

Development and Evaluation of Interactive Humanoid Robots

TAKAYUKI KANDA, MEMBER, IEEE, HIROSHI ISHIGURO, MEMBER, IEEE,
MICHITA IMAI, MEMBER, IEEE, AND TETSUO ONO

Invited Paper

We report the development and evaluation of a new interactive humanoid robot that communicates with humans and is designed to participate in human society as a partner. A human-like body will provide an abundance of nonverbal information and enable us to smoothly communicate with the robot. To achieve this, we developed a humanoid robot that autonomously interacts with humans by speaking and gesturing. Interaction achieved through a large number of interactive behaviors, which are developed by using a visualizing tool for understanding the developed complex system. Each interactive behavior is designed by using knowledge obtained through cognitive experiments and implemented by using situated recognition. The robot is used as a testbed for studying embodied communication. Our strategy is to analyze human-robot interaction in terms of body movements using a motion-capturing system that allows us to measure the body movements in detail. We performed experiments to compare the body movements with subjective evaluation based on a psychological method. The results reveal the importance of well-coordinated behaviors as well as the performance of the developed interactive behaviors and suggest a new analytical approach to human-robot interaction.

Keywords—Body movement analysis, development method, human-robot interaction, psychological evaluation.

I. INTRODUCTION

The recent development of humanoid and interactive robots such as Honda's [1] and Sony's [2] is a new research direction in robotics. The concept of the *partner robot* is rapidly emerging. Partner robots will act as human peers

in everyday life and perform mental and communicational support for human beings as well as physical support. For instance, pet robots have been used successfully in mental therapy for the elderly [3]. Moreover, if robots can communicate with humans by vocal language, the conversational ability of the robots helps humans to retrieve information through a computer network, such as NEC's [4], and creates friendlier relationships with humans. The Matsushita Company has already started a trial to comfort elderly persons with interactive stuffed animal robots.

The partner robots should facilitate effective multimodal communication in order to complete an arbitrary set of tasks together with humans. Interhuman communication employs diverse channels made available by our entire body. By establishing eye contact and observing and possibly imitating gestures, we greatly increase our understanding of others' utterances [5]. It is well known that during conversation, a human immediately detects correspondences between their own body and the body of their partner. This suggests that to produce effective communication skills in an interactive robot, its body should be based on a human's. The previous research on human-robot communication, which is often motivated by cognitive science and psychology, has determined various interactive behaviors that the robot's body should afford. For example, Scassellati developed a robot as a testbed for verifying the effect of joint attention [6]. Matsusaka *et al.* developed a robot that can gaze at the person who is talking with it [7]. Nakadai *et al.* developed a robot that tracks a speaking person [8]. Moreover, by combining the knowledge from cognitive science and robots' behaviors, we can achieve smooth and natural human-robot communication. For example, our robots also utilize their body properties for facilitating interaction with humans [9] and cause people to unconsciously behave as if they were communicating with humans [10].

On the other hand, methods of analyzing partner robots, especially with respect to human-robot interaction, are still

Manuscript received May 26, 2003; revised February 12, 2004.

T. Kanda is with the ATR Intelligent Robotics and Communication Laboratories, Kyoto 619-0288, Japan (e-mail: kanda@atr.jp).

H. Ishiguro is with the ATR Intelligent Robotics and Communication Laboratories, Kyoto 619-0288, Japan, and also with Osaka University, Osaka 565-0871, Japan (e-mail: ishiguro@atr.jp).

M. Imai is with the ATR Intelligent Robotics and Communication Laboratories, Kyoto 619-0288, Japan, and also with Keio University, Kanagawa 223-8522, Japan (e-mail: michita@atr.jp).

T. Ono is with the ATR Intelligent Robotics and Communication Laboratories, Kyoto 619-0288, Japan, and also with Future University Hakodate, Hokkaido 041-8655, Japan (e-mail: tono@atr.jp).

Digital Object Identifier 10.1109/JPROC.2004.835359

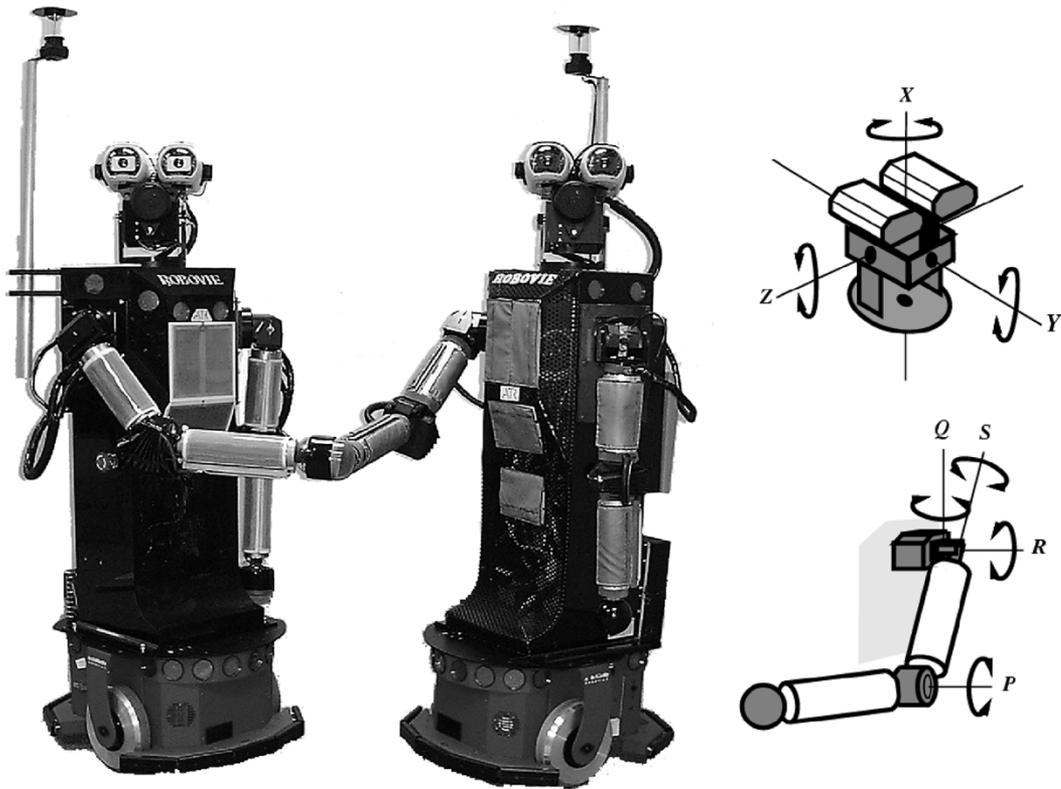


Fig. 1. Robovie—a robot with sufficient physical expression ability.

