Gaze-communicative Behavior of Stuffed-toy Robot with Joint Attention and Eye Contact based on Ambient Gaze-tracking

Tomoko Yonezawa
ATR IRC Lab.
2-2-2 Hikaridai
Seika-cho, Soraku-gun,
Kyoto 619-0288, Japan
yone@atr.jp

Hirotake Yamazoe
ATR IRC Lab.
2-2-2 Hikaridai
Seika-cho, Soraku-gun,
Kyoto 619-0288, Japan
yamazoe@atr.jp

Akira Utsumi
ATR IRC Lab.
2-2-2 Hikaridai
Seika-cho, Soraku-gun,
Kyoto 619-0288, Japan
utsumi@atr.jp

Shinji Abe
ATR IRC Lab.
2-2-2 Hikaridai
Seika-cho, Soraku-gun,
Kyoto 619-0288, Japan
sabe@atr.jp

ABSTRACT
This paper proposes a gaze-communicative stuffed-toy robot system with joint attention and eye-contact reactions based on ambient gaze-tracking. For free and natural interaction, we adopted our remote gaze-tracking method. Corresponding to the user’s gaze, the gaze-reactive stuffed-toy robot is designed to gradually establish 1) joint attention using the direction of the robot’s head and 2) eye-contact reactions from several sets of motion. From both subjective evaluations and observations of the user’s gaze in the demonstration experiments, we found that i) joint attention draws the user’s interest along with the user-guessed interest of the robot, ii) “eye contact” brings the user a favorable feeling for the robot, and iii) this feeling is enhanced when “eye contact” is used in combination with “joint attention.” These results support the approach of our embodied gaze-communication model.

Categories and Subject Descriptors
H.4.3 [Information Interfaces and Presentation]: Communications Applications

General Terms
Gaze-communicative Robot Design and Experimentation

Keywords
stuffed-toy robot, gaze communication, eye contact, joint attention

1. INTRODUCTION
Communicative robots and artificial agents, with their multimodal representations, have been developed over the past decade with the expectation that their anthropomorphism would provide familiar and natural expressions evoking emotional communication. Anthropomorphism is effective in presenting not only appearances but also various non-verbal expressions [4, 22, etc.]. As a channel of multimodal expressions, the gazing behavior of these anthropomorphic media (e.g., virtual agents [5] and humanoid robots [6, 24]) provides cues to help people guess the media’s internal states, just as in human-human communication [7]. For instance, joint attention, one of the gaze behaviors discussed for communicative learning between mothers and toddlers [10], has been tentatively applied to the artificial behaviors of robots [8] to express their internal intentions or emotions.

Since there are many uses of stuffed puppets in the care of dementia patients [11] and traumatized children [15], stuffed toys offer familiar and unforced communication for people in various situations. This adaptability is due to their flexible anthropomorphic characteristics in both avatars/media and partners as used in playing house. In this paper, we focus on the communicative gazing behavior of stuffed toys. The gaze of animal-like stuffed toys has a distinctive ambiguity arising from the characteristics of their eyes, which are composed almost completely of irises without whites; therefore, gaze should be expressed by the direction of their heads. For effective and comprehensive gaze-communication, stuffed toys must behave in a simple and appropriate way.

Aiming to evoke and establish communication between humans and a stuffed-toy robot, we consider “interest” as a trigger and “favorable feeling (feeling of appreciation or liking)” as a motivation of the communication. The gradual change in gazing behavior along with the progress of communication is adaptive to various physical and psychological situations of the user by exciting her/his conscious or subconscious feelings. In this paper, we propose and verify a gaze-communicative stuffed-toy robot system based on a gradual gaze-communication model. The model consists of 1) joint attention action for indirectly getting the user’s interest and 2) eye-contact reaction for directly eliciting the user’s favorable feeling. The system adopts a motion-programmable stuffed-toy-covered robot (IPRobotPHONE [17]) that reacts to the user’s estimated gaze.

For natural and familiar gaze-communication, it is desirable that users be unaware of gaze-tracking. However, conventional gaze-tracking systems need to mount fixtures
2. RELATED RESEARCH

There have been many research efforts on robots and their gestures for communication in connection with gaze. Castellini et al. have developed a control method for a remote robot using the user’s gaze [3]. This research did not refer to the social or psychological effectiveness of gaze but instead focused on controllability using gaze.

