Evaluations of Interactive Guideboard with Gaze-communicative Stuffed-toy Robot

Tomoko Yonezawa, Hirotake Yamazoe, Akira Utsumi, and Shinji Abe
ATR Intelligent Robotics and Communication Laboratories
2-2-2 Hikari-dai Seika, Kyoto 619-0288, Japan {yone,yamazoe,utsumi,sabe}@atr.jp

Keywords

Stuffed-toy Robot Guide, Gaze-communicative Guideboard, Wide-area Gaze-tracking

Introduction

In this paper, we introduce and evaluate an interactive guideboard with a communicative stuffed-toy guide-robot that behaves in correspondence to the user's gazing direction. The proposed system adopts our remote gaze-tracking method, which estimates the user's gaze angles based on image processing. The main purpose of this research is to provide intuitive and natural guidance interaction through an information board system that functions as if it were with a human guide who cares not only about the user's active attitude but also about the user's gaze. The system is applicable to both normal and *disabled/unmotivated* people who do not/cannot make an utterance but need some detailed information from the guideboard. It is also able to provide detailed information on any particular region of the guideboard corresponding to the subconscious interest of the user determined by her/his gazing direction. Our proposed interactive guideboard system combines the effective features of voice guidance following *the user's gaze*, based on our remote gaze-tracking method, and the impact of *anthropomorphism* using various non-verbal expressions (Duffy, 2003, etc.), including gazing behaviors (Fukayama et al., 2002, etc.), while adopting a scheme of human-human communication (Kendon, 1967).

We built a tentative interaction model for a gaze-communicative guide robot that performs in the following three steps: a) *drawing the user's interest or attention*, b) *attracting the user with appropriate communications*, and c) *guiding the user to the detailed information corresponding to the direction of the user's conscious/subconscious gaze*. Based on this model's three stepwise interaction modes, we implemented the interactive guideboard with a gaze-communicative stuffed-toy robot. In this paper, we show the experimental results on the effectiveness of i) *the presence of the anthropomorphic guide robot* in the system, ii) *the attentive behavior* (turns of the robot's head) to the guideboard, and iii) *the gaze-corresponding guidance to the appropriate information based on both the direction of the robot's head toward the content and the information given by of the vocal guidance.*

Related Research

Many research efforts on gaze applications have considered the user's gaze as a conscious control interface for disabled people (Shi et al., 2007, etc.). In contrast to these works, we apply the user's gaze to detect both subconscious and conscious desires for further information; moreover, our approach does not require the user to wear any fixture.

On the other hand, there has been much research on communication robots that utilize gaze. Several studies have examined the social recognition/interactions between humans and robots related to gaze

