Proceedings of the 32nd ISR(International Symposium on Robotics), 19-21 April 2001

Robovie: A robot generates episode chains in our daily life

Hiroshi Ishiguro[1,2], Tetsuo Ono[1], Michita Imai[1],
Takeshi Maeda[1], Takayuki Kanda[1], Ryohei Nakatsu[1]
(1) ATR Media Integration & Communications Research Laboratories
(2) Department of Computer & Communication Sciences, Wakayama University,
930 Sakaedani, Wakayama 640-8510, Japan
E-mail: ishiguro@sys.wakayama-u.ac.jp

Abstract

We have developed a robot called "Robovie". A unique aspect of this robot is the mechanism designed for communication. Robovie can generate human-like behaviors with the actuators and sensors. First of all, we have obtained two important results for human-robot communication in cognitive science; one is importance of physical expressions using the body and the other is effectiveness of the robot's autonomy in robot's utterance recognition by humans. Based on these psychological experiments, we have designed a new architecture and implemented for generating episode chains in the interactions with humans. The basic structure of the architecture is a network of situated modules. Each module consists of elemental behaviors to entrain humans and a behavior for communicating with humans.

1. Introduction

There are two research directions in robotics; one is to develop task-oriented robots that work in limited environments and the other is to develop interaction-oriented robots that collaborate with humans in open environments. Industrial and pet robots are the former ones. They work in factories and limited areas in a house with particular tasks such as assembling industrial parts, behaving like an animal, and so on. On the other hand, the purpose of the robot that we are developing is not to execute particular tasks. We are trying to develop a robot that exists as our partner in our daily life. The fundamental requirement of humans in our daily life is to communicate and recognize the existence each other. Our robot supports such an aspect of our life and provides rich information to humans by using the communication functions. We consider, the robots existing as our partners will be a new information infrastructure for communication.

For realizing the robot, we are tackling to establish a new collaboration between cognitive science and robotics. Cognitive science, especially on ideas of body properties for communication, helps to design more effective robot-behaviors for interacting with humans. On the other hand, the developed robot can be used for verifying theories of cognitive science. We consider this unique interdisciplinary relationship enable us to develop the new type of robot.

This paper, first of all, introduces the developed robot

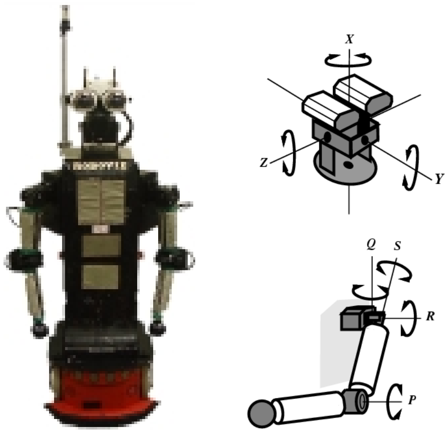

Fig. 1: Robovie

called Robovie. Then it shows two important cognitive experiments. Based on the experiments, the last section discusses a new robot-architecture for generating episode chains in our daily life.

2. Robovie: an Interactive Humanoid Robot

We have developed a robot called "Robovie" shown in Fig. 1. The robot that has a human-like appearance is designed for communication with humans. Like a human, it has various sensors, such as vision, sense of touch, audition and so on. With the human-like body and sensors, the robot can perform meaningful interactive-behaviors for humans.