Several studies have examined social recognition/interactions between humans and robots. Scassellati [16] implemented mind theories of infants toward a humanoid robot, but there is not enough verification of each implementation. Breazeal et al. [2] have suggested that the imitative behaviors of a robot are effective for social communication. Sidner et al. [18] have discussed gaze and various interactions of a robot based on the concept of engagement gesture, and Thomaz et al. [19] have examined affective/cognitive aspects of communication using Leonardo, an embodied robot, based on a social referencing model. These works primarily dealt with the facial and gazing behaviors of robots in imitative interactions based on each particular model. However, joint attention, which is an important component of the gaze behavior, has not been sufficiently discussed from the viewpoint of combining it with the other gazing behaviors. Accordingly, this research is motivated by the need to clarify both the simple effectiveness of human-robot eye contact/joint attention and the interaction of these behaviors.

3. SYSTEM DESIGN

3.1 Gaze-communication Model and Design

For evoking communication, we propose a gaze-communicative stuffed-toy robot system that is aware of the user’s conscious and subconscious gaze. Figure 1 shows our model of gradual gaze-communication. The joint attention behavior of the stuffed-toy robot is expected to indirectly evoke the user’s interest as a communicative trigger. Once the user’s interest is drawn to the stuffed-toy robot, an indirect communication based on joint attention is established. Consequently, the direct reaction of the stuffed-toy robot to the user’s eye contact is not perceived as a sudden approach, and such a reaction gives the user a favorable feeling for continuing the communication.

The reactive stuffed-toy robot is capable of making expressions in voice and gestures with its speaker and two-axes freedom in each joint, i.e. the head and arms.
Unlike conventional methods, our method does not require special action such as instructing users to look at several specific points for eye/head model calibration. This can be realized by combining 3-D face-model reconstruction based on factorization and head/eye model estimation by nonlinear optimization. Iris centers are obtained by fitting an ellipse into the observed image (Figure 4-c), and thus gaze direction can be determined (Figure 4-d). The gaze estimation accuracy of our method is 5° in the horizontal and 7° in the vertical orientation.

4. DEMONSTRATION EXPERIMENTS

Based on our model of gradual gaze-communication between a human and a stuffed-toy robot, we first verify the effect of joint attention on both the interest of the user and the guessed interest of the robot. Next we verify the effect of “joint attention” and “eye-contact reaction” on both favorable feeling of the user and user-guessed favorable feeling of the robot.

Common Settings, Methods, and Instructions: To ensure that the subjects concentrate while looking at the robot(s) for one minute during each experiment, they are initially instructed to briefly describe (within five words) after each experiment what the stuffed-toy robot(s) had in mind. Figure 5 shows the experimental environment. We evaluated the system by the subject’s gaze-tracking data and subjective evaluations of the primary stimuli, i.e. the embodied behavior(s) of the robot(s). As the content of joint attention, we also prepared secondary stimuli in the form of simple animations displayed on monitors that changed every six seconds.

4.1 Interest by Joint Attention

Hypotheses: I) Subjects can guess the “interest” of the stuffed-toy robot in either of two animations. II) The “interest” of the subjects is affected by the robot’s gaze. III) Subjects gaze longer at an animation that is also gazed at by the robot longer than at the other animation.

Stimuli: A stuffed-toy robot placed between two monitors turns its head in three directions: to the left monitor, to the right one, and to the subject. The total duration of each direction of the head was set to be A) about thirty seconds for the more-gazed animation, B) about twenty seconds for the less-gazed animation, and C) about ten seconds for the subject in a common gazing scenario designed with random order and duration. The monitors displayed two animations on each side randomly (left and right).

Procedure and Instructions: Figure 6-A shows the environment of this experiment. Subjects were instructed only to guess what the robot had in mind and to answer these two questions: Q1–Which animation is the robot interested in? Q2–Which animation content are you interested in? The subjects indicated their responses using a seven-point rating scale (7: definitely left, 6: probably left, 5: possibly left, 4: even, 3: possibly right, 2: probably right, 1: definitely right).