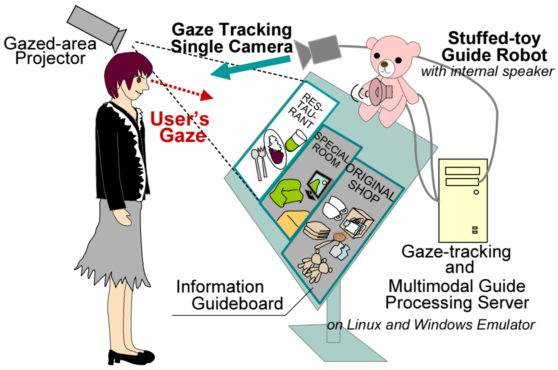

Figure 1. System Structure

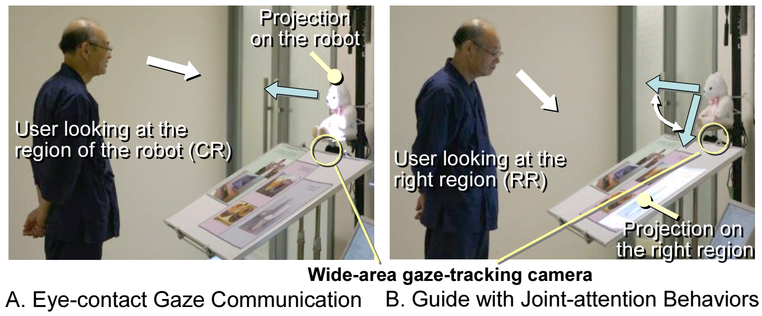

Figure 2. Operation View of Two-mode Guide System

(Breazeal et al., 2005; Sidner et al., 2005, etc.). These works primarily dealt with the facial and gazing behaviors of robots in imitative interactions based on a particular model. However, the appropriate application of joint attention, which is an important attentive component of gaze behavior, has not been sufficiently discussed.

Many research works have shown the importance of anthropomorphic agents in natural/intuitive guidance and assistance to people for various purposes, such as personal guide system (Sumi et al., 1998, etc.) Meanwhile, Katagiri et al. showed how the different behaviors of an agent differently affect the user's knowledge and performance (Katagiri et al., 2001). In this paper, we aim to adopt a kind of persuasive power of the robot's gazing behavior corresponding to the user's gaze and thus attracting the user's gaze toward an intuitive guide system based on our primary model of gaze communication (Yonezawa et al., 2007).

Structure of Gaze-communicative Guide System

Gaze-communicative Guide Robot

To provide natural and intuitive guidance through stepwise interactions, our gaze-communicative guide-board system is implemented with a robot agent in three different reactive modes corresponding to the user's gaze. First, the robot tells the user that it can give guidance on the details of the guideboard, corresponding to the user's gaze, when her/his face is detected. Next, when the user looks at the robot, the robot first reacts to the eye contact as initiation of communication, and then it tells the user the guidance information. Finally, when the user looks at some particular region of the guideboard, the robot provides the hierarchically composed information corresponding to the duration of gazing at the region. The robot also behaves as if it is looking at both the board and the user to express an anthropomorphic effort and intention to give guidance.

The guide system consists of a guideboard, our gaze-tracking system, an illuminating projector, and a stuffed-toy guide robot featuring vocal, assistive and gazing behaviors. The guideboard is divided into three regions, and its assumed scenario is guidance in a hotel lobby. This hotel guideboard has the following parts: the left region (LR) shows information and pictures of the restaurant, the center region (CR) describes specially equipped rooms, and the right region (RR) displays information on the hotel's in-house shop (Figure 1). Figure 2 shows examples of the system operations. The stuffed-toy robot reacts to the user's gaze toward it as the initiation of eye-contact communication (Figure 2-A) and provides the detailed information on each region corresponding to the user's gazing direction (Figure 2-B).

We conducted a demonstration experiment in the lobby of a hotel and found that the users operated the system with their gaze irrespective of age or gender, as we designed. Some of the users, whose gaze were not the actual target, observed the interaction between a particular user and the region described by the robot's vocal information (Figure 3-A and 3-B). On the other hand, children enjoyed the interaction with the robot's behaviors by freely using their gazes (Figure 3-C). These observations showed the gradual transition of interactions from gaze communication to the gaze-corresponding guidance.

Figure 3. Demonstration experiment in lobby of hotel

Wide-area and Calibration-free Gaze-tracking Method

The guide robot needs to perceive the user's gaze, which moves in a *wide-area* space such as a room. However, the current gaze-tracking systems still need fixtures, calibrations, and limitations on the distance to the system (Newman et al., 2000; Ohno et al., 2000, etc.). In our method, we employ a single-camera-based gaze-tracking method using a high-resolution camera (Yamazoe et al., 2008) to detect the user's attention or direction of interest in a *wide area*. This is possible because the method can estimate the user's gaze direction from low-resolution face images (320×240 pixels) with eye-region images (30×20 pixels). Figure 4 shows an overview and the results of the wide-area and calibration-free gaze-tracking method. This method, by applying facial-feature tracking and 3D eyeball-model estimations, offers the advantages of not needing any attachment devices to users and not requiring the user to wear attached devices or to perform preliminary actions such as looking at reference points for calibration. Figure 5 shows examples of gaze-tracking angles relative to the board. The averaged accuracy of the angles for each region is shown in Table 1. Accordingly, this system's robot and illuminated board could react to the user's gaze at sufficient response speed and accuracy for natural gaze communication.