lacking. To effectively develop any systems in general, it is essential to measure the systems' performance. For example, algorithms are compared with respect to time and memory, and mechanical systems are evaluated by speed and accuracy. Without analyzing current performance, we cannot argue advantages and problems. For partner robots, we need to analyze how a robot affects the humans, in particular the psychological aspects; however, no analysis method has yet been established. Thus, it is vital to determine what types of measurements we can apply. Although questionnaire-based methods have been often used, they are rather subjective, static, and obtrusive (that is, we interrupt the interaction when we apply a questionnaire). Less commonly, human behaviors are employed for this purpose, such as distance [11], attitude [12], eye gaze (often used in psychology), and synchronized behaviors [5]. Although these methods are more difficult to apply, the results are more objective and dynamic. However, they are still fragments rather than a systematic analysis method applicable for human-robot interaction.

This paper presents our exploratory approach for the development and evaluation of interactive humanoid robots. We have developed an interactive humanoid robot that has a human-like body as the testbed of this embodied communication. Furthermore, many interactive behaviors have been implemented. This encourages people to treat the robot as a human child. Our evaluation approach is to measure the body movement interaction between a humanoid robot and humans and compare the results with traditional subjective evaluation. Through the evaluation experiments, we

argue the performance of the developed interactive robots and provide perspectives for this new analytical method for human-robot interaction.

II. DEVELOPMENT OF INTERACTIVE HUMANOID ROBOT

We believe that bottom-up construction is currently a better way to build an interactive humanoid robot, since we do not know how to generally utilize a human-like body in human-robot communication. In other words, our *constructive approach*—to continue implementation until the robot behaves more human-like—aims to establish the ideal top-down model of interactive humanoid robot through exploratory bottom-up construction. As well, it provides a testbed for analyzing human-robot interaction, which produces plenty of knowledge for building interactive robots.

A. Hardware of Humanoid Robot "Robovie"

We have developed a robot named "Robovie," shown in Fig. 1. This robot, which has a human-like appearance, is designed for communication with humans. Like a human, it has various sensors, such as vision, sense of touch, and audition. With its human-like body and sensors, the robot performs meaningful interactive behaviors for humans.

Size is important for an interactive robot. So as not to produce an imposing impression on humans, we limited the size to 120 cm, which is the same as the average junior school student. The diameter is 40 cm and the weight is about 40 kg. The robot has two arms [4 * 2 degrees of freedom (DOF)], a head (3 DOF), two eyes (2 * 2 DOF for gaze control), and

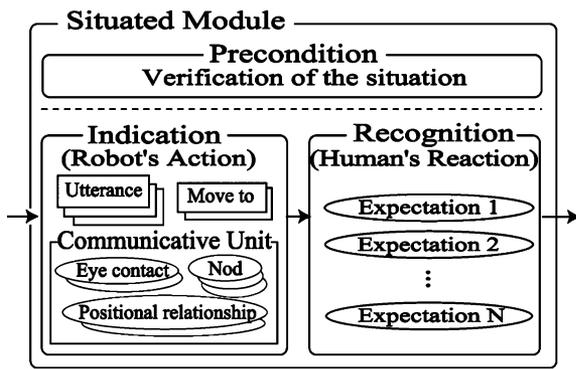


Fig. 2. Software architecture of the developed robot.

a mobile platform (two driving wheels and one free wheel). The robot also has various sensors, 16 skin sensors covering the major parts of the robot, ten tactile sensors around the mobile platform, an omnidirectional vision sensor, two microphones to listen to human voices, and 24 ultrasonic sensors for detecting obstacles. The eyes have a pan-tilt mechanism with direct-drive motors, and they are used for stereo vision and gazing control. The skin sensors are important for realizing interactive behaviors. We developed sensitive skin sensors using pressure-sensitive conductivity rubber. Another important point in the design is the battery life. This robot can work for 4 h and charges the battery itself by autonomously looking for battery-charging stations. With the actuators and sensors, the robot can generate enough behaviors required for communication with humans.

Robovie is a self-contained autonomous robot. It has a Pentium III PC on board for processing sensory data and generating behaviors. 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 the easy and quick development of Robovie's software modules.

B. Software Architecture for Interactive Humanoid Robots

We believe that there is a strong correlation between the number of appropriate behaviors an interactive robot can produce and its perceived intelligence. Our *constructive approach* [13] is to continue to implement behaviors until humans think that the robot has an animated and lifelike existence beyond that of a simple automatic machine. In addition, to supplement the current insufficient sensor-processing ability, we employed an active interaction policy. That is, robots initiate interaction to maintain communicative relationships with humans. It includes the nonverbal information of both robots and humans, which is completely different from linguistic dialogue approaches, such as that in the work of Roy *et al.* [14].

Our robot architecture (shown in Fig. 2) is for implementing a large number of behaviors. The interactive behaviors are designed with knowledge of the robot's embodiment obtained from cognitive experiments and then implemented as *situated modules* with situation-dependent sensory data processing for understanding complex human

behaviors. The relationships between behaviors are implemented as rules governing execution order (named *episode rules*) to maintain a consistent context for communication. As the number of implemented situated modules and episode rules increases, a complicated switching of behaviors will be generated. A development tool named the *Episode Editor* eases the complication of the implementation. It displays the complex relationships and the execution of many simple behaviors visually.

It is a completely bottom-up design, which is quite different from other designs. Developers create situated modules, which execute a particular task in a particular situation, and episode rules that represent their partial execution order. The mechanism of interaction among humans is not yet known, so a top-down design approach is not yet possible.

The basic strategy of implementation is as follows.

- 1) Develop situated modules for various situations.
- 2) Define the basic execution order of the situated modules with episode rules for sequential transition.
- 3) Add episode rules for reactive transitions.
- 4) Modify implemented episode rules and specify episode rules of negation to suppress execution of the situated modules for a particular long-term context.

