Human-Robot Dynamic Social Interaction

Junji Yamato†*, Rodney Brooks, Kazuhiko Shinozawa, and Futoshi Naya*

Abstract

To determine whether a physical robot produces a more direct emotional coupling with human beings than a computer-generated graphical image of a similar robot (agent), we performed interactive experiments comparing human-agent and human-robot interactions. We found that robots were better communication partners with humans under some conditions. At the MIT Artificial Intelligence Laboratory, we then constructed a robot that has human-like facial expressions and shoulder and neck gestures. After moving it to NTT Communication Science Laboratories in Kyoto, we measured the responses of human subjects under various conditions. Experiments to measure the effects of eye contact and shared-attention are currently under way.

1. Introduction

Recent rapid increases in computational power have enabled the development of embodied conversational agents that interact with their users in a natural and friendly manner through speech recognition, synthesized voice, and action displays. Also being developed are personal robots that can serve as our communication partners or “pets”. Because robots are corporal and exist in the real world, rather than on a screen, they can communicate effectively if they have been designed appropriately.

A computer can be given a personality by using minimal superficial cues [1], and it can display a software agent on its screen to interact with users. But how can a robot be given a personality? What are the differences and similarities between agents and robots? To determine the key factors in designing interactively communicating robots, we experimentally evaluated the interactions between agents and users, and between robots and users. We focused on the effect that recommendations made by the agent or robot had on user decisions, and designed a “color name selection task” for this purpose [2]. We used two robots as the robot/agent for comparison. The first was a small animal-like robot developed by NTT Cyber Solutions Laboratories with an on-screen animated agent based on the Microsoft (MS) agent. The other was the K4 robot developed by MIT Artificial Intelligence Laboratory.

2. Experiments

2.1 Task

The task we designed to quantitatively measure the effect of agent/robot recommendations was as follows. Subjects were shown colored squares one at a time on a computer display and given two candidate names for each color. Most of the colors and candidate names were unfamiliar to ordinary people, such as carmine or vermilion. The subjects were asked to name each color as it was displayed. The answer was not obvious, and most subjects had no prior reference. The agent or robot suggested which color to choose, and the subject could either accept or reject the recommendation. The subjects were told that the experiment was a color-name recognition test, and were not given any explanation of the presence of the agent or robot. After the subject named the color, the agent/robot expressed pleasure if the subject chose its

† NTT Communication Science Laboratories
Atsugi-shi, 243-0198 Japan
E-mail: yamato@eye.brl.ntt.co.jp

recommendation and disappointment if not. We expected the subject to more readily accept the recommendation when the communication between the agent/robot and subject was better and when the subject felt the agent/robot was more familiar or reliable.

2.2 Small animal-like robot developed by NTT

We used a robot originally developed by NTT Cyber Solutions Laboratories and developed the experiment system. It had six motors for controlling the neck, eyelids, mouth, both arms, and waist. The robot and experimental system for presenting the colors are shown in Fig. 1. The agent system was developed at NTT East R&D Center based on the MS agent. The robot and agent both used the “Fluet” Japanese speech synthesizer developed by NTT Cyber Space Laboratories [3].

2.3 Two situations

We expected that the robot would have more influence on subjects’ decisions because it had a physical real three-dimensional body in the real world. It shared the same physical space as the subjects, while the agent existed only on the computer screen. This could give the robot an advantage in establishing better communications in general. However, we thought it might depend on the situation so we carefully prepared the communication environment and equalized other factors that were unrelated to the environment. We chose two situations for both the agent and robot: the “virtual world” and the “real world”.

(1) Experiment 1: virtual world

Colors were presented by displaying squares on the computer screen. Subjects chose one color name from two choices using a radio button style of selection and then clicked the OK button with a mouse. Figure 1 shows the experimental setup for the robot (left) and agent (right). Both used a common set of colors, voices, and scripts for the agent and robot. We compared three conditions in this experiment: ① no recommendation, ② recommendation by the robot, and ③ recommendation by the agent. There were 30 subjects who each tried all three conditions.