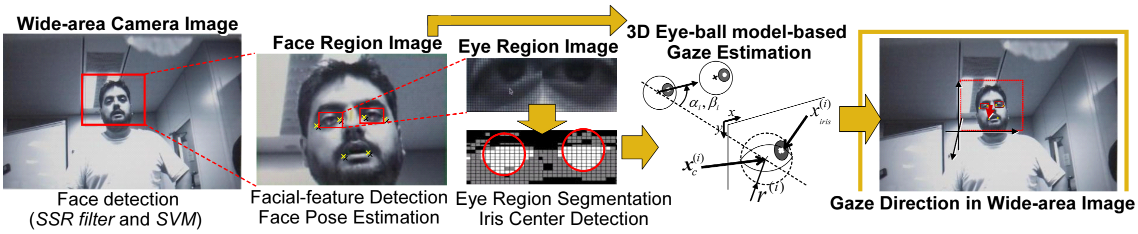

Figure 4. Wide-area Gaze-tracking by Single Camera

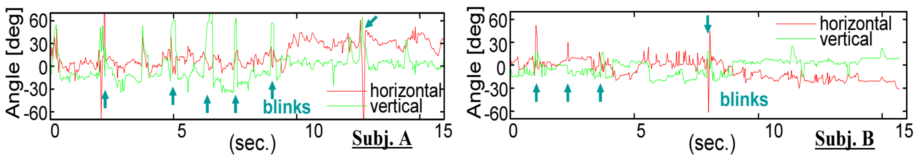

Figure 5. Examples of estimated gaze angles

Table 1. Gaze-tracking accuracies

(region)	*LR*	*CR*	*RR*	robot
horizontal	3.81°	3.51°	2.43°	1.40°
vertical	6.89°	5.44°	5.79°	9.11°

Evaluations of Gaze-communicative Robot on Guideboard

Based on our assumed stepwise interaction model, we examined the effectiveness of the gaze-communicative robot and the gaze-corresponding guidance in the following three types of subjective evaluations performed by 25 people aged from twenties to thirties (13 females and 12 males). These evaluations were made to verify the effectiveness of i) *the robot's presence* in the user's passive guidance (EX1), ii) *the gazing behaviors of the robot* (EX2), and iii) *the gaze-corresponding guidance* (EX3) using both voice and motions of joint attention.

To simplify the contents of the guidance, the guideboard shows only two figures: a triangular pole in *LR* and a cube in *RR*, and the voice guidance simply speaks the name of the object on the guideboard. The subjects evaluate each stimulus using a five-point rating scale to evaluate the relevance (5: very relevant, 4: somewhat relevant, 3: even, 2: somewhat irrelevant, 1: irrelevant) of the following statements: `(a) certainty factor`, `(b) attractive factor`, `(c) naturalness`, and `(d) before-after change of interest`. The short sequences of the guidance were shown to the subjects as the experimental stimuli in randomized order for counterbalance.

