

Effect of Shared-attention on Human-Robot Communication

Junji Yamato

NTT Communication Science Laboratories
3-1 Morinosato-Wakamiya, Atsugi,
Kanagawa 243-0198 JAPAN
yamato@eye.brl.ntt.co.jp

Kazuhiko Shinozawa, Futoshi Naya

ATR Intelligent Robotics and Communication
Laboratories
2-2-2 Hikaridai Keihanna Science City
Kyoto 619-0299, JAPAN
shino@atr.co.jp

ABSTRACT

In our pursuit of ways to quantitatively evaluate communication between humans and robots, we recently focused on the effect of shared attention on human decision-making. We used a head robot that can make facial expressions and has human face tracking capability, and designed the interaction so that the robot often looked at the same objects the subjects looked at. Subjects were asked to select a color name from two candidates while the robot made a recommendation. The ratio that subjects selected the recommended candidate was used as a measure of the robot's persuasive communication ability. We found that the matching ratio was correlated with the period time that shared attention (SA) was established for a group of submissive subjects. There was a significant difference in the matching ratio between a high-SA group and a low-SA group. This suggests that establishing shared attention is important for human-robot communication.

Author Keywords

Human-robot interaction, communication robot, shared attention, Social agent, interaction design

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User interfaces.

INTRODUCTION

The recent growth in computer power has enabled the development of embodied conversational agents. Also being developed are personal robots or pet robots that can serve as communication partners. A computer can be given a personality by using minimal superficial cues [5]. So, how

about the robots, and what are the differences and similarities between agents and robots? We have experimentally investigated these questions using agent and robots in a simple color-naming task. We have found that both an agent and a robot can make influences user decision-making and that the foot-in-the door technique also works for agent [12,13]. Another finding is that spatial consistency is important; an agent had more influence when color samples were displayed on a screen and robot had more influence when actual color plates were used [7,8]. While a robot had less influence when color samples were displayed on a screen. So, became clear that a crucial condition is that the robot has to share the same space with users in order to perform as a good persuader. In this paper, we focused on why this is so.

Our hypothesis is this. This condition is good for establishing shared attention of human and robot. When a robot, color sample plates, a button box, and a user are all in the same physical space, not separated in a screen and physical space, shared-attention to the color plate or button box are established naturally through the interaction. On the other hand, when color sample and selection buttons were displayed on screen, it is difficult for user to feel that robot is actually putting attention on these objects while the robot talking about the objects. Shared-attention is considered a key factor in the theory of mind [1], and some robots that are designed for communication have some features to achieve this [3,6]. However, a little quantitative evaluation has been made for shared-attention between humans and robots really matters in human-robot communication. We conducted experiments to measure the effect of shared-attention in human-robot communication by measuring the effect on human decision-making in the color name selection task. The effect was quantitatively measured and statistically significant for a specific type of subject.

EXPERIMENT

The experiment was designed to quantitatively measure the influence of robot's recommendation on human decision-making. We used the color name selection task. Subjects looked at the color sample plate, and were asked to select

the color name from two candidates. All colors in the task were ambiguous and some names were not so familiar to ordinary people. For example, carmine or vermilion were displayed as candidates for a bright red color plate, although all words were actually Japanese. The answer was not obvious and most subjects had no prior reference. Before a subject made decision on any given question, a robot in front of the subject made a recommendation supporting one candidate. Subjects could accept the recommendation or not based on their own preference. The experimental set ups are shown in Figure 1.

The ratio that subjects selected the same candidate the robot recommended (matching ratio) was the measure of the robot's persuading capability or the measure of the goodness of the communication between a subject and the robot. The robot recommended the same candidates for all subjects. Which one it would recommend was determined beforehand based on a preliminary experiment. We chose candidates so that expected average matching ratio would become around 0.5 for no-recommendation condition.