[Hardware] Fig. 1 shows the developed robot. It is a humanoid-type robot that moves two driving wheels. The size is important as an interactive robot. Not to give an awful impression to humans, we have decided the size as 120 cm, which is same as a junior school student. The diameter is 40 cm and the weight is about 40 Kg. The robot has two arms (4*2 DOF), a head (3 DOF), two eyes (2*2 DOF for gaze control), and a mobile platform (2 driving wheels and 1 free wheel). The robot farther has various sensors, skin sensors covering the whole body, 10 tactile sensors around the mobile platform, an omnidirectional vision sensor, two microphones to listen human voices, and 24 ultra-sonic sensors for detecting obstacles. The eye has pan-tilt mechanism with direct-drive motors and they are used for direct-drive motors and they are used for stereo vision and gazing control. The skin sensors are important for realizing interactive behaviors. We have developed a sensitive skin sensors using pressure sensitive conductivity rubber. Another important point in the design is the battery life. This robot can work 4 hours and charges the battery by autonomously looking for battery stations. With the actuators and sensors, the robot can generate almost all behaviors needed for communication with humans.

[Software] Robovie is a self-contained autonomous robot. It has a Pentium III PC on board for processing sensory data and generating gestures. The operating system is Linux. Since the Pentium III PC is sufficiently fast and Robovie does not require precise real-time controls like a legged robot, Linux is the best solution for easy and quick development of Robovie's software modules.

3. Two Cognitive Experiments

With this robot, we have performed two experiments for human-robot communication in cognitive science. As the results, we have obtained two important ideas: one is importance of physical expressions using the body and the other is effectiveness of robot's autonomy for robot-voice recognition by humans. In other words, the ideas are based on "joint attention" between the robot and a human.

3.1 Mutual Entrained Gestures in Human-Robot Communications

Mutual entrained gestures are important for smooth communications between Robovie and a human. We have performed psychological experiments to ensure it. The aim of the experiments was, concretely speaking, to investigate correlations between body movements and utterance understanding in human-robot communications. The detail is summarized as follows.

[Experiments] We focused on the interaction between a subject and a robot while it teaches a route direction to the subject, and investigated the appearance of the subject's gestures and the level of the utterance understanding by using several different gestures in the teaching.

[Subjects] For this experiments, we asked collaboration to thirty undergraduate and graduate students as the subjects, and randomly divide them into six groups. The subjects had not previously visited this experimental environment.

[Environment] Fig. 2 shows hallways in our laboratory. Points S and R denote the initial positions of a subject and the robot, respectively. The robot taught a route direction to the lobby B at A.

[Procedure] The experiments consist of the following three phases

1. The subject and the robot move from S to A and from R to A, respectively.
2. At A, the subject asks a question "Tell me the way to the lobby", and the robot begins to explain the route. The robot says "Go forward, turn right, turn left, turn right, turn left, and then you will arrive at the destination." While speaking, it performs gestures in one of the six levels described in [Condition]. The purpose of this experiment is to investigate relations between the six levels of robot's gestures and emerged human's gestures.
3. The subject tries to go to the lobby. When the subject arrives at the lobby or it gives up by losing the way, the experiment finishes.

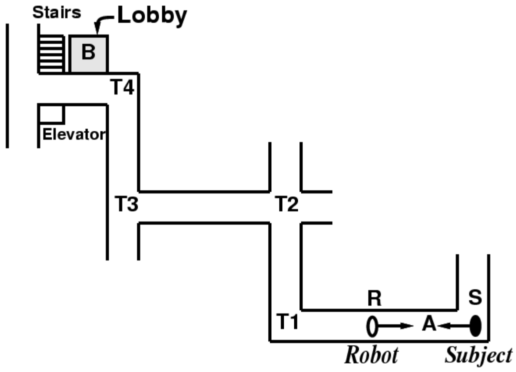

Fig. 2: The structure of the experimental environment

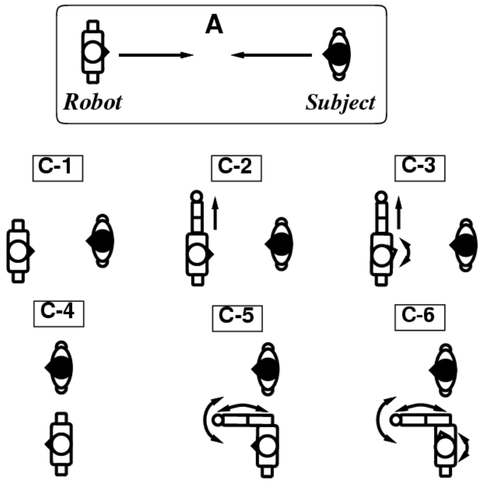

Fig. 3: Six levels of robot's gestures

[Conditions] As conditions of the experiments, we have prepared the six levels of robot's gestures as shown in Fig. 3.