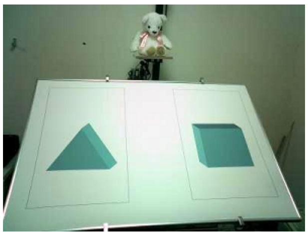

Figure 6. Simple board in evaluations

Conditions: We prepared different conditions to compare them and to verify the effectiveness of each criterion in every experiment. In `EX1`, we prepared the condition α_1 as the automatic vocal guidance without the robot and the condition β_1 as the automatic vocal guidance with showing the robot. For `EX2`, we prepared the condition α_2 as a motionless robot with vocal guidance and the condition β_2 as the robot's vocal guidance with its gazing behavior corresponding the guiding content. In `EX3`, we prepared four conditions to compare two different criteria: gaze-corresponding vocal guidance and gaze-corresponding gazing direction of the robot. So we prepared the condition {`r+,v+`} in that the robot gazes at the figure which the subject looks at, and the vocal guidance explains the figure; for example, the robot talks "This triangle cookie is cocoa flavored." The condition {`r+,v−`} was prepared as the robot gazes at the same figure with the subject, but the content of vocal guidance is the opposite figure. In the condition {`r−,v+`}, the robot gazes at the opposite figure but the vocal guidance corresponds to the subject's gazing figure. In the condition {`r−,v−`}, both the robot's gazing direction and the content of vocal guidance are corresponding to the opposite figure of the subject's gaze.

Instructions: The subjects stood in front of the guideboard. The distance from the guideboard to the subjects was about four feet. In `EX1` and `EX2`, the subjects were instructed only to observe the guideboard system. In `EX3`, the subjects were instructed to select and look at one figure and observe the guideboard system. The subjects were also instructed to do same things for the other figure.

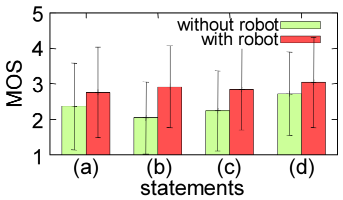

Figure 7. MOS for the robot's presence

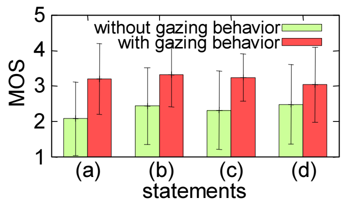

Figure 8. MOS for gazing behaviors

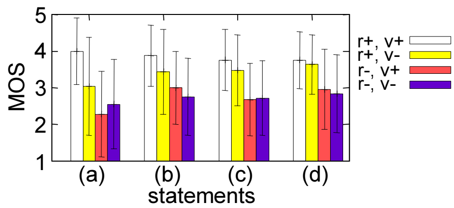

Figure 9. MOS for gaze-corresponding guide

Table 2. ANOVA of MOS for Robot's Presence

	(a)	(b)	(c)	(d)
$F_{(1,24)}$	9.6	27.9	12.0	5.37
p	<.01	<.01	<.01	.029

Table 3. ANOVA of MOS for Robot's Gazing Behaviors

	(a)	(b)	(c)	(d)
$F_{(1,24)}$	23.1	27.9	23.3	7.19
p	<.01	<.01	<.01	.013

Table 4. Two-factor ANOVA of Gaze-corresponding Guide

dir.		(a)	(b)	(c)	(d)
r+ vs r- (robot)	$F_{(1,24)}$	11.6	14.3	23.8	14.2
	p	<.01	<.01	<.01	<.01
v+ vs v- (voice)	$F_{(1,24)}$	3.68	4.11	0.61	1.13
	p	0.067⋆	0.054⋆	0.442	0.298
intrr.	$F_{(1,24)}$	9.17	0.342	1.30	0
	p	<.01	0.563	0.266	–

dir.: direction, *intr.*: interaction

Results: The MOSs (means opinion scores) comparing different stimuli are shown in Figure 7–9 as the results of `EX1`–`EX3`, and their ANOVA results ($df. = 24, \alpha = .05$) are marked with underlines for each significance, with $\star$ for significant tendencies (Table 2–4). `EX1` is a comparison of 1) the guideboard without the robot and 2) the guideboard using the robot (Figure 7). The results of ANOVA for `EX1`

(Table 2) show every significant difference. The results of EX2, comparing 1) the guidance without the robot's motion and 2) the guidance with the robot's gazing behaviors at the corresponding region of the guideboard, also show every significant difference in each statement (Figure 8, Table 3).

