Variable Body Image – Evaluation Framework of Robot's Appearance using Movable Human-like Parts

Hirotaka Osawa and **Yuji Matsuda**
Graduate School of Science and Technology
Keio University

Ren Ohmura and **Michita Imai**
Faculty of Science and Technology
Keio University

Abstract

Many HRI researchers try to make guidelines for an appearance of the robot. However, these guidelines are insufficient to design and to evaluate a robot's appearance separately. We propose new evaluation framework called "variable body image" to solve this problem. This framework changes the appearance of a robot with divided human-like robotic parts and makes it possible to evaluate the appearance during interaction. We designed and implemented hardware and software of movable human-like parts to achieve our framework. Using our framework, we evaluated the types of robot appearance that is effective in human-robot interaction by comparing interactions between a participant and a cart equipped with different human-like parts. We used user preference ratings for the SRT (system response time) to evaluate the effectiveness of the appearances. Our result suggests that a user treats a robot as an anthropomorphic agent even if there are no anthropomorphic parts. The result also suggests that the eyes are more effective in an interaction as compared to the mouth.

Introduction

Many studies have been conducted to improve human-robot interaction. These studies are called HRI (human-robot interaction) studies. One of the important issue of HRI studies is to select appropriate appearance of a robot. Many humanoids are modelled on human appearance . However, MacDorman et al. revealed that a robot with human-like appearance sometime gives uncanny impression to users(MacDorman and Ishiguro 2006). Some studies also suggests that abstracted appearance is more appropriate for interaction(Kozima *et al.* 2003)(Sato *et al.* 2006)(Okada *et al.* 2000). These studies suggest that we need more precise framework to design robot's appearance than the approach to just imitate human appearance.

Many HRI researchers try to make guidelines for an appearance of the robot. DiSalvo et al. challenged to make the guidelines of a robot's appearance by comparing many robot's photos(DiSalvo *et al.* 2002).

However, these guidelines are insufficient to evaluate a robot's appearance separately. For example, it is impossible to determine "where is appropriate positions of eye parts of my robot" or "what part is useless with my robot" by

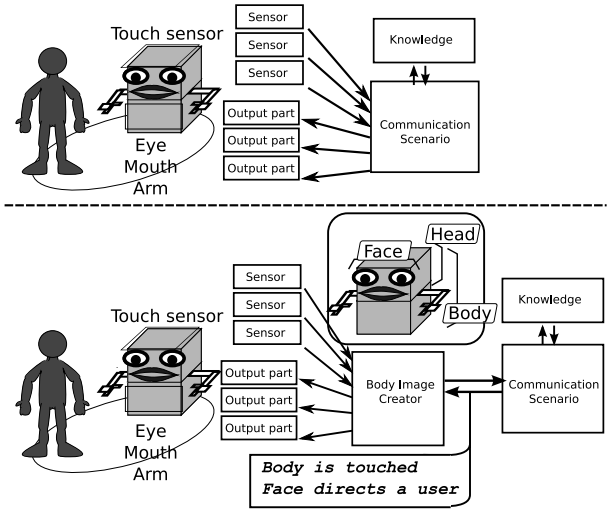

Figure 1: Conventional robot design (up) and robot design with variable body image (down)

these guidelines. To evaluate the apperance of the robot separately, we need more flexible parts that works in real time.

We propose new evaluation framework called "variable body image" to solve above problem. This framework changes the appearance of a robot with divided human-like robotic parts and makes it possible to evaluate the appearance during the interaction. These parts makes it possible to compare each appearance during human-robot interaction. Figure 1 shows how an anthropomorphic robot is designed in the base of our framework. In conventional robot design, a communication scenario is designed according to fixed positions of robotic parts (Fig. 1 up). On the other hand, each part is movable in our framework, and a user accepts imaginary body image according to places of attached parts. The interaction between human and robot flexibly changes according to accepted imaginary body image (Fig. 1 down). In our framework, we can evaluate various appearances of the robot with single set of parts.

We developed movable human-like parts hardware and software to achieve our framework. We also conducted an experiment with these parts to evaluate robot's appearance according to our framework. To evaluate each face con-

dition, we used users' preference for response time of the robot.