The visualization support function of the Episode Editor will be helpful, especially at step 4.

1) *Communicative Units*: The *communicative unit* (*communicative sensory-motor unit*) is a very basic unit that achieves sensory-motor action for natural and effective human-robot communication. Experiments in cognitive science produced several essential ideas about the robot's body property, such as in the study by Ono *et al.* [5]. Each communicative unit (as shown in Fig. 3) is retrieved from these ideas. Concretely, we have implemented "eye contact," "nod," "positional relationship," and "joint attention" (gaze and point at object) (Fig. 4).

Although there are not many implemented ideas to date, we can continuously develop such communicative units along with such cognitive knowledge. We, therefore, believe that the communicative ability of the robot will increase along with the development of communicative units.

2) *Situated Modules*: In linguistic research, an adjacency pair is a well-known term for a unit of conversation where the first expression of the pair requires the second expression to be of a certain type (greeting and response, question and answer, and so on.) Similarly, we consider that human-robot interaction can be primarily achieved with action-reaction pairs. That is, when a human acts toward the robot, it reacts to the human's action; when the robot acts toward a human, the human reacts to its action.

The situated module realizes the action-reaction pair as an interactive and reactive behavior (indication-recognition pair) in a particular situation. With the active interaction approach, the robot mainly acts and humans react by sequential transition. Deviation from the basis is treated by reactive transition and *reactive modules*. The robot executes just one situated module at a time.

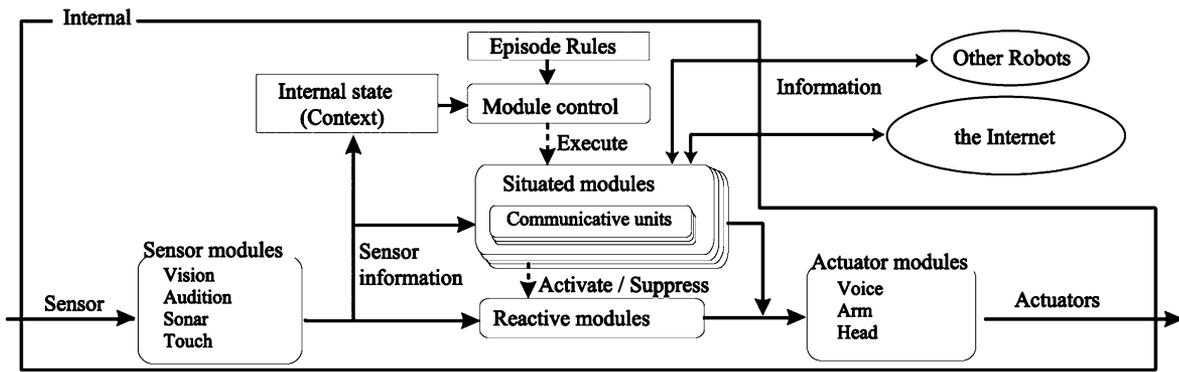


Fig. 3. Situated module and communicative units.

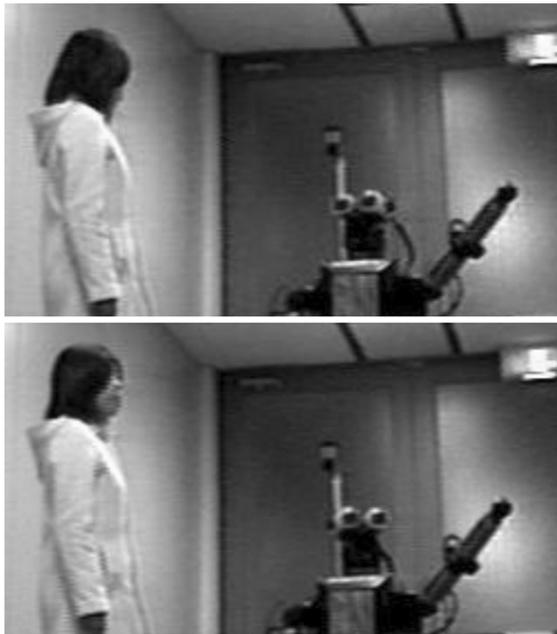


Fig. 4. A scene of joint attention (eye contact and pointing to share the attention).

When developers create a situated module, they combine the communicative units first. Then, they supplement with other sensory-motor units such as utterances and positional movements for particular interactive behaviors. (Fig. 5)

a) Precondition, indication, and recognition parts: Each situated module consists of precondition, indication, and recognition parts (Fig. 3) and is implemented using communicative units. By executing the precondition, the robot system checks whether the situated module is in an executable situation or not. For example, the situated module that produces talking about the weather by retrieving weather information from the Internet is not executable (precondition is not satisfied) when the robot system cannot connect to the Internet.

By executing the indication part, the robot takes an action to interact with humans. For example, for the handshake module, it says, "Let's shake hands," and offers its hand.

The recognition part is designed to recognize several kinds of human reactions toward the robot's action caused by the

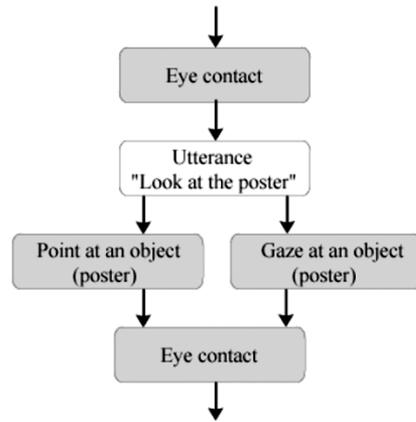


Fig. 5. Development of a situated module with communicative units.

indication part. That is, it is an expectation of the human's reaction. The situated module itself produces the particular situation, and then it can recognize complex human behaviors in the particular situation.

b) Sequential and reactive transition of situated modules and reactive modules: After the robot executes the indication part of the current situated module, it recognizes the human's reaction by the recognition part. Then, it ends the execution of the current situated module, records the result value corresponding to the recognition result, and transits to the next executable situated module [Fig. 6(a)]. The next module is decided on by the result value of the current situated module and the execution history of the situated modules. This sequential transition is ruled by episode rules.