C-1: The robot does not move.

C-2: The robot raises the left arm leftward when speaking "Go right" and rightward when speaking "Go left".

C-3: In addition to C-2, the robot turns the eyes to the subject while talking.

C-4: The robot stands side by side and directs the body along the hallway.

C-5: In addition to C-4, the robot raises the right arm forward, rightward and leftward when it teaches the directions.

C-6: In addition to C-5, the robot turns the eyes to the subject while talking.

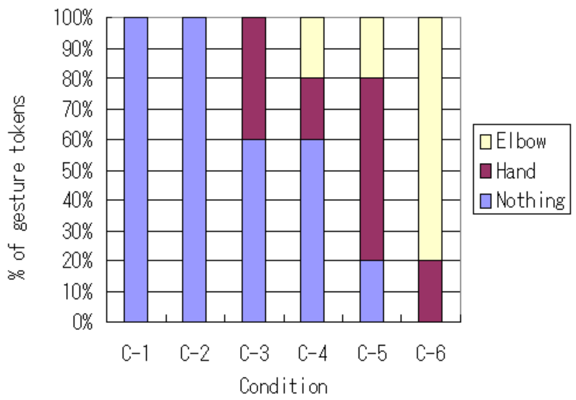

Fig. 4: Results of subjects' body movements in human-robot interaction

Fig. 5: Scenes of human-robot interactions

	C-1	C-2	C-3	C-4	C-5	C-6
Time to destination	69.5	71.3	67.7	70.2	66.8	65.4
Number of subject not arriving	1	2	3	0	0	0

Table 1: Average time to arrive at the destination and the number of failed subjects

Fig. 4 shows the ratio of subjects' body movements under the six levels. We have classified the body movements into three categories: no body movement (Nothing), hand movements (Hand) as shown in the left photo of Fig. 5, and raising hands up to the elbow level (Elbow) as shown in the right photo of Fig. 5. Fig. 4 shows a significant changes of subject' gestures against the conditions (χ^2 = 25.210, $p<0.01$). As the level changes from 1 to 6, the subjects perform bigger gestures. Moreover, the average numbers of times that the subjects gaze the robot were as follows: 0.8 (C-1), 1.0 (C-2), 2.0 (C-3), 1.2 (C-4), 1.0 (C-5), and 3.8 (C-6).

Further, we recorded the time that the subjects spent to move from A to B in Fig. 2. Table 1 shows the average time and the number of subjects who did not arrive at B. Regarding the average times, there is no significant difference among the conditions, but the average time in C-6 is shorter than others. A more noteworthy point is that a considerable number of subjects could not arrive at the destination in C-1, C-2, and C-3. The reason found in the questionnaire is that they could not understand robot's utterance. Especially, they confused to understand the meaning of "left" and "right". However, in C-4, C-5, and C-6, there is no subject who could not arrive at the destination. This means that they could obtain a joint viewing point by the robot's gestures.

We conclude these experimental results as follows:

1. Many and various behaviors of the robot induce various human communicative gestures. In other words, the subject's gestures are increased by entrainment and synchronization with the robot and a relationship between the robot and the subject is established from the mutual gestures.
2. The emerged mutual gestures help to understand robot's utterance.
3. The joint viewpoint represented by the robot gestures allows the subject to understand the utterance.