Background of Study

In this section, we note related studies that researched how and why human perceive robots or artificial machines as anthropomorphic agents.

Related Studies about Robot Appearance

Many studies suggest that the photorealistic appearance is not needed for a communicative agent like a virtual agent or a robot. Yee et al. suggest that it is not appropriate to make the real appearance of a virtual agent. Mori hypothesized his notion by the word of "uncanny valley" that users feel more fear for anthropomorphic robot when it becomes more similar to human(Mori 1970). MacDorman shown the existence of uncanny valley by morphing of a robot's photo and a human photo(MacDorman and Ishiguro 2006). Their studies show that over-imitation of human is needless for HRI. Then, what appearance is required to engage human-robot interaction?

DiSalvo et al. compared each scale of many robot's anthropomorphic parts by photos(DiSalvo *et al.* 2002). According to this result, they suggest specific feature of anthropomorphic robot as below.

1. Wide head, wide eyes
2. Features that dominate the face
3. Complexity and detail in the eyes
4. Nose, mouth, eyelids and skin
5. Humanistic form language

User's Anthropomorphic Reactions for An Object

Human basically treats an object as a communicative subject and regards them as if it had its own body. Reeves and Nass noted this tendency of human on The Media Equation(B. Reeves and C. Nass 1996). This tendency is found on users even if they understand object's function deeply. There are several hypotheses for a reason of this tendency. Denette divides human's stance for behavior of the system into physical stance, design stance and intentional stance(Denette 1989). He also noted that human has a tendency to find intention in something happen. He hypothesized that this tendency is raised during evolution.

These studies suggest that people have a strong tendency innately to regard something as anthropomorphic subject. This result suggest that people can interact with a robot using their innate imagination even if it lacks for anthropomorphic parts.

Cognitive Rationality

There are also requirement to simplify robot design from cognitive aspect. Sonoyama noted that useless anthropomorphic parts generate excessive expectation of a user(Sonoyama 2003). He suggest that users become disappointed when they notice that human-like humanoid cannot bring or manipulate objects in spite of its decorated hands.

Table 1: Specification of eye-like parts

Scale	120mm × 160mm × 50mm
Weight	180g
TFT Controller	ITC-2432-035
Wireless module	ZEAL-Z1(19200bps)
Microcnotroller	Renesas H8/3694
Connection method	Velcro tape
Cover	Sponge sheet,Plastic board

Some studies have suggested that even if a robot has incomplete facial features, it can interact with humans appropriately. Attachable small communication robot called Cero has just arm and faceless head for gestures(Severinson-Eklundh *et al.* 2003). Keepon has succeeded in developing a simple robot face that has eyes and a dot in place of the nose and can make facial expressions(Kozima *et al.* 2003). Smartpal and Muu have no face (Matsukuma *et al.* 2006)(Okada *et al.* 2000). Through the development of Smartpal, which has no face, its developer has shown it is not necessary for robots to have a face for effective human-robot interaction

Design and Implementation

Implementation of Humanoid Parts for Anthropomorphization

We designed and implemented eye-like and arm-like parts as below to achieve our framework.

Our anthropomorphized object did not need to manipulate other objects. Instead of manipulation, anthropomorphic devices must be simple and light so they can be easily attached. We developed human-like robotic devices and attached them to our target by using hook and loop fasteners.

The eye-like parts are consisted of a TFT LC Panel. They were used to determine the positions of the pupils and irises using the 3-D coordinates of the places they were attached to and their direction vectors. The eye-like parts were 2-cm wide. They were thin and could be attached anywhere. They can be used to gaze in any directions as if the implemented eye of the object were watching. The mouth-like part are consisted of six servo motors and has enough degree of freedom for emotions. Eye-like and mouth-like parts can express six emotions - pleasure, fear, sadness, disgust, anger, and surprise - described on Ekman's work(Ekman and Davidson 1994).

These devices are connected to control PC via Bluetooth. They are easily attach to the object separately with velcro tape. Specifications of parts are presented in Tables 1 and 2, and the parts are depicted in Fig. 2 and Fig. 3.

Software: Body Image Creator

Figure 4 shows the system of Body Image Creator module. Body Image Creator has two functions as below.