Results: We summarized each result corresponding to the user’s gaze at the longer-duration gazed monitor (LM), the opposite monitor to LM (OM), and the stuffed-toy robot (SR). One subject’s data could not be captured in this experiment.

Evaluation of Robot’s/Subject’s Interest: The summarized distributions of subjective evaluations for Q1 and Q2 are shown in Figure 7. To verify the significance of the distributions, we
4.2 Favorable Feeling by Robot’s Behavior

4.2.1 Effect of Reaction to Eye Contact

Hypotheses: I) Subjects guess that the stuffed-toy robot has a “favorable feeling” for the subjects from its eye-contact behavior. II) Subjects develop a “favorable feeling” for the robot by the eye-contact behavior.

Stimuli: We prepared two behaviors as follows. B1: reacting corresponding to the subjects’ gaze at the robot (with eye-contact reactions) and B2: reacting at randomized intervals (without eye-contact reactions). The reaction for each eye contact between the robot and the subject is randomly selected from the prepared sets of motions and utterances.

Procedure and Instructions: The subjects were instructed to guess the mind of the stuffed-toy robot and then to compare two different behaviors of the robots’ set (Figure 6-B). After each one-minute experiment, they evaluated Q3 (which robot was assumed to have a favorable feeling for the subject) and Q4 (toward which robot the subject had a favorable feeling) on the seven-point rating scale from left to right with a brief description of what is in the robot’s mind.

Results: The summarized distributions of Q3 and Q4 are shown in Figure 10. Table 3 shows the results of T-tests (prd.=4). There was a significance of Q3 and a significant tendency of Q4 in the T-test. Hypothesis I was supported, and II was supported to an extent. It is possible that the reaction to the user’s eye contact causes a favorable feeling of communication.

4.2.2 Effect of User-initiative Joint Attention

Hypotheses: I) Subjects guess a “favorable feeling” of the stuffed-toy robot for the subjects from its behavior of joint attention. II) Subjects develop a “favorable feeling” for the robot by the joint attention.

Stimuli: We prepared two behaviors as follows. B3: turning the face of the stuffed-toy robot to the position of the subjects’ gaze (with joint attention) and B4: randomized directions at randomized intervals (without joint attention). The robot does not utter anything in either B3 or B4 behavior.

Procedure and Instructions: The same procedure and instructions as the previous experiment (Section 4.2.1) were prepared with the same evaluation items, Q3 and Q4.

Results: The summarized distributions of Q3 and Q4 are shown in Figure 11. There was no significance of Q3 or Q4 in the T-tests (prd.=4) as shown in Table 4, and neither hy-
Figure 11: Favorable Feeling by Joint Attention

Table 4: Analyses of Favorable Feeling #2

<table>
<thead>
<tr>
<th>comparing</th>
<th>t.f. of</th>
<th>ave.</th>
<th>stdev.</th>
<th>$t_{(21)}$</th>
<th>p(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3 &amp; B4</td>
<td>robot(Q3)</td>
<td>3.86</td>
<td>1.25</td>
<td>-.513</td>
<td>.613</td>
</tr>
<tr>
<td></td>
<td>subject(Q4)</td>
<td>4.09</td>
<td>1.41</td>
<td>-.302</td>
<td>.766</td>
</tr>
</tbody>
</table>

Figure 12: Favorable Feeling by Combined Interaction

Table 5: Analyses of Favorable Feeling #3

<table>
<thead>
<tr>
<th>comparing</th>
<th>t.f. of</th>
<th>ave.</th>
<th>stdev.</th>
<th>$t_{(20)}$</th>
<th>p(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B5 &amp; B6</td>
<td>robot(Q3)</td>
<td>3.81</td>
<td>1.29</td>
<td>-.68</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td>subject(Q4)</td>
<td>4.10</td>
<td>1.37</td>
<td>.32</td>
<td>.74</td>
</tr>
<tr>
<td>B5 &amp; B7</td>
<td>robot(Q3)</td>
<td>4.62</td>
<td>1.07</td>
<td>2.648</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>subject(Q4)</td>
<td>4.62</td>
<td>1.28</td>
<td>2.210</td>
<td>.04</td>
</tr>
</tbody>
</table>

4.2.3 Combined Effects

Hypotheses: I) The robot’s eye-contact reaction affects the subject’s guessed “favorable feeling” of the robot when it is used in combination with joint attention behavior. II) The eye contact affects the subject’s “favorable feeling” for the robot in combination with joint attention. III) The joint attention affects the subject’s guessed “favorable feeling” of the robot when it is used in combination with eye-contact reaction. IV) The joint attention affects the subject’s “favorable feeling” for the robot in combination with eye contact.