3.2 Joint Attention in Human-Robot Communication

The experiment shown in the last sub-section clarified the importance to share a joint viewing point in human-robot communication. The results suggest proper robot behaviors in the development of everyday robots. The concept of the joint viewing point can be extended as the concept of joint attention and it gives more proper robot behaviors for interacting with humans.

The relevance theory [3] proposes a communication model for recognizing situations and the humans' experiences. It employed a new term named mutual manifestness that represents mental state where two or more humans recognizes the same situation or recall similar experience. The relevance theory regards human's communications as a process of gaining mutual manifestations by passing messages to others. This concept of mutual manifestness is same as that of focus of attention; and it is called "joint attention" in social psychology [4] in the case where people frequently focus on the same object while communicating each other.

However, mechanisms of focus of attention proposed so far are insufficient for developing a speech generation system depending on the situations. The following three difficulties have to be overcome in human-robot communication.

1. How to draw a human's attention to the target to which a robot is paying attention.
2. How to make a human realize the intention of the robot.
3. How to utilize the human's attention in a robot mechanism for communication.

Difficulty 1 is attributed to the lack of an expression when the robot pays an attention. Without the expression, the human cannot realize where the robot is paying its attention. That is, the human and the robot are not in a state of mutual manifestation in terms of the relevance theory. The lack of a human's attention in a conversation is a crucial

	Saw a poster	Saw Robovie's hand
With eye-contact	6	0
Without eye-contact	1	5

Table 2: Comparison of the number of humans who looked at a poster pointed out by the robot

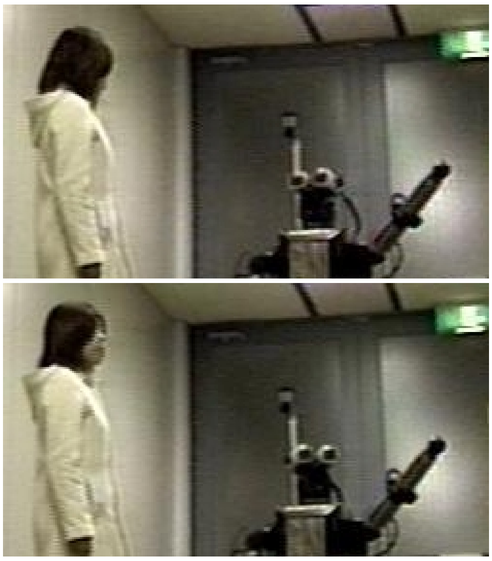

Fig. 6: Eye contact between subjects and the robot

problem. For example, when a guide robot in a museum focuses on an artwork and begins to explain it to a human by using demonstrative pronouns, the human may not be able to understand what the robot is explaining if the human is not focusing on the artwork. This difficulty is overcome by adding attention expression behaviors to the robot. Here, we have implemented two behaviors: gazing head motion to face the target and hand gestures to point at it.

Difficulty 2 comes from Difficulty 1. The relevance theory insists that the occurrence of a state of mutual manifestation depends on the inference of the speaker's communicative intention. For example, when the robot says that "Take this away." In front of a box in order to proceed forward, the human has to use the robot's communicative intention to interpret the robot's utterance. If the human pays its attention to the box, the situation can be recognized as a state of mutual manifestation with the robot. This difficulty is overcome by employing an eye contact behavior. The robot turns the head direction to face the human to promote the relationship with the human. The eye contact inspires the human to guess the robot's intention and to become aware of the robot's attention manifested by the attention expression.

Difficulty 3 is attributed to the joint viewing point problem as discussed in the previous sub-section. By sharing the joint viewing point, a human can easily recognize the robot's utterance even if it omits some concrete words. This effect is not only for sharing the joint viewing point, but also to have a proper positional relation among a robot, a human, and a target. For example, when the robot ask to move a box locating in front away, if they have the proper position relation, it can say just "Move it away". The function to share a joint viewing point and to establish a proper positional relation is effective for smooth communication.