(2) Experiment 2: real world

In this experiment, colors were automatically presented as physical color plates the same size as the ones in the virtual world. The machine presenting the plates was designed to be small and user-friendly. We selected an achromatically colored body to avoid any influence of contrast. The subject selected the color name by pressing the corresponding button and then pressing the OK button with his/her finger rather than using a mouse once a decision was made. The box was designed so that the button locations matched those of the buttons on the computer screen for the virtual world. Figure 2 shows the color plate and button box. In this experiment, we tested the same three conditions with 31, 30, and 27 subjects, respectively. The experimental setups for conditions ② and ③ are shown in Fig. 3.

3. Results and discussion

3.1 Experimental results

(1) Experiment 1

The subjects were influenced by the agent’s recommendation. As we can see from Fig. 4, the mean selection ratio of the group of subjects under condition ③ was higher than that of ①. The difference was statistically significant ($p<0.01$). However, the robot’s recommendation did not influence the subject’s decision. This was not what we expected before the experiment. Although the robot shared the space with human subjects, it was not necessarily good for

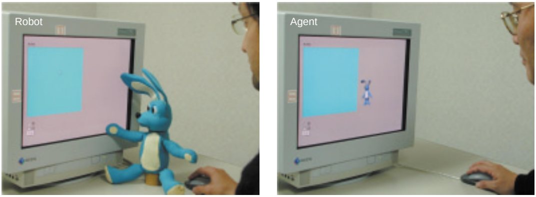

Fig. 1. Experiment 1: Comparing robot and agent recommending a color on the screen.

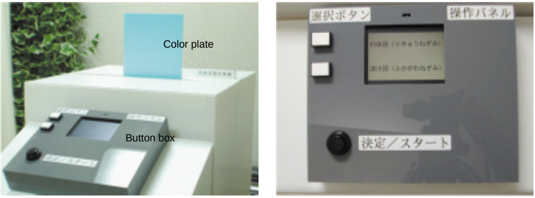

Fig. 2. Real-world settings.

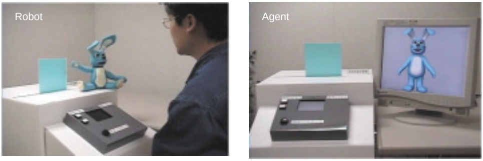

Fig. 3. Experiment 2: Comparing robot and agent recommending a color name in the real world.

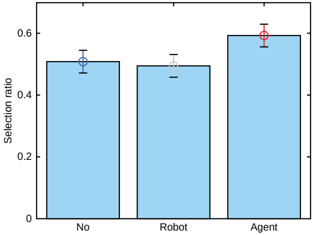

Fig. 4. Results of experiment 1.
Effect of recommendation: agent>robot ($p < 0.01$).

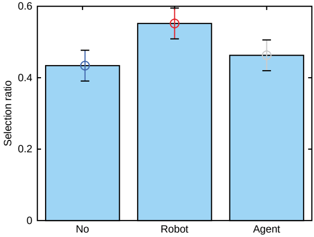

Fig. 5. Results of experiment 2.
Effect of recommendation: robot>agent ($p < 0.05$).

communication. This suggests that this advantage of the robot is not always effective in every situation, but depends on some other conditions within the communication environment.

(2) Experiment 2

Figure 5 shows the mean selection ratios in each group in the real world. The difference between the three groups was significant (F=6.725, p=0.002), as analyzed by ANOVA. There were significant differences between ① “no recommendation” and ② “robot” (p=0.003) and between ② “robot” and ③ “agent” (p=0.042) in a multiple comparison test using Scheffe’s method [4].