First, Body Image Creator generates virtual body image of a target by the geometric data of the target and each coordiante of the human-like parts.

Table 2: Specification of mouth-like part

Scale	350mm × 50mm × 40mm
Weight	400g
Motor	Micro-MG × 6
Wireless module	ZEAL-Z1(9600bps)
Microcontroller	Renesas H8/3694
Connection method	Velcro tape
Cover	Aluminum, sponge, rubber

Figure 2: Eye-like parts

Second, Body Image Creator translate instructions (like ”Point right side” or ”Tap a head”) to low level commands for each device (eye, arm, mouth and speaker) using generated virtual body image.

Experiment

To evaluate our framework, we conduct a experiment to find the basic anthropomorphic features necessary for effective human-robot interaction. We focused on a face of a robot and conduct an experiment to reveal what facial part plays the major role in human-robot interaction.

Evaluation of An Appearance by System Response Time

We use the SRT (system response time), introduced by Shiwa et al., to evaluate the effectiveness of the robot appearance. SRT is the delay caused by a robot when responding to a user's utterance. Shiwa et al. have reported that user preference ratings in human-robot interaction decrease monotonically in propotion to increasing the delay cause by the response of a GUI (graphical user interaface) screen. In the interaction with a robot, however, this decrease is not observed less than 2 s. Shiwa et al. have observed that the user preference ratings are almost the same for all users when they interact with robots which responds to them within 0 or 1 s, and the ratings decrease after 1 s.

We have used the findings of Shiwa et al. (user preferences for SRT) to evaluate the human likeness of a robot. If users find that the robot is anthropomorphic, their preference ratings will decrease after 1 s. On the other hand, their preference ratings will decrease monotonically immediately, if users find that there is no anthropomorphic elements in the interaction (like explanation with GUI screen).

In this experiment, we focus on a short SRT to observe the boundary of an SRT to avoid negative feelings from users.

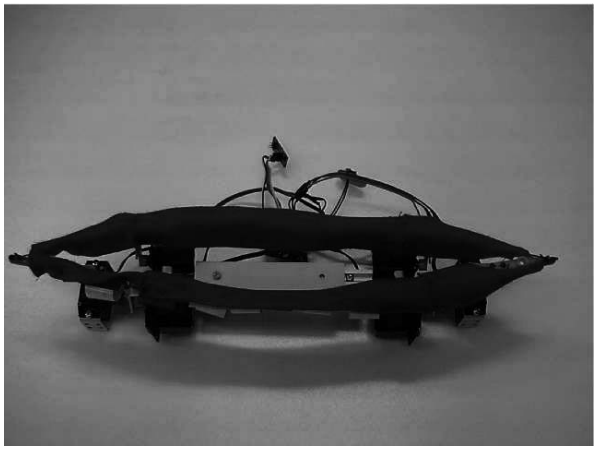

Figure 3: Mouth-like part

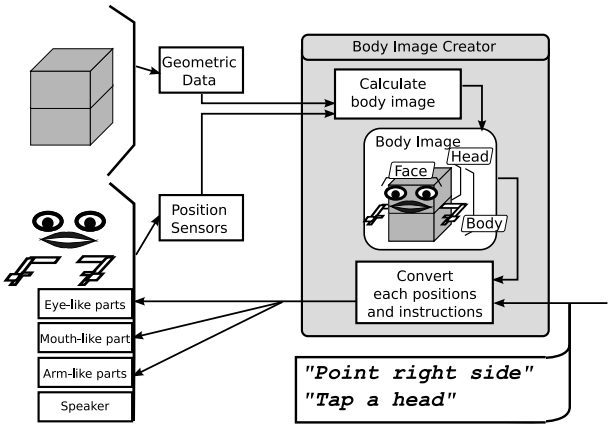

Figure 4: Body Image Creator

Method

Participants Twenty-six native Japanese-speaking university students (22 men and 4 women, average age: 22.8 years) participated in the experiment.

Settings We developed the following five appearances for a robot in this experiment:

- Display only The cart had display device and it had no voice.
- Voice only: The cart had no anthropomorphic appearance but it had a human voice.
- Mouth only: The cart had a mouth-like structure and a human voice.
- Eyes only: The cart had eye-like structures and a human

Figure 5: Each appearances

voice.