Thus, the human-robot communication for a consistent context can be realized by sequential transition. However, there are important problems in communication that are unresolved by sequential transition: interruption and deviation. For example, when humans talk to each other and suddenly a telephone rings, they will stop the conversation to respond to the telephone call. Such an interruption and deviation is dealt with by reactive transition and reactive modules. Reactive transition is also ruled by episode rules [Fig. 6(b)]. If the reactive transition is assigned for the current situation and the precondition of the corresponding next situated module

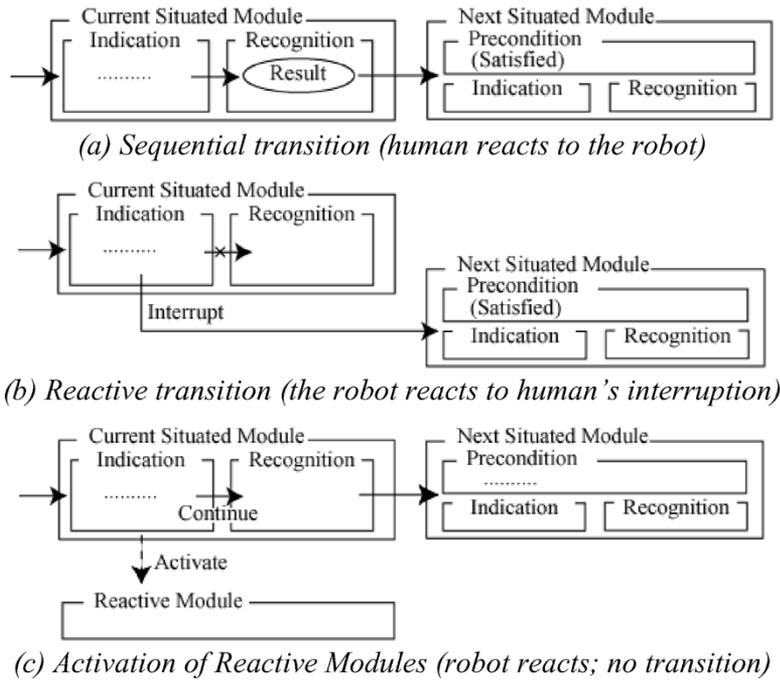


Fig. 6. Transition of situated modules.

is satisfied, the robot quits executing the current situated module and instantly transits to the next situated module.

Reactive modules are also prepared for an interruption, but in this case the robot does not quit the execution of the current situated module when a reactive module is activated [Fig. 6(c)]. Instead, the robot system executes the reactive module in parallel with the current situated module. For example, we implemented a reactive module to make the robot gaze at the part of its body being touched. When the robot talks to a human and the human suddenly touches the arm of the robot, the robot gazes at the arm to indicate that it has noticed it, but continues talking. Similar to subsumption architecture [15], upper hierarchy modules (situated modules) can suppress lower ones (reactive modules).

c) *Other components of the architecture:* The architecture has components for communication using computer networks. This is a new information infrastructure that lets robots keep humans informed by communicating with them in natural language. For example, when the robot and humans talk about weather, the robot obtains weather information from the Internet. If the forecast is rain, it says, “It will rain tomorrow.” The *internal state* represents the context based on the execution history of the situated modules. Inputs from sensors are preprocessed at *sensor modules* such as speech recognition. *Actuator modules* perform low-level control of actuators.

3) *Episode Rules:* Episode rules guide the robot into a new episode of interaction with humans by controlling transitions between situated modules. All episode rules are compared with the current situated module and the execution history of the situated modules to determine which situated module to execute next. The system performs the comparison in the background of the current situated module’s execution and prepares the next executable module list. After the cur-

Table 1
Grammar of Episode Rules

1. $\langle \text{ModuleID}=\text{result_value} \rangle \dots \langle \dots \rangle \text{NextModule}$
 2. $\langle \text{ModuleID1}=\text{result_value1} \rangle \langle \text{ModuleID2}=\text{result_value2} \rangle \dots$
 3. $(\dots)\{n,m\} \dots$
 4. $! \langle \dots \rangle \text{NextModule}$
 5. $\wedge \langle \text{ModuleID}=\wedge \text{result_value} \rangle \text{NextModule}$
- (1: basic structure of describing executed sequence,
2: “OR”, 3: repetitions, 4: negation of *episode rule*,
5: negation of Module ID and result value)

rent module’s execution, the robot checks the preconditions of each situated module in the list. If the precondition is satisfied, it transits to the next situated module. Each episode rule has a priority. If some episode rules conflict, the episode rule with higher priority is used.

Table 1 indicates the basic grammar of the episode rule. Each situated module has a unique identifier called a *Module ID*. $\langle \text{ModuleID} = \text{result_value} \rangle$ is the rule that refers to the execution history and the result value of the situated modules, and then $\langle \text{ModuleID1} = \text{result_value1} \rangle \langle \text{ModuleID2} = \text{result_value2} \rangle \dots$ means a referring rule of the previously executed sequence of situated modules (Table 1, item 1). $\langle \dots \rangle \mid \langle \dots \rangle$ means a selective group (OR) of the executed situated modules, and then (\dots) means the block that consists of a situated module, a sequence of situated modules, or a selective group of situated modules (Table 1, item 2). Similar to the regular expression, we can describe the repetition of the block as $(\dots)\{n,m\}$, where n gives the minimum number of times to match the block and m gives the maximum (Table 1, item 3). We can specify the negation of the whole episode rule with an exclamation mark (!). For example, $! \langle \dots \rangle \dots \langle \dots \rangle \text{NextModuleID}$ (Table 1, item 4) means that the module of NextModuleID will *not* be ex-