3.2 Discussion

The results of the two experiments are summarized in Fig. 6. We found some evidence of the different influences that the agent and robot had on human decision-making. The results revealed that the effect of an agent/robot's behavior depends on the situation and communication environment it has with humans. A robot exists in the physical world because it has a physical body. An on-screen agent exists in the virtual world because it has only a two-dimensional body on a computer screen. Each has more influence when the other related objects, i.e., color plates and button box, are in the same world. This suggests that environmental consistency is a key factor in subject behavior. What features are important for improving the communications of a robot that has a physical body in the real world? In other words, how should we design robot behavior to make it more natural and helpful to human beings? One key factor is the visual (eye) orientation of the robot. This is because the real world is three-dimensional not planar like the computer screen, and the visual orientation represents the focus of the robot's attention. For example, eye-contact and attention-sharing are considered to be important features of communications that display and recognize the attention of participants. As a result, we designed and built a new robot that can establish eye contact and share attention so that we could measure these effects quantitatively.

4. New robot built by MIT

The new robot was built by MIT as a version of Kismet [5]. It is known as K4. The MIT team also developed new low-level software for motor control and improved its vision for social interaction.

The original Kismet robot was developed earlier at MIT. It had two steerable eyes mounted in a head on a multiple-degrees-of-freedom neck, moveable ears, moveable eyebrows, a moveable jaw, and moveable lips. A key component of Kismet's software was its visual attention system. Images were processed at full resolution on parallel processors looking for regions of skin color, saturated color, or motion. The resulting images were added together in image coordinates and the part of the image with the highest score was chosen as the target for a saccade, a rapid movement of the eyes. Besides a visual attention system we also used a stereo system based on the mergence of the two cameras to detect whether people were nearby

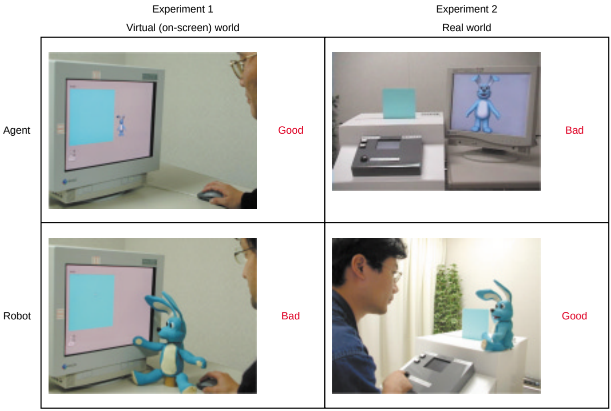

Fig. 6. Summary of results for two experiments.

and run face and eye detection software. This software system ran on a number of parallel X86 processors under the real-time operating system QNX. Using the other degrees of freedom we gave Kismet a simple system for displaying emotions through facial expressions and head posture. Kismet itself had no real face, not even a skin to cover its internal motors. Nevertheless, when we asked naive subjects to interact with it we found that they seemed to engage in social interactions with the robot just as they did with other people.

After seeing how people reacted to Kismet, a natural question is what makes them react in such a way. One can hypothesize a number of reasons, including perhaps (a) that people will interact in this way with any artifact no matter how it behaves or looks, (b) that people interact with Kismet because it is cute looking with big pink ears, (c) that people interact with Kismet because its intentions are clear and it responds to people in an appropriate social manner, or (d) that its physical embodiment in the space shared with humans is very important. Hypothesis (a) seems to be far from the truth. We hoped that building K4 would produce a platform where we could test (b), (c), and (d).

The challenge in building K4 was to come up with a more expressive robot that could have a skin covering all its internals. This would hide the internal mechanism and prevent it distracting human subjects. It would also make it easier to have an articulated three-dimensional graphical model of the face for comparison tests.