- Eye and mouth: The cart had both eye- and mouth-like structures (like a human face) and a human voice.

We attached these robotic parts to a pushcart, as shown in Fig. 5. We use a pushcart as a base of the robot. These robotics parts are controlled and they move during the experiment.

Procedures The experiment employed a within-subject design to avoid effect individualities of participants. First, the order of five appearance settings was counterbalanced, and within each of the settings they engaged in all the different SRT trials. We made four pattern of SRT list that had randomly shuffled delay time. We changed SRT according to this list in each sentence. Before the experiment, the assistant instructed participant to watch the cart during interaction. This instruction suggest a participant to keep his/her intention for the interaction.

The participants engaged in a simple conversation with the cart using six different dialogs listed in Table. 3. We complied the list of dialogs for the interaction taking into consideration the typical tasks that robots execute, e.g., moving objects.

After 4 dialog, the participant got out the room and the assistant modified the face of the cart. When the face of the cart is prepared, the participant got in the room and conducted the experiment again.

The participants were asked to repeat the following actions 20 times (four different SRTs $\times$ five types of appearances) for each interaction. A participant sit down the chair 1.5 m away the cart (Fig. 6). We set this distance according to Hall's social distance(Hall 1966). He/she approaches the cart and engages in a conversation with it by using the six dialogs given in Table 3. Figure 7 shows communication flow of the experiment. When the cart responded, the participants filled out each question about their preference for SRT. One sequence of the experiment finished when the participants answered the sixth question.

Operation The response of the cart is controlled by a human operator, as in the Wizard of Oz method. This Wizard of Oz method is essential since it is necessary to control timing precisely, and it is difficult to robustly determine exactly when a dialog will end. Figure 8 shows the experimental room. The operator sits corner of the room and presses a key immediately after a dialog is delivered by the participants. The corner was hidden by a box and the participants could't see the operator. The operator watched the participant via a hidden camera. This operation triggers the cart

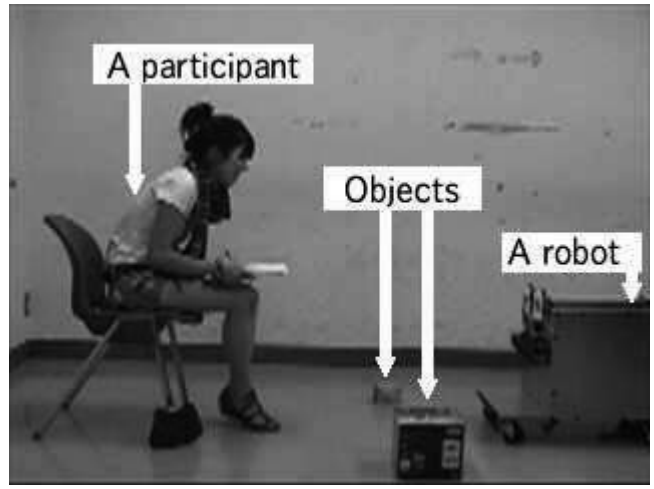

Figure 6: Experiment scene

Table 3: Conversation between user and a robot

User(action)	Robot(response)
"Hello"	
	"Hello"
"Please move the box."	
	"Which one?"
"That one!" (with direction of the right object)	
	"Okey, I understand."
"When do you move it?"	
	"About 6 P.M."
"Yes, please."	
	"Okey."
"Thank you."	
	"You're welcome."

response, which is delayed according to the specified SRT. Since the participants utter predefined dialogs (Table. 3), the operator can easily press the key immediately after each participant delivered a dialog. Delay time included calculation delay and connection delay. The operator could predict the end of the participant's talk precisely, because each dialog is already determined as Table 3.

SRT Parameters On the basis of the studies carried out by Shiwa et al., we carried out the experiment for four SRTs: 0, 1, 2, and 3 s. We compared our results with those obtained by Shiwa et al. in the case of the SRT equal to 2 s (Shiwa *et al.* 2008).