The results of EX3 comparing correspondence (+) or non-correspondence (−) of the guiding behaviors with joint attention(r) and voice guide (v) are shown in Figure 9. Here, r+ means the robot gazes in the same direction as the user's gaze, and r− means the robot gazes in the opposite direction to the user's gaze. In addition, v+ means the appropriate contents of the vocal guide for each figure, such as "This is a tent" for *LR* and "This is a building" for *RR*, and v− indicates the inappropriate contents of the voice guide. Two-factor ANOVAs show that the robot's gazing direction corresponding to the user's gazing position obviously increases the quality of the guidance, compared with the only somewhat significant tendencies of the vocal guide's contents (Table 4).

Discussion

As shown in the previous section, our results verified the effectiveness of i) the robot's presence, ii) the robot's gazing behaviors toward the guided region, and iii) guidance corresponding to the user's gaze. These results indicate the beneficial effect of gaze-communicative guidance in enhancing the attractiveness and reliability of a guide system, based on positive impressions such as affection and naturalness.

Consequently, the anthropomorphic presence, the anthropomorphic behaviors, and gaze-corresponding interaction of the gaze-communicative guide make a positive impression on the users of this system. These results suggest the possibility of removing the psychological or physical burdens, that are actually experienced by disabled or hesitant people in asking for further detailed information beyond that on the guideboard.

Conclusion

This paper introduced and evaluated an implementation of an interactive guideboard system using a communicative stuffed-toy guide robot that behaves in correspondence to the user's gazing direction, which is estimated by our remote gaze-tracking method. The primary experiment on the system demonstrated an intuitive and enjoyable system for guidance interaction, performing at sufficient speed and accuracy. Furthermore, we verified both the effectiveness of *the anthropomorphic guide agent* and *the gaze-corresponding guidance* through analyses of subjective evaluations. As future works, we are considering not only vocal guidance and the robot's gazing behaviors with simple illumination but also other modalities such as the robot's sign language and the interactive information on the board to expand the target users toward deaf people.

Acknowledgements

This research is supported by National Institute of Information and Communications Technology. We would like to thank Dr. Norihiro Hagita and Dr. Kenichiro Shimokura for this opportunity, and other ATR members for cooperating our experiments.

References

Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., & Blumberg, B. (2005). Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Autonomous Robots*, *11*, 31–62. (Issues 1–2)

Duffy, B. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3), 177–190.

Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., & Hagita, N. (2002). Messages embedded in gaze of interface agents – impression management with agent's gaze–. *Proc. ACM SIGCHI*, *1*, 41–49.

Katagiri, Y., Takahashi, T., & Takeuchi, Y. (2001). Social persuasion in human-agent interaction. *IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 64–69.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22–63.

Newman, R., Matsumoto, Y., Rougeaux, S., & Zelinsky, A. (2000). Real-time stereo tracking for head pose and gaze estimation. *Proc. FG2000*, 122–128.

Ohno, T., Mukawa, N., & Kawato, S. (2000). Just blink your eyes: A head-free gaze tracking system. *Proc. CHI2003*, 950–951.

Shi, F., & Gale, A. (2007). Environmental control by remote eye tracking. *COGAIN 2007*, 49–52.

Sidner, C., Lee, C., Kidd, C. D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, *166*, 140–164.

Sumi, Y., Etani, T., Fels, S., Simonet, N., Kobayashi, K., & Mase, K. (1998). C-map: Building a context-aware mobile assistant for exhibition tours. *Community Computing and Support Systems, LNCS 1519*, 137–154.

Yamazoe, H., Utsumi, A., Yonezawa, T., & Abe, S. (2008). Remote gaze direction estimation with a single camera based on facial-feature tracking without special calibration actions. *Eye-tracking Researches and Applications*, 245–250.

Yonezawa, T., Yamazoe, H., Utsumi, A., & Abe, S. (2007). Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. *Proc. of ICMI07*, 140–145.