International Conference on Control, Automation and Systems 2007
Oct. 17-20, 2007 in COEX, Seoul, Korea

Museum Guide Robot with Effective Head Gestures

Kazuhisa Sadazuka[1], Yoshinori Kuno[1], Michie Kawashima[1], and Keiichi Yamazaki[2]
[1]Graduate School of Science and Engineering, Saitama University
{sadazuka, kuno}@cv.ics.saitama-u.ac.jp
kawashima411@nifty.com
[2]Graduate School of Cultural Science, Saitama University
yamakei@post.saitama-u.ac.jp

Abstract We are currently developing a museum guide robot with an emphasis on nonverbal behavior in order to foster "friendly" human-robot interaction. In this paper, we discuss an interdisciplinary approach to robot guide development. We first examined head gestures of human guides through a sociological, conversational analytic approach, and discovered how human guides coordinate their head movements along with their talk when explaining exhibits. Second, we developed a robot system based on these findings. Third, we evaluated human-robot interaction through an experimental, sociological approach. The results suggest that robot head turning at interactionally significant points may lead to increased "engagement" of visitors with the robot. Based on this result, we have developed a guide robot that moves its head at interactionally significant points while monitoring visitor's head movements.
Keywords: Guide robot, human-robot interaction, nonverbal behavior, face direction, gaze

1. INTRODUCTION

We are currently working on an exhibit/museum guide robot with an emphasis on "friendly" human-robot interaction displayed through nonverbal behaviors. There have previously been several museum guide robot projects, e.g., [1], mainly focusing on the autonomy of the robots without emphasizing interaction with humans. In interacting with humans, it may be important for the purposes of presenting a "friendly" robot to utilize nonverbal behavior (e.g. head turning, pointing gestures). Sidner et al. [2] conducted research in a situation utilizing a guide robot that was designed to explain some innovative items. The research by Bennewitz et al. [3] presented a humanoid guide robot that interacts with multiple persons. This robot can direct the attention of its listeners toward objects of interest through pointing and gaze. In addition, Shiomi et al. [4] conducted a longitudinal study on human-robot interaction at Osaka Science Museum. The aforementioned studies, while very informative, have not examined how gestures and other body movements can potentially be coordinated with talk in human-robot communication.

Recent research has attempted to employ gestures in robots such as head movements in human-robot communication by studying human communication with a focus on gesture, head movement, and eye gaze. In particular, Sidner et al. [2] developed a penguin robot and examined how users responded to the robot under two conditions, both within the context of the robot explaining an exhibit: 1) the robot continuously gazes towards the visitor/user, 2) the robot moves its head and arms occasionally during the explanation. Under the second condition, user attention increased, as users responded to the robot's head movement and gaze direction by changing their own gaze and head direction. In another study, Breazeal [5] focused on emotion. Her work suggests the importance of nonverbal interaction between humans and robots. In summary, previous research on human-robot communication has suggested a link between robot head movement and gesture, and listener attention and response.

This paper describes an on-going collaborative research project by researchers in engineering and sociology. In a previous paper [6] we presented the results of an analysis of the behavior of human guides and visitors through experimental, sociological methods. We also showed preliminary experimental results regarding the effects of robot head turning using a simple prototype robot. In this paper, we report on further experiments using a humanoid robot. We then discuss a guide robot that we have developed based on these experiments.

2. PREVIOUS WORK

This section briefly describes our previous work [6]. As mentioned earlier, we began our project by observing human behaviors. In particular, we performed two experiments in which we asked guides to explain exhibits to visitors. We carried out a total of 11 sessions each 15-30 minutes in duration. We focused on points in the talk in which the guide turned his head towards the visitor. We identified 136 cases, and made detailed transcripts of these cases, using the methods of conversation analysis [7]. We then classified the cases (Table 1). In the table, a transition relevance place (TRP) is a point in ordinary talk at which a recipient response (such as a head nod or verbalization) becomes relevant [7], such as when the current speaker finishes a sentential unit. We found that TRPs are the most frequent point at which human guides turn their head and gaze towards the visitor.

Based on the results of this experiment, we developed a prototype museum guide robot that moves its head towards the visitor at TRPs. We attached a plastic head on a pan-tilt-zoom camera, and used the pan-tilt

978-89-950038-6-2-98560/07/$15 ©ICROS

mechanism of the camera to move the head.

We performed the following experiment using 16 participants. The robot explained an artwork in one of two modes: 1) *proposed mode*—the robot turns its head towards the visitor at predetermined points, and 2) *fixed mode*—the robot continuously gazes towards the exhibit without turning its head towards the visitor. In the *proposed mode*, the robot turned its head 7 times for each trial at predetermined points, whereas in the *fixed mode*, the robot turned its head 0 times. We quantitatively examined how often participants turned their heads: 1.6 times on average for the *fixed mode* and 4.1 times on average for the *proposed mode*. The number of head turns increased significantly in the *proposed mode* ($p< 0.01$, paired t-test).