Measurement

Since the evaluations were repeated 40 times, we only measured one evaluation. The participants were asked to rate the anthropomorphic appearance of the cart on a 1-to-7 scale in the questionnaires, where 1 was the lowest grade and 7 was

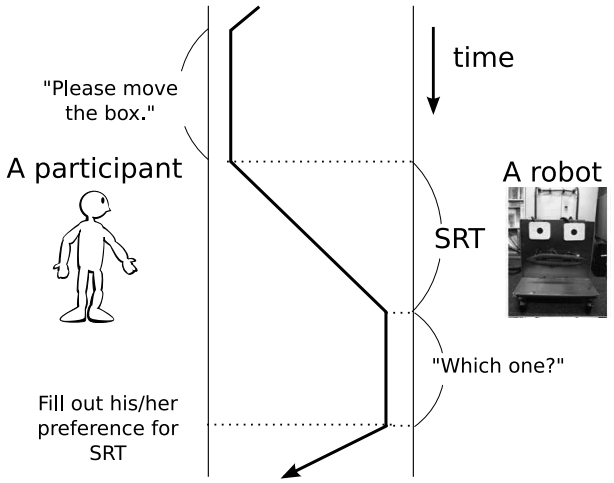

Figure 7: Communication flow

the highest grade. Further, the participants were asked to specify their preference for SRT.

Hypothesis and Predictions

On the basis of our previous study and related robot studies, e.g., the study carried out by Keepon, we hypothesized that users prefer a slightly delayed response even if the facial features of robots are incomplete. We thought that the preference graph for the SRT in the case of the five types of appearances would have the same shape and would not decrease until the SRT becomes equal to 1 s. Further, we thought that the eye-like structures would play the major role as compared to the mouth-like structures in the interaction (on the basis of the study carried out by Bateson et al. on eye-like structures(Bateson *et al.* 2006)). We thought that in the case of a delayed SRT (1-3 s), users would find the robot appearance to be anthropomorphic because such delays are natural in human-human interactions.

On the basis of the three abovementioned considerations, we made the following predictions:

1. When a user interacts with a robot with these four voice appearances, the user preference rating will not decrease monotonically along with an increase in the SRT for SRT less than 1 s. On the other hand, the user preference rating will decrease monotonically in display only condition.
2. The user preference ratings will show a difference between the appearances including eyes ("eyes only" and "eyes and mouth") and the appearances not including eyes ("mouth only" and "voice only").
3. In the case of a delayed SRT (1-3 s), the differences between the user preference ratings in all five cases will be clearer.

Results

Figure 9 shows the relation between the user evaluations and the SRTs for display only condition. Figure 10 shows the relations for other four conditions. The error bars indicate the standard error of each element in the user preferences.

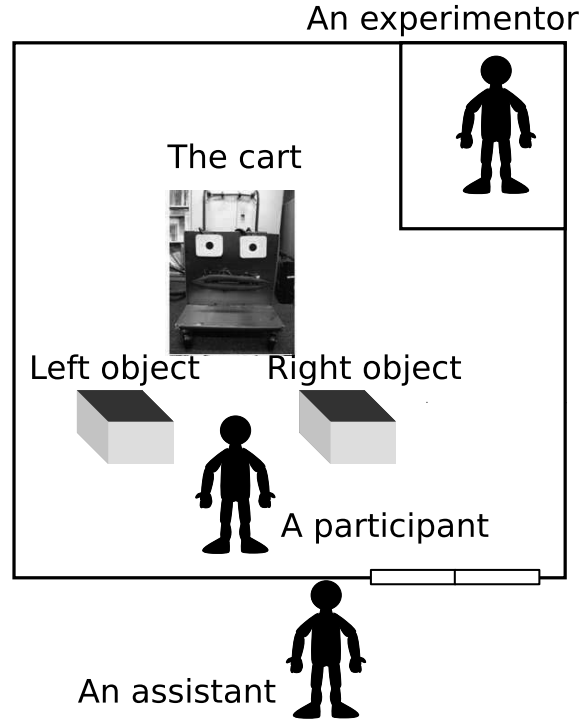

Figure 8: Experimental room

Verification of Prediction 1 The results were analyzed by the two-way repeated-measures analysis of variance (ANOVA; $p < 0.05$) technique using two within-subject factors, the SRT and the elements of the interaction. Significant differences were found in the each SRT factor of the five appearances ($F(3, 75) = 44.232$, $p < 0.001$ for "display only;" $F(3, 75) = 33.712$, $p < 0.001$ for "voice only;" $F(3, 75) = 30.266$, $p < 0.001$ for "mouth only;" $F(3, 75) = 21.585$, $p < 0.001$ for "eye only;" and $F(3, 75) = 41.396$ $p < 0.001$ for "eye and mouth").