Stimuli: We prepared the following three behaviors. B5: reacting to eye contact with joint attention, B6: reacting to randomized timing (without eye-contact reactions) with joint attention, and B7: reacting to eye contact without joint attention. In other words, B5 is the combination of B1 and B3, B6 is the combination of B2 and B3, and B7 is the combination of B2 and B4.

Procedure and Instructions: In comparing the pair of B5 and B6, the same procedures and instructions as those in the previous experiments were prepared with the same evaluation items, Q3 and Q4. We also conducted another experiment to compare the pair of B5 and B7.

Results: The summarized distributions of Q3 and Q4 comparing B5–B6/B5–B7 are shown in Figure 12. The results of the T-tests (prd.=4, Table 5) show the significance of the comparative evaluation between B5 and B7. Hypotheses III and IV were thus verified, so it is conjectured that joint attention is effective when used in combination with eye contact; on the other hand, hypotheses I and II were not supported.

5. DISCUSSIONS

The results of Section 4.1 show that joint attention helped the subjects understand the stuffed-toy robot’s interest by its more-gazed direction with only ten seconds difference between LM and OM in a one-minute experiment. The evaluation of the subject’s interest in the contents (Q2) was not affected by the robot’s gaze, but their behaviors such as gazing duration and gaze switches show significant influences. This indicates that the subjects’ subconscious gaze behaviors were drawn by the gaze of the robot. Thus it is conjectured that the gaze-communicative stuffed-toy robot works effectively to gain the user’s interest.

The results of Section 4.2 show different characteristics for each gaze reaction (joint attention and eye-contact reaction) in activating a user’s favorable feeling. Eye-contact reactions provided both the subject’s guessed favorable feeling of the robot toward the subject and the subject’s own favorable feeling toward the robot. Those feelings were enhanced when the eye-contact reactions were used in combination with joint attention; on the other hand, joint attention did not elicit any favorable feeling when it was used alone. These results suggest that the stuffed-toy robot can activate a favorable feeling by effective use of eye-contact reactions in combination with joint attention.

As we predicted in the system design, eye contact provides a direct evocation of a favorable feeling, but an indirect evocation of “joint attention” does not give an obviously favorable feeling, as shown when this factor was used alone. In the practical design stage, the various reactions of a stuffed-toy robot could be made more effective for emotional communication by properly understanding and applying each reaction. In contrast to traditional gaze-tracking systems, ambient gaze-tracking facilitates natural gaze communications in handling various and unrestricted behaviors in real space. While confirming our gaze-communication model with the function of A) joint attention for gaining interest and B) eye-contact reaction for evoking communication by a favorable feeling, as shown in Figure 1, it would be effective to add C) combination of joint attention and eye contact for maintaining the gaze-communication arising from the user’s favorable feeling.

As a characteristic of anthropomorphic representation with various multiple modalities, the verified effectiveness of the gazing behaviors in this research is expected to be exploited for building a new multimodal interaction model of embodied robotics.

6. CONCLUSIONS

This paper proposed a gaze-communicative system for a stuffed-toy robot by using “joint attention” and “eye-contact reaction” based on ambient gaze-tracking. The demonstration experiments show that i) joint attention draws a user’s subconscious interest, ii) “eye contact” brings the user’s favorable feeling for the robot, and iii) the feeling is enhanced when “eye contact” is used in combination with “joint attention.” These results support our gradual gaze-communication model for embodied stuffed-toy robots. Based on the results, we propose expanding our model by combin-
ing the two gaze reactions into a continuous communicative step. As future work, it is necessary to evaluate the naturalness of the gaze-communicative reactions as an important clue to multimodal interaction.

7. ACKNOWLEDGEMENTS

This research was supported in part by the National Institute of Information and Communications Technology of Japan. The authors would like to thank Dr. Noriko Suzuki for her extensive advice on the demonstration experiments.

8. REFERENCES