Table 1. Number of cases that human guides turned their heads in the two experiments (Total 136 times. Counted multiple if multiple conditions were satisfied).

	Number of occurrence
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

3. EXPERIMENTS WITH HUMANOID ROBOT

Our previous experiment with the prototype robot suggests that participants move their heads more frequently when the robot turns its head (*proposed mode*) than when the robot does not turn its head (*fixed mode*). However, from this experiment we cannot be certain whether participants' head movements are driven by the robot's head movements at appropriate points or are simply mimicking the robot's motions. In order to probe this issue, we performed further experiments by using a robot Robovie-R ver. 2 [8], which can move its head more smoothly than the prototype robot. In these experiments, we prepared two modes for the robot motion: 1) *random mode*—robot turns its head at unsystematic points), 2) *proposed mode*—robot turns its head at interactionally significant points, such as at TRPs.

We performed a preliminary experiment [9] before attempting a larger scale experiment to discover any problems. The robot was programmed to explain two posters regarding our research projects. We had twelve participants, each of whom underwent both *random mode* and *proposed mode*. Half of the participants underwent the *random mode* first, while the other half underwent the *proposed mode* first. In addition, half of the participants underwent poster 1 first and then poster 2, while the other half underwent the opposite order. We videotaped the experiments as shown in Fig. 1.

(a) Facing towards the poster.

(b) Facing each other.

Fig.1 Experimental scenes.

The experimental results suggest the importance of robot head turning in eliciting visitor head turns. However, we found several points that needed to be modified. We performed a medium-sized experiment in which we modified the following.

1. Participants

In our first experiment with the humanoid robot, we used only 12 participants. To compensate for the small number of participants, we asked each participant to undergo both the *proposed mode* and *random mode*. This might have affected the result, although we considered the effect of order by dividing the participants into two groups. In the second experiment we invited 46 participants and asked each to undergo either the *proposed mode* or *random mode*, rather than both.

2. From synthesized speech to recorded speech

In the previous experiment, we used synthesized speech. Some participants commented that the speech had been a little hard to understand. In addition, the sociologists among us pointed out the importance of prosodic information such as intonation and pitch at TRPs. Synthesized speech cannot easily express such paralinguistic information. In the second experiment we recorded an explanation by a human female and used the recorded speech.

3. Standing position

The current robot moves exactly as programmed. Thus, even though the robot turns its head towards a visitor, the robot’s face may not be towards the visitor if the visitor is standing in a position different from the expected position. We found that participants who did not respond well to the robot in the previous experiment often stood at positions where they were not facing towards the robot when the robot turned in the visitor’s direction. In the second experiment we asked each participant to stand at a fixed point. This problem suggests that the robot should be able to detect the visitor’s position and turn in that direction. Although we are developing such a robot system, for now we asked the participants to stand at a predetermined position.

4. Exhibit

In the previous experiment we used posters depicting our research. We were concerned that the participants might not be able to understand the contents easily. In the second experiment, we used a photograph of a sub-tropical plant.

We performed an experiment based on the above guidelines. We invited 46 participants, and among them 24 (13 males and 11 females) underwent the *proposed mode* and 22 (12 males and 10 females) the *random mode*. All participants were Japanese graduate and undergraduate students, ages 19 to 24. Fig. 2 shows an experimental scene. The robot explained the sub-tropical plant (“air plant”) shown in the photograph. We videotaped the experiment and analyzed it in the same way as in the previous experiment. Participant responses to robot head turning were observed as follows:

A. only turning towards the robot,
B. nodding while continuing to look at the exhibit,
C. nodding after turning towards the robot.

Fig.2 Experimental scene in the second experiment.

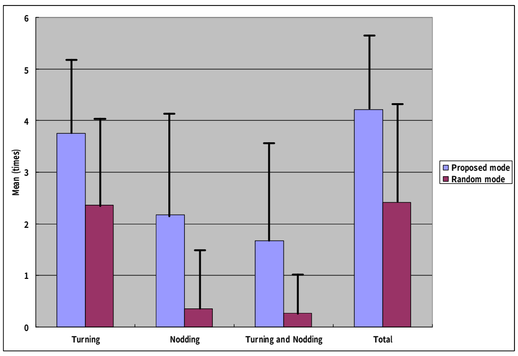

Fig. 3 Average number of responses.

In the figure, ‘Turning’ is the average number of A and C responses; ‘Nodding’ is the average number of B and C responses; ‘Turning and Nodding’ is the average of C responses; and ‘Total’ is A, B, and C combined. All these values are significantly different (t-test, $p<0.01$) (t-values, 2.90 for Turning, 3.79 for Nodding, 3.12 for Nodding and Turning, and 3.58 for Total).

Fig. 3 indicates that the averages in the *proposed mode* and *random mode* for both ‘Nodding’ and ‘Turning and Nodding’ are different. These results suggest that robot head turning may lead to heightened “engagement” of visitors with the robot. However, the standard deviations are rather large. Responses greatly vary among individuals. Some participants did not turn or nod towards the robot even in the *proposed mode*. However, the averages in the *proposed mode* are significantly larger than those in the *random mode*.