The SRT parameters in the four cases were compared using the Ryan method. We found significant differences among four pairs (voice only, mouth only, eye only, and eye and mouth) of SRT parameters ($p < 0.05$ for every pair), except the 0-1 s pair ($p > 0.10$). We found significant differences among all pairs of SRT parameters in display only condition ($p < 0.05$ for every pair).

Further, the ratings did not decrease monotonically for SRTs less than 1 s. Thus, prediction 1 has been supported.

Verification of Prediction 2 In order to compare the appearances of the robot, three within-subject factors - appearance, SRT, and elements of interaction - were analyzed using the three-way repeated-measures ANOVA.

1. A significant difference was observed between the appearance factors ($F(3, 75) = 6.615$, $p < 0.001$).
2. A significant difference was found in the interactions for different appearance factors and SRT factors ($F(9, 117) = 2.561$, $p < 0.01$).

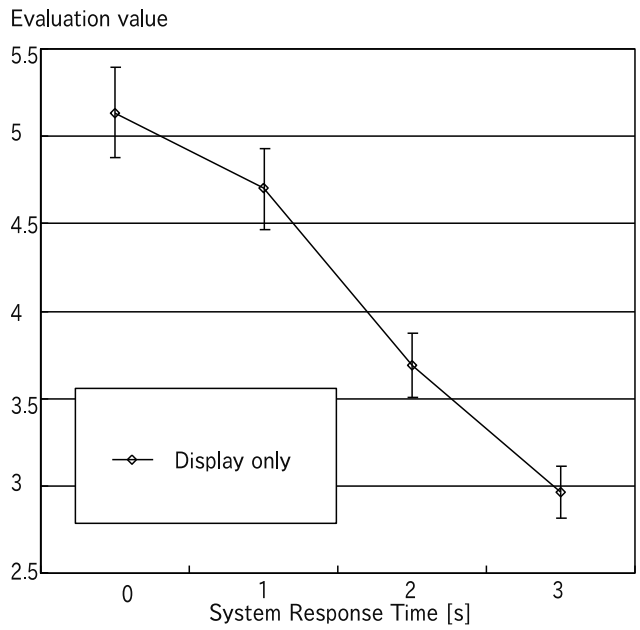

Figure 9: A relation between evaluation and SRT for display only condition

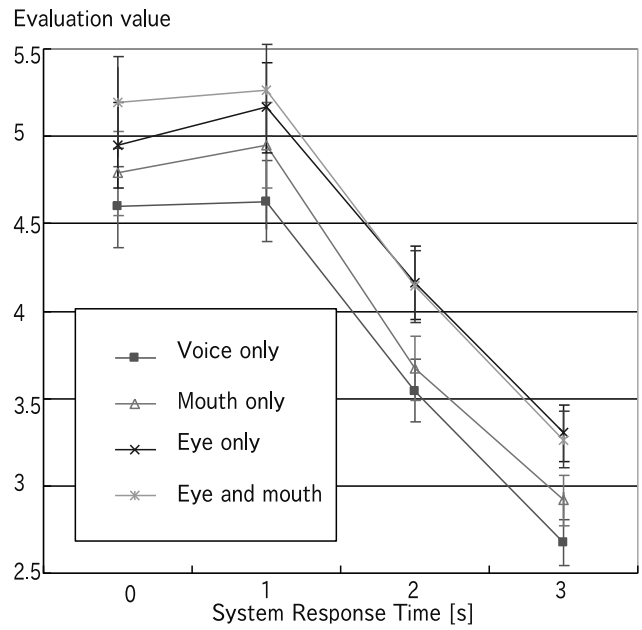

Figure 10: Relations between evaluation and SRT for other four conditions

The first analysis, in which multiple comparisons were made using the Ryan method, revealed significant differences among all pairs of appearance factors, except the "eyes and mouth" and "eyes only" pair and the "mouth only" and "voice only" pair. That is, there was a significant difference in the user preference ratings when the robot had only eye-like structures ("eyes only" and "eyes and mouth" pair) and the robot did not have eye-like structures ("mouth only" and "voice only" pair). Thus, prediction 2 has been supported.