We have verified the effect of the robot's behaviors discussed above. First of all, we have prepared two groups each of which consists of six subjects: one was given Robovie with eye contact, and the other was given Robovie without eye contact. Robovie performed the attention expression for both groups. The target of the attention expression was a poster on a wall. The experiment recorded the number of subjects who looked at the poster according to the attention expression.

The experimental procedure is as follows. At first, the robot passes in front of the subject, and stops in front of the poster, where both the robot and the poster are in the subject's sight. At the location, the robot turns to the subject and points to the poster with its arm while speaking "please look at this". Here, the robot perfumes eye contact to the subject.

Table 2 shows the results. The results indicate that the subjects with eye contact (the upper photo in Fig. 6) look at the poster (the lower photo in Fig. 6), and the subjects without eye contact look at the robot arm instead of the poster. That is, eye contact is significantly effective for achieving joint attention (χ^2=8.57, $p<0.01$); and the robot behaviors designed based on the discussions of the difficulties 1-3 are proper for establishing a communicative relationship with a human.

4. A Robot Architecture for Generating Episode Chains

From the psychological experiments discussed in Section 3, we have obtained four ideas as follows:

1. Rich robot's behaviors induce various human communicative gestures that help utterance understanding.
2. Attention expression by the robot guides the human's focus to the robot attention.
3. Eye contact by the robot indicates robot's intention of communication to the human.
4. Sharing of a joint viewing point and a proper positional relation establish a situation where the human can easily understand robot's utterance.

Based on these ideas, we have designed a new architecture of the robot and implemented to the developed robot

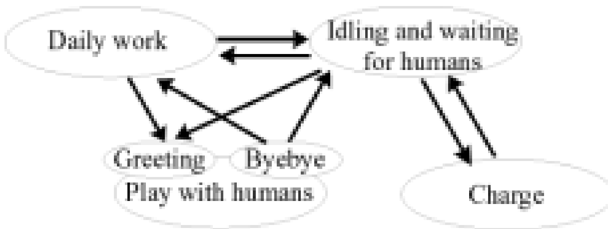

Fig. 7: The meta-structure of the architecture

"Robovie". The basic structure of the architecture is a network of situated behavior modules. Fig. 7 shows the metastructure of Robovie's software. All of the behaviors are classified into four categories; and Robovie performs behaviors belonging to one of them. A unique point is that the category "Play with humans" has two sub-categories of greeting to say "Hello" or "Bye" when switching the category.