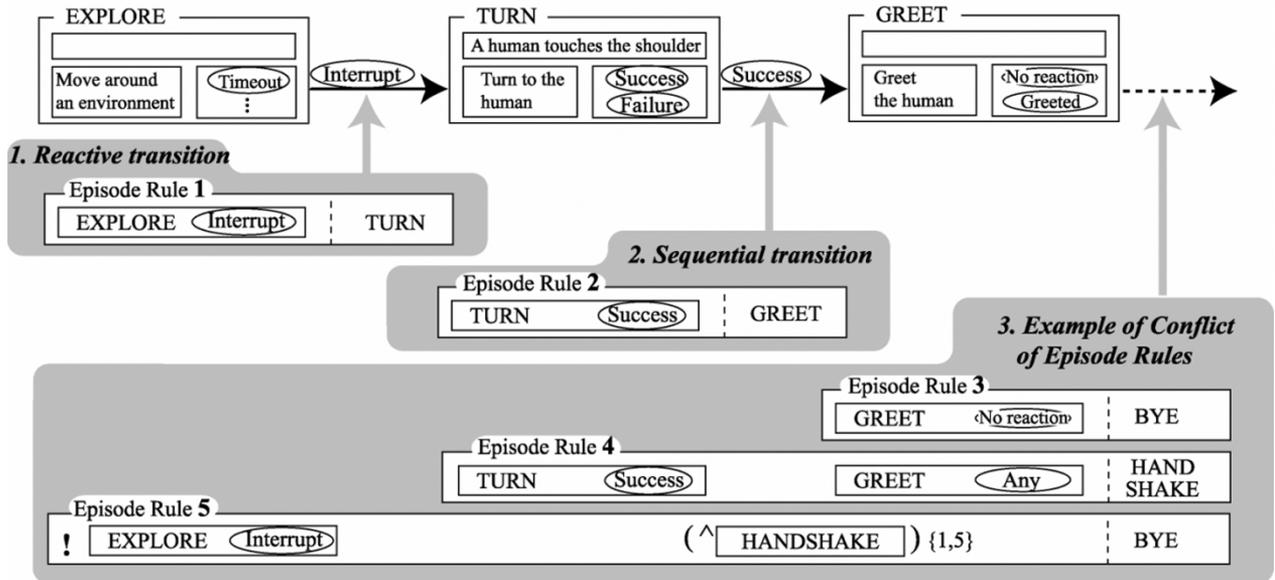


Fig. 7. Illustrated example of the transitions of the current situated modules ruled by episode rules.

executed when the episode rule matches the current situation specified by $\langle \dots \rangle \dots \langle \dots \rangle$. The negation of ModuleID and the result value can be written as a caret character (^) (Table 1, item 5).

Fig. 7 is an example of a transition. At first, the robot explores the environment by EXPLORE. Then, a human touches the shoulder of the robot. This action causes a reactive transition ruled by episode rule 1 (Fig. 7). The robot turns to the human by executing the situated module TURN. After the execution of TURN, it starts to greet him/her by executing GREET. This sequential transition is caused by episode rule 2 (Fig. 7).

Next, we explain the conflict of episode rules (Fig. 7, rule 3). When GREET results in “no reaction,” BYE is the candidate of the next situated module selected by episode rule 3. Meanwhile, HANDSHAKE is the candidate selected by episode rule 4. Episode rule 5 is a negative episode rule to suppress the transition to BYE (it specifies that the robot should not say goodbye before the handshake once its exploration has been interrupted). If the priority of episode rule 5 is higher than that of episode rule 3, BYE is not the candidate of the next execution.

4) *Episode Editor:* We developed the Episode Editor, which visually displays the complex relationships and execution of many simple situated modules. This enables us to intuitively develop a large number of situated modules and relationships among them (episode rules). The Episode Editor has three main functions.

a) *Implementation of episode rules:* The Episode Editor has the obvious function of editing episode rules. In Fig. 8, half of the left of the screen displays the episode rule that a developer is editing. Each box indicates one block of situated modules. If there is a complex block assigned by (...), the block is expanded in the left-second box of the screen. The rightmost box on the left screen indicates the next executable situated module. The right half of the screen shows the detail of the selected box in the left screen, where developers can indicate the identifier of the situated modules

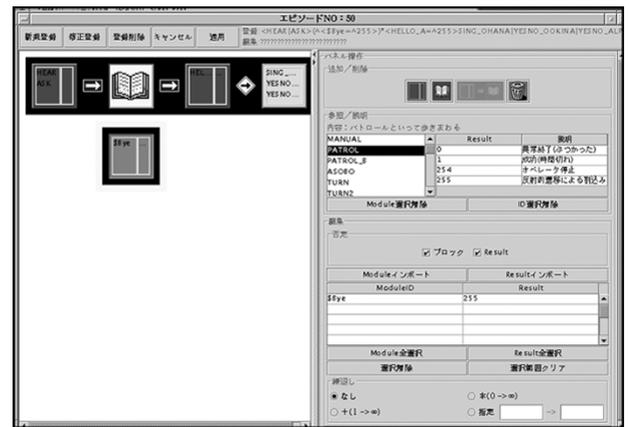


Fig. 8. Editing screen of the episode editor: developing or modifying an episode rule.

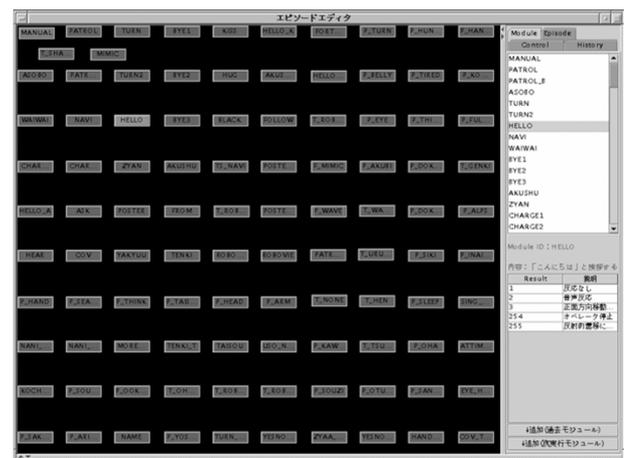


Fig. 9. Main screen of the episode editor: all situated modules (gray boxes) are displayed in rows and lines before calculation of positions.