K4's mechanical system has a very similar mechanism to Kismet's for eye control. There is a single tilt axis that controls both eyes, but they have independent pan axes. The tilt axis is attached to a cable driven differential system that rolls and tilts the head. More degrees of freedom provide yaw for the head and further redundant degrees of freedom allow very life-like motions. Behind the eyeballs are foveal cameras. A person looking at the robot naturally assumes that the gaze direction they estimate from the physical appearances of the robot corresponds to what it is looking at and this is indeed true—what the eyes seem to be looking at is what the cameras are looking at. In building K4 we carefully mimicked the appearance of the human visual system; indeed, it has the same functionality as the human system. However, K4 has extra concealed wider-angle cameras in its face. These are used for finding peripheral targets to which Kismet may saccade under the influence of its visual attention system. They compensate for the lack of wide-angle view within the cameras behind the eyeballs—they work in an entirely functional way and their placement is designed to avoid confusing the human subject with visible components. As far as a naive human subject can determine, the robot is seeing both peripheral and foveal views through its eyeballs.

The fact that the gaze direction in a real robot determines what it sees makes it very hard to accurately simulate a robot as a purely on-screen computer graphics (CG) system. Any such on-screen robot will still need a camera to perceive the person who is interacting with the on-screen entity, but the camera cannot be placed at the position of the eye in the screen image. Rather, a fixed camera, or multiple cameras, must be mounted elsewhere, and software systems must then generate a virtual view for higher-level processing that corresponds to the view that would be seen from the CG eyes in whatever direction they are currently pointing, otherwise the functional coupling between apparent and actual gaze directions will be lost.

The K4 robot also has more extensive facial expression features than Kismet. One of our goals was to make certain features move much faster than those of Kismet, such as its eyebrows. To do this we used DC servomotors on K4 rather than the model airplane motors used for Kismet. Besides allowing the facial expressions to change at much more human speeds, this method also produces much less motor noise when K4 responds to commands from its internal emotional model.

The K4 robot also needed extensive software to control it. Unfortunately the original Kismet used sixteen PC-class machines running in parallel. Most of that processing was used in the vision systems including the attention system. Each image processed was at least 2^{16} = 65,536 pixels. To significantly reduce the amount of processing we built a new portable vision processing system based on log-polar representations of images. This reduced the number of pixels to about 2048 without sacrificing any resolution at the foveal center of the camera image. The visual attention system works just as well with log-polar coordinates, but there were a number of technical challenges in making the stereo component work in this coordinate system. An integrated log-polar vision system was built that could run on just a single processor.

We built the motor control system in two layers. At the bottom layer it was tuned to the exact kinematics of the K4 robot. But above that level we built a robot-independent system. The saccading mechanism and the ability of the robot's eyes to perform a smooth pursuit were built at the robot-independent level.

- **Experiments using the robot K4**

K4 has a variety of capabilities in terms of both expression and recognition, and we used these to investigate two aspect of human-robot communication. The first was eye contact, which is very important in human-to-human communications and is a well-known cue for gaining attention and attracting interest. We expected that a robot with eye contact would be more familiar and comfortable for humans to interact with. These kinds of user feelings do affect their decisions and can be measured with color-name experiments. The other aspect was shared-attention.

When a human and a robot look at the same object and are both aware of this, shared-attention is established. Humans feel, in this situation, that the robot is truly paying attention to the object and we expect that this feeling will improve the communication between them and give the robot more influence. We redesigned the experiment to use K4 and developed some software to evaluate this human-robot communication. Figure 7 shows K4 and the experimental setup. We can see eye contact between K4 and the human subject in Fig. 8. This is achieved through the human-face-tracking abilities of K4, based on the low-level vision modules developed by MIT.

5. Conclusion

NTT and MIT shared a common research interest in the interaction between robots and human beings, and made efforts to clarify how effective interactions could be established between them. The results discussed in Section 3 represent the first step in quantitatively evaluating human-robot social interactions. Effective communication between robots and humans depends on the environment, and a consistent environment is more important than the features of the agent/robot itself. In the next step, MIT Computer Science and Artificial Intelligence Laboratory's K4 robot is expected to play an even bigger role in achieving eye contact and shared-attention through its visual-perception and gesture-expressing capabilities. An experiment with K4 is currently under way, and is expected to reveal new factors in human robot communication. Although this joint research project has ended, NTT is continuing to investigate how human-robot/agent communication can be made more natural and efficient, currently focusing on the effect of gaze. So the collaboration with MIT, especially the experiments using K4, was an important catalyst for this research activity.