Verification of Prediction 3 In the second analysis, multiple comparisons were made using the Ryan method among the five appearance parameters and SRT parameters. The results revealed significant differences among all pairs of appearance parameters, except between the "eyes and mouth" and "eyes only" pair and the "mouth only" and "voice only" pair.

No significant difference was found for the SRT equal to 0 s. Significant differences were found in the interactions in the case of the 1-3, 1-4, and 2-4 s pairs for the SRT equal to 1 s ($p < 0.05$), in the case of the 1-4 and 2-4 s pairs for the SRT equal to 2 s ($p < 0.05$), and in the case of the 1-3 and 1-4 s pairs for the SRT equal to 3 s ($p < 0.05$).

These results show that a difference was found only if the response time was large (1-3 s). Thus, prediction 3 is also supported.

Discussion

Why is a delayed response preferred irrespective of the robot appearance?

The user preferences for SRT are affected by the expectations of the users (Nagaoka *et al.* 2003), which are based on the system models they have in their minds. In the interaction with a communication robot, they use the conversation model of humans to anticipate the robot's reaction due to its anthropomorphic features (as Shiwa has reported).

The first analysis shows that this model can also be applied to robots with incomplete anthropomorphic features. We assume that users has wide-ranging area for recognition of face than our faces. To confirm this hypothesis, we have to carry out experiments to check user preferences for SRTs between 0 s and 1 s.

The third analysis also supported the hypothesis that if a robot takes more time for executing a task, it has better anthropomorphic features. In order to confirm this hypothesis, we have to carry out experiments with more number of participants.

Why are eye-like structures more important than mouth-like structures?

Basically, the fundamental function of the human eye is to see, and that of the human mouth is to talk. These two features play important roles in human interactions. However, we have found that eye-like structures play the major role in the user preferences during human-robot interaction.

We have hypothesized that this difference is generated by the difference in the importance of the eyes and the mouth in information dissemination. Biological studies have shown that the human eye can not only see but can also determine what a person is looking at (Kobayashi and Kohshima 2001). The mouth also plays an important role in disseminating information.

We can determine where a person is looking at from his/her eye movements. On the other hand, information can

be disseminated verbally through the mouth. In other words, a person may lie, but his/her eyes cannot lie. If a robot does not have eyes, it is extremely difficult to predict whether it is paying attention or to guess its response. We think that above confusion decreased user's preference of interntion timing in "Mouth only" and "Voice only" appearance.

Our result suggests that robots with eyes, e.g., Keepon, are suitable for human-robot interaction because there is no significant difference between the "eyes and mouth" and "eyes only" pairs. This result is also useful for designing new robots. To corroborate our results, we plan to carry out experiments including more participants.

Finally, we have noted that cultural difference of users may possibly affect their preferences. Some studies have shown that the mouth plays the major role in interactions in the Western culture (Morris 1996). To confirm the contribution of cultural difference to the user preferences, we plan to carry out experiments including participants from different cultures.

Improvement of Design of Anthropomorphzed Object

Our result has shown that robot eyes are important in determining the user preferences for SRT. This result will help in the design of a better system including anthropomorphic agent, especially by using our direct anthropomorphization method.

In our previous study, we have proposed a new type of human-robot interaction in which the target, which can be a common object, is anthropomorphized by using attachable humanoid parts (Hirotaka Osawa 2006). This type of interaction does not depend on the cultural background or the literacy level of the users because this interaction takes place through body language and gestures. Users understand the robot instructions through the body movements of the robots even if the users do not have prior knowledge about the robots.

In the abovementioned anthropomorphization method, the eyes are more important than the mouth. In the future, for example, it will be easier for users to understanding complex machines efficiently if these machines have eye-like structures.