Subjects were 28 people (14 male and 14 females), aged from 21 to 29. The average age was 24.0 and the standard deviation was 1.83. There was no instruction as to how they should handle the robot's recommendations. Each subject saw 30 color plates in total, and the order of presentation was the same for all subjects.

Besides performing the color name selection task, subjects took a personality-profiling test to categorize them according to certain personality traits for detailed analysis. We employed the TEG (Tokyo University Egogram)[10], which was developed by the Medical School of the University of Tokyo based on transactional analysis [9] using principal component analysis of a large-scale public survey. The TEG consists of 60 questions and measures five personality factors: CP (critical parent), NP(nurturing parent), A(adult), FC(free child), and AC(adapted child). Each factor is scaled from 0 to 20.

Achievement of shared-attention was evaluated from recorded video after the experiments. The gaze direction of the robot was classified into four classes: subject, color plate, button box, and other. Figure 2 shows the robot for each class of gaze. The gaze of the subjects was classified as robot face, robot (other), color plate, button box, or other. When both looked at the color plate or button box simultaneously, shared-attention was considered to be achieved. Our original expectation was that the effect of robot's recommendation, measured by the matching ratio, would correlate with the period of time shared-attention was established.

Robot

We used a head robot with a human face tracking feature [11]. The robot was build by the MIT Artificial Intelligence Laboratory as a new version of Kismet [2]. It has two eyes with pan and tilt control, two eyebrows with up and down



Figure 1 Experimental set ups.

control, eyelids, a mouth with expressive lips (two degrees of freedom (DoF) for each), and two fan-like ears for expression of its emotional state. It also has neck movements with three DoF. The robot has a video camera in each eye, and another one in the center of its face.

The vision system of the robot can extract and track the skin color region in the captured image. We used this feature to establish eye contact with the subject. The speech of the robot was generated by the "Fluet" Japanese speech synthesizer [4] developed by NTT.

RESULTS

The matching ratio was 0.57. This was higher than a reference group in a no-recommendation condition. However, there was no statistically significant difference.

There was no correlation between the matching ratio and the amount of shared-attention time (SA time) for all subjects. This implies shared-attention did not influence on human-decision making in general.

However, an interesting interaction was found. A high-AC group of subjects whose AC factor of TEG test was more than 11 (75 percentile in standard distribution) demonstrated a strong correlation between the matching ratio and SA time (Speaman's $r=0.51$, $p=0.051$. see Figure 3). On the other hand, for a low-AC group, the correlation was negative, but it was not statistically significant.

Based on this finding, the matching ratios of the high-SA group and low-SA group among high-AC subjects were compared. As shown in Figure 4, the matching ratio for the high-SA group was higher. The difference was statistically significant by t-test ($p<0.05$)