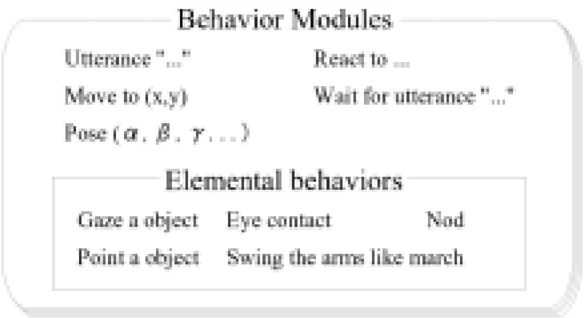

Fig. 8: Behavior modules

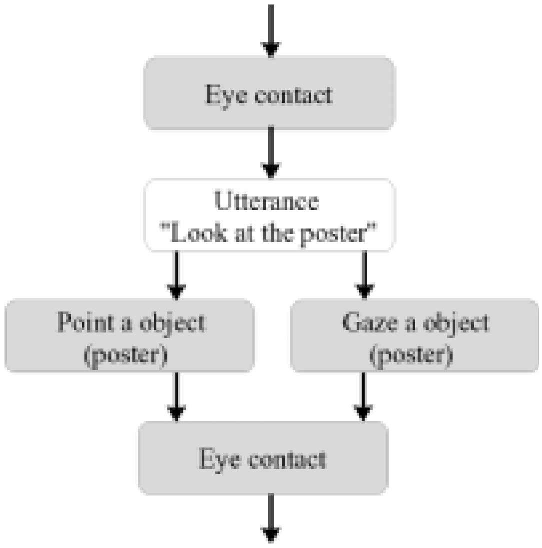

Fig. 9: An example of behavior module

The behavior models belonging to the category include elemental behaviors for communications as shown in Fig. 8. The elemental behaviors that implement the above-mentioned ideas are the most important point in this architecture. The robot behaviors developed so far do not have the function to entrain humans into the communication. By combining the elemental behaviors and other task-oriented behaviors, we can realize various interactive behaviors. Fig. 9 shows an example of the interactive behavior that the robot asks a human to look at a poster.

Fig. 10 shows the all-over software architecture. Basically, this is an extension of the architecture based on situated modules [5]. The architecture proposed in our previous work has two merits: easy development of behavior modules and robust execution by dynamic switching of the behavior network. With keeping the merits, we have extended the architecture.

Episodes between a robot and humans are emerged through interactive behaviors and contextual chains of the behavior. The behavior modules shown in Fig. 10 form a network based on these execution orders, and the network can generate various sequential orders among the interactive behavior modules. By switching the execution order based on sensory input, the robot can generate various episode chains depending on the situations. This episode chain is not a still behavior sequence. A behavior module in a previous robot is activated based on sensory input. On the other hand, our robot controls the behavior sequence based on predefined weak orders even if it does not sufficient sensory data and continuously entrains the human. We consider the episode chains will represent the robot's autonomy.

These two ideas are also important to implement sensory data processing. In previous robotics, robots needed to perform perfect sensor data processing to execute particular tasks. The robot in this paper, however, entrains humans into the interaction loops by the interactive behaviors and it does not require perfect sensory data processing. Humans, rather, adapt to the robot's ability. With the architecture, the robot can continuously generate rich behaviors for communication even if the sensory data processing is not perfect.