(Module ID), result values, repetition, and negation. This editing screen is opened from the main screen (Fig. 9),

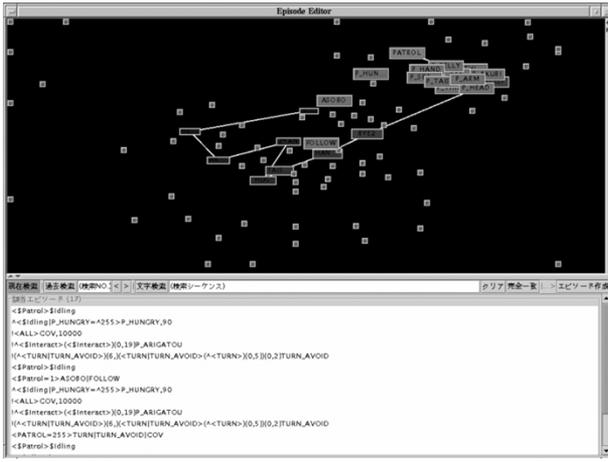


Fig. 10. Displaying the current status and search result. Upper screen displays the current status, lower screen shows the search result of the episode rules.

which has two windows: a list of all implemented episode rules and the search result of the episode rules. Developers can select one of the episode rules from the lists to edit.

b) *Visualization of situated modules and episode rules*: The Episode Editor has two types of visualization support functions for past episodes: the position of situated modules for the entire execution history and the size and color for the recent history and current state. The situated modules are placed according to their relationships (Fig. 10), which are calculated from the execution history. For example, if situated module X is frequently executed after Y , then X and Y have a strong relationship and appear near to each other in the main screen. The force of this relationship between two situated modules is calculated by the following formula, which is known as the spring model method:

$$F_{ij} = K_{ij} \cdot D_{ij} - R \cdot D_{ij}^{-7}$$

where i and j are situated modules; F_{ij} , D_{ij} , and K_{ij} are the force, distance, and spring constant between i and j , respectively; and R is the constant for the force of repulsion. The spring constants are retrieved from the execution history. The position of each situated module is iteratively calculated until all positions converge.

The Episode Editor also helps our visualization of the recent execution history of situated modules (Fig. 10: upper half of the screen). The recently executed situated modules are connected and colored blue (dark gray in the figure). The sizes are small because the modules were executed in the past. The current situated module is large and colored red (light gray in the figure). The candidates of the next execution are also large and colored green (gray in the figure). Other situated modules are displayed very small.

c) *Searching episode rules*: The third function of the Episode Editor is to search implemented episode rules. There are three types of searches: current search, recent history search, and direct search. In the current search (Fig. 10), the Episode Editor shows the current status (upper half of the screen) and the episode rules (bottom half) match with the current status and history. In the history search, a user can



Fig. 11. Interactive behaviors.

specify a point in the history to search the episode rules that match a specified time. In the direct search, a user can select situated modules from the main screen to specify the situation (such as $\langle TURN = success \rangle \langle GREET = greeted \rangle$) to search the episode rules.

C. Realized Interactive Behaviors

Robovie’s task is to perform daily communication in the manner of children. The number of developed situated modules has reached 100: 70 are interactive behaviors such as handshake (Fig. 11, upper left), hugging (Fig. 11, upper right), playing paper–scissors–rock (Fig. 11, lower left), exercising (Fig. 11, lower right), greeting, kissing, singing a song, engaging in a short conversation, and pointing to an object in the surroundings; 20 are idling behaviors such as scratching its head and folding its arms; and ten are moving-around behaviors, such as pretending to patrol an area and going to watch an object in the surroundings.

Basically, the transition among the situated modules is implemented as follows: the robot sometimes asks humans for interaction by saying, “Let’s play, touch me,” and exhibits idling and moving-around behaviors until a human acts in response; once a human reacts to the robot (touches or speaks), it starts and continues the friendly behaviors while the human reacts to these; when the human stops reacting, it stops the friendly behaviors, says good-bye, and restarts its idling or moving-around behaviors.

III. EVALUATION OF HUMAN–HUMANOID ROBOT INTERACTION

In this section, we report on an evaluation of the interaction between humans and the developed robots. The robot uses its human-like body and voice for accomplishing smooth and natural communication with humans, behaves like a human child, and attempts daily communication with humans. Our strategy is to analyze the human–robot interaction in terms of body movements using a motion-capturing system. We performed experiments to compare the body movements with subjective evaluation, which is based on a psychological method. In addition to the evaluation of the interaction, we intend to discover knowledge on embodiment that partner robots can utilize to encourage humans to interact with them.

Table 2
The Adjective Pairs Used for Subjective Evaluation and the Mean, Resulting Mean, and Standard Deviation

Adjective-pairs		Mean	Std. Dev.
Good	Bad	4.88	0.95
Kind	Cruel	4.85	1.29
Pretty	Ugly	5.08	0.93
Exciting	Dull	4.46	1.61
Likable	Unlikable	4.77	1.03
Evaluation score		4.81	0.92

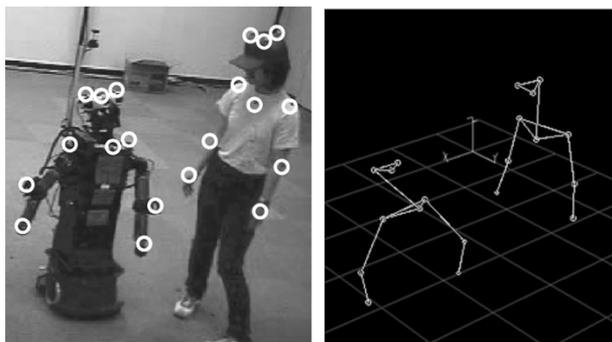


Fig. 12. Attached markers (left) and obtained 3-D.

A. Experiment Settings

We used 26 university students (19 men and seven women) as our subjects. Their average age was 19.9 years. First, they were shown examples of how to use the robot. Then, they freely observed the robot for ten minutes in a rectangular room 7.5 m by 10 m. As described in Section II-C, the robot autonomously tried to interact with the subjects. At the beginning of the free observation, the robot asked the subjects to talk and play with it, and then the subjects usually began touching and talking to the robot.

After the experiment, the subjects answered a questionnaire about their subjective evaluations of the robot with five adjective pairs shown in Table 2, which was compared to the body movements. We chose these adjective pairs because they had high loadings as evaluation factors for an interactive robot in a previous study [10].