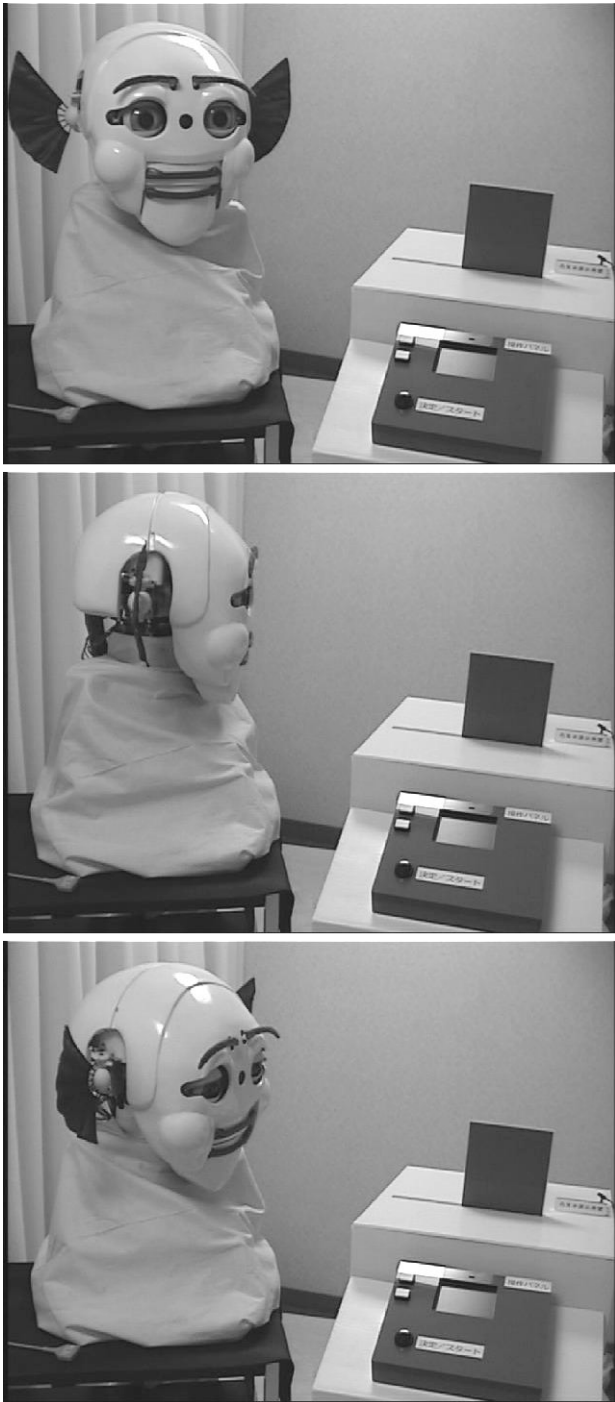


Figure 2. Robot's gaze.

(From the top, subject, color plate, and button box)

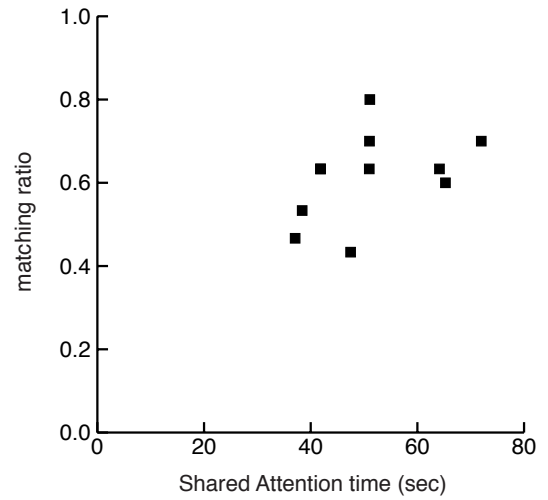


Figure 3. Correlation of SA-time and matching ratio.

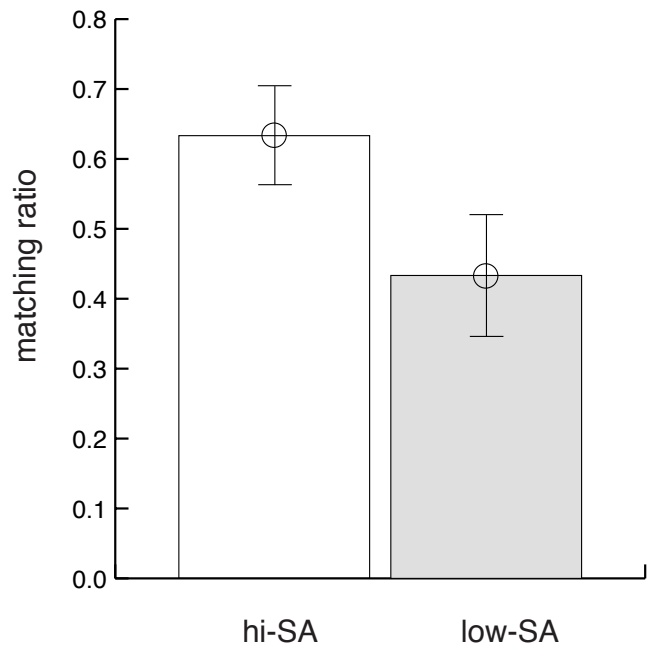


Figure 4. Matching ratios of high-SA and low-SA groups.