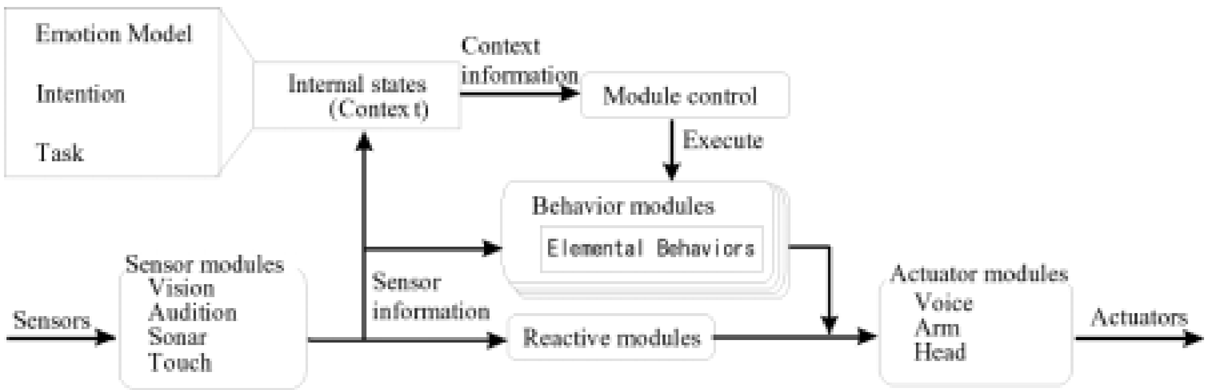

Fig. 10: Software architecture based on behavior modules

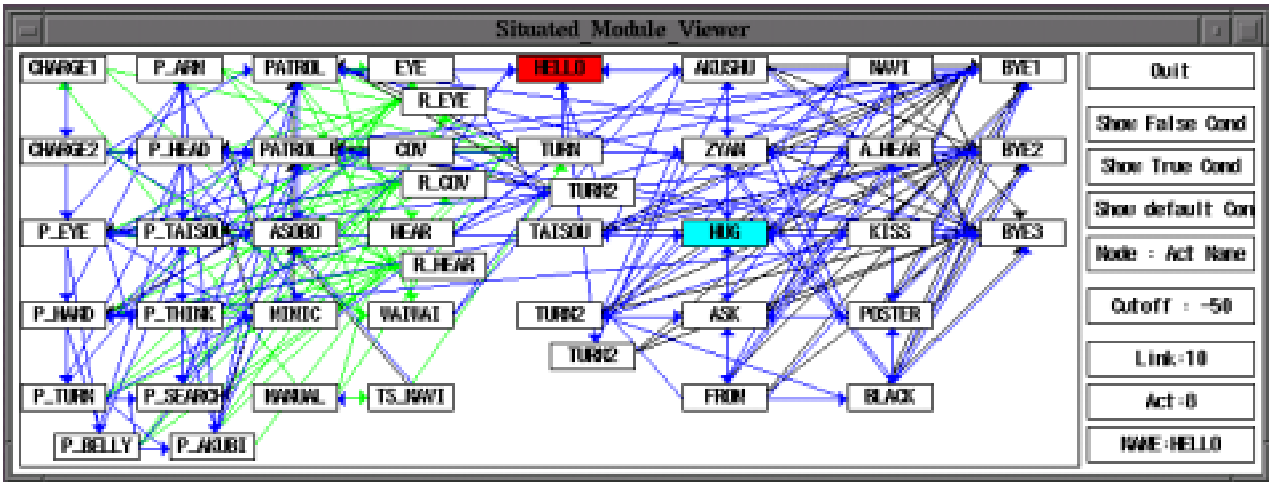

Fig. 11: All behavior modules and their relationships

Finally, Fig. 11 shows all of developed modules and their relationships for demonstration in exhibitions. The robot behaviors generated from the various interactive behaviors and the complicated network has given human-like impressions to attendees in many exhibitions. The playing behaviors Robovie takes are as follows: “greeting”, “hand-shake”, “playing the game of ‘paper, stone and scissors’ ”, “hugging”, “kiss”, “short conversation”, “exercise”, “pointing the poster”, and “saying good bye”. Robovie also takes idling behaviors such as “scratching its head”, “folding its arms”, and so on.

5. Conclusion

This paper has reported on a new humanoid robot called “Robovie”. The unique aspect of Robovie is the mechanism designed for communication. Robovie can generate human-like behaviors with the actuators and sensors. In the designing, we have performed two psychological experiments and developed the behaviors obtained from them. Our next step is to implement more interactive behaviors to Robovie and try to establish more sophisticated relationships between the robot and humans.

We have started this project on August 1999. After the development of Robovie on July 2000, Robovie has appeared in many robot exhibitions and been reported by almost all major newspapers and several TV programs in Japan. These are not only advertisements but also valuable chances to gather comments from ordinary people. For developing a robot work in our daily life, these activities bring much information in addition to the cognitive experiments. For more detail of this project, please refer to the following pages:

http://www.mic.atr.co.jp/~michita/everyday-e/

References

[1] David McNeill, “Psycholinguistics: A New Approach,” Harper & Row, 1987.

[2] T. Ono, M. Imai, and R. Nakatsu, “Reading a robot's mind: a model of utterance understanding based on the theory of mind mechanism,” Advanced Robotics, Vol.14, No.4, pp.311-326, 2000.

[3] D. Sperber and D. Wilson, “Relevance: Communication and Cognition,” Oxford: Basil Blackwell, 1986.

[4] C. Moore and P. J. Dunham, “Joint Attention: Its Origins and Role in Development,” Lawrence Erlbaum Associatcs, Inc., 1985.

[5] H. Ishiguro, T. Kanda, K. Kimoto, and T. Ishida, “A Robot Architecture Based on Situated Modules,” Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 1617-1623, 1999.