part by the Ministry of Internal Affairs and Communications under SCOPE-S and JSPS New Research Initiatives for Humanities and Social Sciences (Cultural Policies and Museums in Japan).

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] 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.
- [4] 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.
- [5] R. Sigwart, K.O. Arras, S. Bouabdallah, D. Burnier, G. Froidevaux, X. Greppin, B. Jensen, A. Lorotte, L. Mayor, M. Meisser, R. Philippsen, R. Piguél, 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] ATR Intelligent Robotics and Communication Laboratories, <http://www.irc.atr.jp/productRobovie/robovie-r2-e.html>