Future Prospects of Our Study

The final goal of conventional HRI studies is to develop robots with a human-like appearance. For example, studies on the uncanny valley (MacDorman and Ishiguro 2006) just compared appearances between machine and human.

However, our result shows that the possible appearance region of robot is wider than the region between man and machine. For example, "eye" only design may be efficient for instructions than human total face. We believe that better anthropomorphic designs have to be developed for robots for specific interactions.

Conclusion

In this study, we propose new evaluation framework for robot's appearance named variable body image.

We developed movable human-like parts hardware and software to achieve our framework. We also evaluated the types of robot appearance that is effective in human-robot interaction by studying the interaction between several participants and robots which have different appearances. To evaluate the effectiveness of the appearances, we used user preference ratings for the SRT (system response time), on the basis of studies carried out by Shiwa et al (Shiwa *et al.* 2008).

Our result suggests that a user also treats a robot as an anthropomorphic agent even if anthropomorphic parts are incomplete. The result suggests that the eyes are more effective in interactions as compared to the mouth. Better anthropomorphic appearance leads to better human-robot interaction if the robot needs delay between each sentence.

Acknowledgments

The first author was supported by the JSPS Research Fellowships for Young Scientists.

References

B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Univ. of Chicago Press, 1996.

Melissa Bateson, Daniel Nettle, and Gilbert Roberts. Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2:412–414, 6 2006.

Daniel C. Denette. *The Intentional Stance*. The MIT Press, March 1989.

Carl F. DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. In *DIS '02: Proceedings of the 4th conference on Designing interactive systems*, pages 321–326, New York, NY, USA, 2002. ACM.

Paul Ekman and Richard J. Davidson. *The Nature of Emotion: Fundamental Questions*. Oxford University Press, 1994.

Edward T. Hall. *The Hidden Dimension*. Doubleday & Company, Inc., Garden City, NY, 1966.

Michita Imai Hirotaka Osawa, Jun Mukai. Anthropomorphization of an object by displaying robot. In *Proceedings of IEEE International Symposium on Robot and Human Interactive Communication*, pages 4166–4173, Hatfield, United Kingdom, 9 2006.

H. Kobayashi and S. Kohshima. Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of human evolution.*, 40(5):419–435, 2001.

Hideki Kozima, Cocoro Nakagawa, and Hiroyuki Yano. Attention coupling as a prerequisite for social interaction. In *RO-MAN '03: IEEE International Workshop on Robot and Human Interactive Communication*, pages 109–114, San Francisco, CA, USA, October 2003.

Karl F. MacDorman and Hiroshi Ishiguro. The uncanny advantage of using androids in social and cognitive science research. *Interaction Studies*, 7(3):297–337, 2006.

Kenji Matsukuma, Hiroyuki Handa, Dai Kono, Seiichiro Okada, and Zenta Nakamoto. Sensor applications for smartpal platform robot. *Yaskawa Technical Review*, 69(4):210–214, 2006.

Masahiro Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.

Desmond Morris. *Bodytalk: A World Guide to Gestures*. Jonathan Cape, 1996.

Chika Nagaoka, Masashi Komori, M R Draguna, Satoshi Kawase, Makiko Yuki, Tomotsugu Kataoka, and Toshie Nakamura. Mutual congruence of vocal behavior in cooperative dialogues; comparison between receptive and assertive dialogues. *Human Interface*, 2003:167–170, 2003.

Michio Okada, Shoji Sakamoto, and Noriko Suzuki. Muu: Artificial creatures as an embodied interface. In *Proceedings of the 27th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000)*, page 91, 2000.

M. Sato, A. Sugiyama, and S. Ohnaka. Auditory system in a personal robot, papero. *Consumer Electronics, 2006. ICCE '06. 2006 Digest of Technical Papers. International Conference on*, pages 19–20, Jan. 2006.

Kerstin Severinson-Eklundh, Anders Green, and Helge Huttenrauch. Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous Systems*, 42(3-4):223–234, 2003.

Toshiyuki Shiwa, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. How quickly should communication robots respond? In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 153–160, New York, NY, USA, 2008. ACM.

Takasuke Sonoyama. The meaning and the possibility of a robot design as a network device. *Proceedings of the Annual Conference of JSAI*, 17(2):2D1, 2003.