B. Measurement of Body Movements

We employed an optical motion-capturing system to measure the body movements. The motion-capturing system consisted of 12 pairs of infrared cameras and infrared lights and markers that reflect infrared signals. These cameras were set around the room. The system calculates each marker's 3-D (three-dimensional) position from all camera images. The system has high resolution in both time (120 Hz) and space (accuracy is 1 mm in the room)

As shown in Fig. 12, we attached ten markers to the heads (subjects wore a cap attached with markers), shoulders, necks, elbows, and wrists of both the robot and the subjects. By attaching markers to corresponding places on the robot and subjects, we could analyze the interaction of body movements. The three markers on the subjects' heads detect the individual height, facing direction, and potential eye

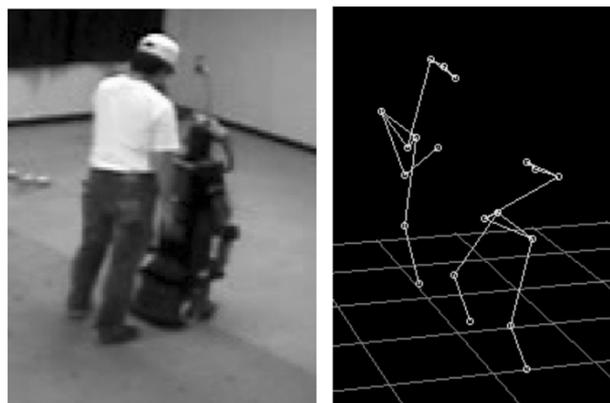


Fig. 13. Scene of eye contact.

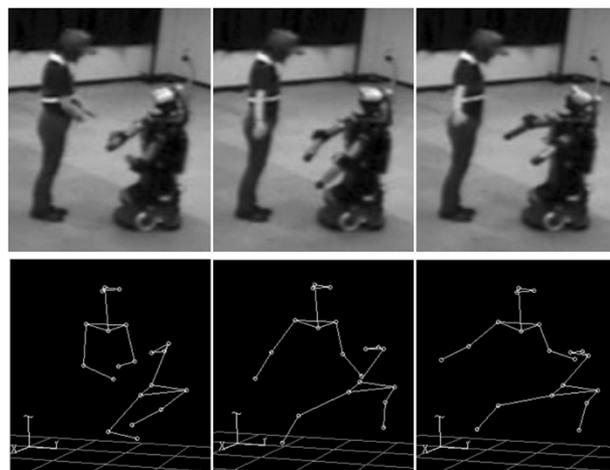


Fig. 14. Scene of synchronized body movements.

contact (Fig. 13) with the robot. The markers on the shoulders and neck are used to calculate the distance between the robot and subjects and the distance moved by them. The markers on the arms provide hand movement information (the relative positions of the hands from the body) and the duration of synchronized movements (Fig. 14: the period where the movements of the hands of the subject and robot highly correlate). We also analyzed touching behaviors via an internal log of the robot's touch sensors.

C. Results

Comparison between the body movements and the subjective evaluations indicates meaningful correlation. From the experimental results, well-coordinated behaviors such as eye contact and synchronized arm movements proved to be important. This suggests that humans make evaluations based on body movements.

1) *Subjective Evaluation: "Evaluation Score"*: The semantic differential method is applied to obtain subjective evaluations with a one-to-seven scale, where seven denotes the most positive point on the scale. Since we chose adjective pairs that had high loadings as evaluation factors for an interactive robot, the results of all adjective pairs represent subjective evaluation of the robot. Thus, we calculated the *evaluation score* as the average of all adjective pairs' scores.

Table 3
Results for Body Movement

	Mean	Std. Dev.
Distance (m)	0.547	0.103
Eye contact (s)	328	61.8
Eye height (m)	1.55	0.124
Distance moved (m)	35.2	17.0
Distance moved by hands (m)	108	29.5
Synchronized movements (s)	7.95	6.58
Touch (num. of times)	54.9	20.8

Table 4
Correlation Between Subjective Evaluation and Body Movements

	Evaluation	Dist.	E. C.	E. H.	D.M.	D.M. H.	S. M.	Touch
Dist.	-0.04	1.00						
E.C.	0.57	-0.47	1.00					
E.H.	0.08	-0.39	0.29	1.00				
D.M.	-0.32	0.20	-0.43	0.02	1.00			
D.M.H.	0.01	-0.04	-0.21	-0.09	0.49	1.00		
S. M.	0.54	-0.05	0.28	-0.05	0.15	0.61	1.00	
Touch	0.21	-0.45	0.49	-0.07	-0.15	0.35	0.41	1.00

Table 2 indicates the adjective pairs used, the averages, and the standard deviations.

2) *Correlation Between Body Movements and Subjective Impressions*: Table 3 displays the measured body movements. Regarding eye contact, the average time was 328 s, which is more than half of experiment time. Since the robot's eye height was 1.13 m and the average subject eye height was 1.55 m, which was less than their average standing eye height of 1.64 m, several subjects sat down or stooped to bring their eyes to the same height as the robot's. The distance moved was farther than what we expected, and it seemed that subjects were always moving little by little. For example, when the robot turned, the subjects would then correspondingly turn around the robot. Some subjects performed arm movements synchronized with the robot's behaviors, such as exercising.

Next, we calculated the correlation between the evaluation score and the body movements (Table 4). Since the number of subjects is 26, each correlation value whose absolute value is larger than 0.3297 is significant. We highlight these significant values with boldface in the table. From the calculated results, we found that eye contact and synchronized movements indicate higher significant correlations with the evaluation score.

According to the correlations among body movements, the following items showed significant correlations: eye contact–distance, eye contact–distance moved, synchronized behaviors–distance moved by hands, and synchronized behaviors–touch. However, these items (distance, distance moved, distance moved by hands, and touch) do not significantly correlate with the evaluation score. That is, only the well-coordinated behaviors correlate with the subjective evaluation. Isolated active body movements of subjects, such as standing near the robot, moving their hands energetically, and touching the robot repetitively, do not correlate to the subjective evaluation.

Table 5
Standardized Partial Regression Coefficients Obtained by Multiple Linear Regression Analysis

	Coefficient	Value
Distance	α_{dist}	0.173
Eye contact	α_{ec}	0.476
Eye height	α_{eh}	0.019
Distance moved	α_{dm}	-0.228
Distance moved by hands	α_{dmh}	-0.029
Synchronized movements	α_{sm}	0.535
Touch	α_{touch}	-0.186

3) *Estimation of Momentary Evaluation: "Entrainment Score"*: The results indicate that there are correlations between subjective evaluation and body movements. We performed multiple linear regression analysis to estimate the evaluation score from the body movements, which confirms the above analysis and reveals how much each body movement affects the evaluation. We then applied the relations among body movements to estimate a momentary evaluation score called the *entrainment score*.