Sidner et al. point out that human responses like nodding may occur even though the robot is not programmed to respond to humans [10]. They note that it is not natural for the robot to nod as a reaction to a participant’s nodding, and developed a robot that can detect participants’ nodding and then respond by nodding back. The number of participant nods increased significantly in the experiment in which the robot nods followed participant nodding.

In our experiment, even though the robot is not designed to detect participants’ reactions, the robot can conduct more smooth communication by moving its gaze direction towards the visitor at points where he or she would normally nod in relation to the progressivity of talk, as well as towards the exhibit after the participant nods towards the robot. This finding suggests that participants treat the robot as if it is able to respond in accordance with participants’ nonverbal behavior.

4. GUIDE ROBOT SYSTEM

Based on the findings from the above experiments, we have developed a museum guide robot that turns its head towards visitors at interactionally significant points during the explanation. In addition, the robot continuously monitors human’s gestures with its camera system to respond with an appropriate action.

Fig. 4 shows the camera system consisting of three cameras on the robot’s chest. The human field of view is wide enough to notice another’s movements. Also, humans can typically move their heads rapidly without causing problems with vision. Since it is difficult to implement such a motion mechanism and a vision system in robots, we placed three fixed cameras on the robot’s chest. The robot can obtain stable images within a wide field of view. Thus, even when the robot is looking at the exhibit, the robot can detect a human face and head movements. We use an Open CV Haar Cascade Classifier [11] to detect human faces. Fig. 5 shows an example of face detection.

We demonstrated the robot at an exhibition held by the Graduate School of Science and Engineering, Saitama University, in November 2006. When the robot continuously detects a human face for two seconds, the robot asks, “May I explain the exhibit?” If the robot finds the same visitor looking towards the robot after this question, the robot begins the explanation. While talking, the robot continues monitoring the visitor’s face. If it finds that the visitor keeps looking at the robot for three seconds, the robot turns its head towards the visitor. If the visitor turns back towards the exhibit against the robot’s action, the robot also turns back towards the exhibit and continues its explanation. Otherwise, the robot asks, “Do you have any questions?” As shown in the bottom row of Table 1, human guides turn towards the visitor when the visitor asks them a question. Before and during the question, the visitor often continues looking at the guides. Natural language understanding technology, however, is not yet able to process and respond to questions from visitors. In the demonstration, the robot calls for human assistance, saying, “If you have a question, I will call a human guide.”

At this point of development, we are starting to work on a system that can be controlled remotely by human assistance so the robot can answer visitor’s questions. The robot works autonomously, and when the robot is asked a question it cannot answer, the robot turns into the remote control mode, as the robot presented in [12].

Fig. 4 Three cameras on the chest.

Fig. 5 Detection of a face.

5. CONCLUSION

Head movements play an important role in face-to-face communication. We have analyzed human head gestures through an experimental, sociological approach and have developed a prototype robot based on the results. We have also analyzed communication between this robot and human visitors. Based on these results, we have developed a museum guide robot that moves its head at interactionally significant places while monitoring the visitor’s head movements.

ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Internal Affairs and Communications under SCOPE-S.

REFERENCES

[1] I. Nourbakhsh, C. Kunz, and T. Willeke, “The Mobot museum robot installations: a five year experiment,” in *Proc. IROS 2003*, 2003, pp.3636-3641.

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

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

[4] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, “Interactive humanoid robots for a science museum,” in *Proc. HRI 2006*, 2006, pp. 305-312.

[5] C. Breazeal, “Emotion and sociable humanoid robots,” International Journal of Human-Computer Studies, vol. 59, 2003, pp.119-155.

[6] Y. Kuno, H. Sekiguchi, T. Tsubota, S. Moriyama, K. Yamazaki, and A. Yamazaki, “Museum guide robot with communicative head motion, ” in *Proc. RO-MAN 06*, 2006, pp.33-38.

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

[8] Intelligent Robotics and Communication Laboratories, http://www.irc.atr.jp/index.html.

[9] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki, and H. Kuzuoka, “Museum guide robot based on sociological interaction analysis,” *CHI2007 Conference Proceedings*, pp. 1191-1194, 2007.

[10] C.L. Sidner, C. Lee, L.P. Morency, and C. Forlines, “The effect of head-nod recognition in human-robot conversation,” in *Proc. HRI 2006*, 2006, pp.290-296.

[11] G. Bradsky, A. kaehler, V. Pisarevskky, “Learning based computer vision with Intel’s open source computer vision library,” *Intel Technology Journal*, vol.09, 2005.

[12] H. Kuzuoka, K. Yamazaki, A. Yamazaki, J. Kosaka, Y. Suga, and C. Heath, “Dual ecologies of robot as communication media: thoughts on coordinating orientations and projectability,” in *Proc. CHI 2004*, 2004, pp.183-190.