References

[1] B. Reeves and C. Nass, "The Media Equation," Cambridge University Press, 1996.

[2] J. Yamato, K. Shinozawa, F. Naya, and K. Kogure, "Effects of conversational agent and robot on user decision," IJCAI-01 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents, Seattle, U.S.A., Aug. 2001.

[3] O. Mizuno and S. Nakajima, "Synthetic speech/sound control language: MSCL," In 3rd ESCA/COCOSDA Proceedings of International Workshop on Speech Synthesis, Jenolan Caves, Australia, pp. 21-26, Nov. 1998.

[4] K. Shinozawa, F. Naya, J. Yamato, and K. Kogure, "Differences between the effect of robot and on-screen agent recommendations on user decision," In SICE System Integration Division Annual Conference (SI2002, in Japanese), Kobe, Japan, Vol. 1, pp. 363-364, Dec. 2002.

[5] C. L. Breazeal, "Social Machines: Expressive Social Exchange Between Humans and Robots," Sc.D. thesis, MIT, 2000.

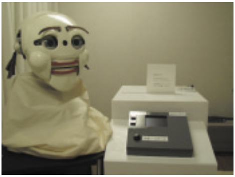

Fig. 7. Robot built by MIT (K4).

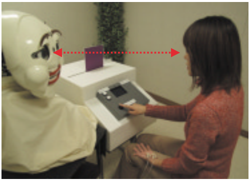

Fig. 8. Eye contact between K4 and subject.

Junji Yamato

Senior Research Scientist, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E., M.E., and Dr.E degrees in precision machinery engineering from the University of Tokyo, Tokyo in 1988, 1990, and 2000, respectively. He also received the S.M. degree in electrical engineering and computer science from Massachusetts Institute of Technology in 1998. His research interests have included computer vision, gesture recognition, and human-robot communication. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), IEEE, and the Association for Computing Machinery. He is a co-author of the book "Analyzing Video Sequences of Multiple Humans—Tracking, Posture Estimation and Behavior Recognition," Kluwer Academic Publishers, 2000.

Rodney Brooks

Professor of Computer Science and Engineering, Director of the MIT Computer Science and Artificial Intelligence Laboratory.

He received the B.Sc. and M.Sc. degrees in mathematics from the Flinders University of South Australia in 1975 and 1977, respectively, and a Ph.D. in computer science from Stanford University in 1981. He was a post-doctoral researcher at Carnegie Mellon University and MIT and a faculty member at Stanford before joining the MIT faculty in 1984. His research interests have included model-based computer vision, Lisp compilers, robot motion planning, mobile robotics, micro-robotics, planetary exploration robots, and humanoid robotics. He is a Fellow of both the American Association for Artificial Intelligence and the American Association for the Advancement of Science.

Kazuhiko Shinozawa

Senior Researcher, Department 1, Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International (ATR).

He received the B.E. and M.E. degree in electrical engineering from Keio University, Kanagawa, Japan in 1988 and 1990, respectively. In 1990, he joined NTT Human Interface Laboratories, where he engaged in solving optimization problems and predicting weather radar images by using artificial neural networks. In 1998, he moved to NTT Communication Science Laboratories. Since then, he has been conducting research for building a robot that can communicate with humans naturally and smoothly. He was one of the principal investigators for this joint project. He moved to ATR in May 2003.

Futoshi Naya

Researcher, Department 1, Intelligent Robotics and Communication Laboratories, ATR.

He received the B.E. degree in electrical engineering and M.S. degree in computer science from Keio University, Kanagawa, Japan in 1992 and 1994, respectively. In 1994, he joined NTT Communication Science Laboratories, Tokyo, Japan. His research interests include multi-agent-based robot control, communication robots, and human-robot tactile interaction. He is a member of IEEE, the Robotics Society of Japan, the Society of Instrument and Control Engineers, and IEICE.