Discussion

This result shows that, as in human-human communication shared-attention is important for human-robot communication. Shared-attention had an influence on user's behavior, but not always a positive one. The kind of influence depended on the personality type of the user. Because high-AC people are considered to be submissive or obedient, this result might sound trivial: the high-AC group was obedient, so of course it chose the recommended one. But it was not so simple. The average matching ratios of high-AC group and low-AC group were not so different; 0.59 and 0.54. Furthermore, the average SA time of the high-AC group was less than that of the low-AC group; 50.2 sec versus 53.4 sec. Distributions of the two groups

mostly overlapped. But one has strong positive correlation, and the other negative. This means that the behavior of the high-AC group was influenced by how much time they established shared-attention with the robot. This indicates the effect of shared-attention in communication

Another point is that even a pseudo shared-attention worked well in this experiment. The robot can track a subject's face to establish eye contact using the skin-color detection feature, but it cannot detect subject's eye direction. Therefore, based on preliminary experiments, we designed the robot's behavior so that it looked at where subjects tend to look. So the shared-attention in these experiments was rather stochastically achieved. This suggests that a robot can pretend to follow a user's gaze using an information source other than actual eye direction in real time, which is still difficult to robustly measure from video images. This is important in designing effective interactive robot actions, especially when the robot has limited performance in terms of image acquisition or image processing.

CONCLUSION

We measured the effect of shared-attention on human-robot communication and found that the amount of time that shared-attention was achieved has a positive correlation with the strength of the effect on human decision-making. Although the results are limited to a specific type of subjects (a high-AC factor (adapted child) group), they could be useful in designing more persuasive communication robots.

REFERENCES

1. Baron-Cohen, S. (1995). *MIND BLINDNESS*. Cambridge, MA: The MIT Press.
2. Breazeal, C. (2002), *Designing Sociable Robots* Cambridge, MA, MIT Press.
3. Kozima, H., Nakagawa, C., and Yano, H. Emergence of Imitation Mediated by Objects, *Proc. The Second International Workshop on Epigenetic Robotics (2002)*.
4. Mizuno, O. & Nakajima, S. (1998) Synthetic Speech/Sound Control Language: MSCL. 3rd ESCA/COCOSDA *Proc. International Workshop on Speech Synthesis*, pp 21-26.
5. Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Stanford, CA: CSLI Publications.
6. Scassellati, B. Theory of Mind for a Humanoid Robot. *Proc. the first IEEE/RSJ International Conference on Humanoid Robotics*, (2000)
7. Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2001). Quantitative evaluation of effect of embodied conversational agents on user. *Proc. HCI International '01*.
8. Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2001). Difference in Effect of Recommendation made by Robot and Screen Agent. *Proc. SI2002. SICE* (In Japanese)
9. Stewart, I., and Joines, V. (1987) *TA Today: A New Introduction to Transactional Analysis*. Lifespace Publishing, Nottingham, England
10. TEG Research Group in the University of Tokyo Medical school, *TEG: Tokyo University Egogram (New Ver.)*, Kaneko Shobou(2000).
11. Yamato, J., Shinozawa, K., Brooks, R., and Naya, F. Human-Robot Dynamic Social Interaction. *NTT Technical Review* 1, 6(2003), 37-43.
12. Yamato, J., Shinozawa, K., Naya, F. and Kogure, K. Effects of Conversational Agent and Robot on User Decision. *The IJCAI-01 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents*
13. Yamato, J., Shinozawa, K., Naya, F. and Kogure, K. Evaluation of Communication with Robot and Agent: Are robots better social actors than agents? *Proc. INTERACT'01* IOS Press (2001), pp.690-691