As a result of the multiple linear regression analysis, standardized partial regression coefficients were obtained, as shown in Table 5. The obtained multiple linear regression is as follows:

$$E = \alpha_{dist} \bullet DIST + \alpha_{ec} \bullet EC + \alpha_{eh} \bullet EH + \alpha_{dm} \bullet DM + \alpha_{dmh} \bullet DMH + \alpha_{sm} \bullet SM + \alpha_{touch} \bullet TOUCH + \alpha_{const} \quad (1)$$

where DIST, EC, EH, DM, DMH, SM, and TOUCH are the standardized values of the experimental results for the body movements. Since the evaluation was scored on a one-to-seven scale, evaluation score E is between one and seven. The multiple correlation coefficient is 0.77; thus, 59% of the evaluation score is explained by the regression. The validity of the regression is proved by analysis of variance ($F(7, 18) = 3.71, P < 0.05$).

The coefficients (Table 5) also indicate the importance of well-coordinated behaviors. Eye contact and synchronized movements positively affected the evaluation score; on the contrary, distance, distance moved, and touch seem to have negatively affected the evaluation score. In other words, the subjects who just actively did something (stood near the robot, moved around, and touched repeatedly), especially without cooperative behaviors, did not evaluate the robot highly.

Because we can momentarily observe all terms involved in the body movements of the regression (1), we can estimate a momentary evaluation score by using the same relations among body movements as follows:

$$E(t) = \alpha_{dist} \bullet DIST(t) + \alpha_{ec} \bullet EC(t) + \alpha_{eh} \bullet EH(t) + \alpha_{dm} \bullet DM(t) + \alpha_{dmh} \bullet DMH(t) + \alpha_{sm} \bullet SM(t) + \alpha_{touch} \bullet TOUCH(t) + \alpha_{const} \quad (2)$$

where designations such as $DIST(t)$ are the momentary values of the body movements at time t . We named this momentary evaluation score the *entrainment score*, with the

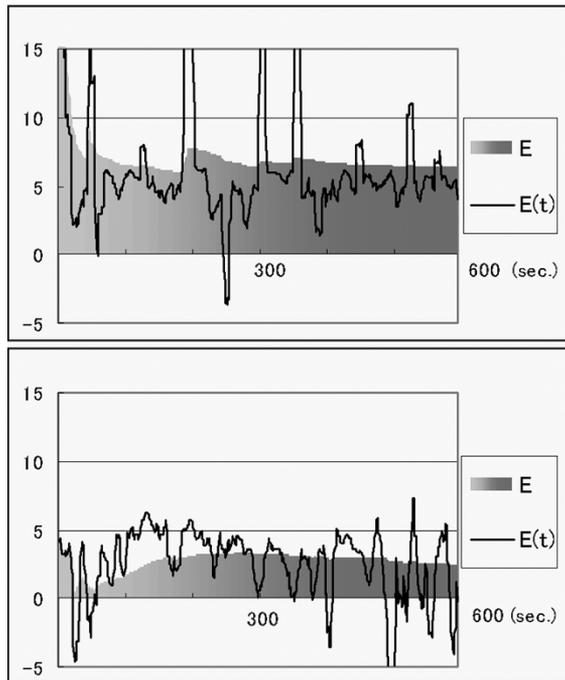


Fig. 15. Illustration of entrainment score.

idea that the robot entrains humans into interaction through its body movements and humans move their body according to their current evaluation of the robot. The evaluation score and entrainment score satisfy the following equation, which represents our hypothesis that the evaluation forms during the interaction occurring through the exchange of body movements:

$$E = \int_0^t \frac{E(t)}{t}. \quad (3)$$

Let us show the validity of the estimation by examining the obtained entrainment score. Fig. 15 shows the entrainment scores of two subjects. The horizontal axis indicates the time from start to end (600 s) of the experiments. The solid line indicates the entrainment score $E(t)$, while the shaded region indicates the average of the entrainment score $E(t)$ from the start to time t (this integration value increased the estimation of E at the end time).

The upper graph shows the score of the subject who interacted with the robot very well. She reported after the experiment that “it seemed that the robot really looked at me because of its eye motion. I nearly regarded the robot as a human child with an innocent personality.” This entrainment-score graph hovers around five and sometimes goes higher. This is because she talked to the robot while maintaining eye contact. She performed synchronized movements corresponding to the robot’s exercising behaviors, which caused the high value around 200 s.

At the other extreme, the lower graph is for the subject who became embarrassed and had difficulty interacting with the robot. The graph sometimes falls below zero. In particular, at the end of the experiment, it became unstable and even lower. He covered the robot’s eye camera, touched it like he was irritated, and went away from the robot. We believe these two

Table 6
Worst Five Situated Modules Based on Average Entrainment Score

ID	Contents	Evaluation
TICKLE	Tickle	-2.09
APOLOGIZE	Apologize	-1.96
NOT_TURN	Say, “I’m busy,” and refuse to play together	-0.51
SLEEP_POSE	A pose of sleeping	-0.42
FULLY_FED	A pose of being fully fed	0.32

Table 7
Best Five Situated Modules Based on Average Entrainment Score

ID	Contents	Evaluation
EXERCISE	Exercise	5.75
ASK_SING	Ask humans, “May I sing a song?”	5.59
CONDUCTOR	Pose imitating a musical conductor	4.85
WHERE_FROM	Ask humans, “Where are you from?”	4.55
LET’S_PLAY	Say, “Let’s play, touch me”	4.24

examples demonstrate the validity of the entrainment score estimation.

4) *Evaluation of the Implemented Behaviors:* In the sections above, we explained the analysis of body movement interaction. Here, we evaluate the implemented behaviors. Although the application of this result is limited to our approach, our findings also prove the validity and applicability of the entrainment score.

We calculated the evaluation score of each situated module based on the average of the entrainment score while each module was being executed. Tables 6 and 7 indicate the worst and best five modules, respectively, and their scores. The worst modules were not very interactive. SLEEP_POSE and FULLY_FED do not respond to human action and exhibit behavior similar to the sleeping pose. NOT_TURN is the behavior for brushing off a human’s hand while saying, “I’m busy,” when someone touches on its shoulder. The best modules were rather interactive modules that entrain humans into the interaction. EXERCISE and CONDUCTOR produce the exercising and imitating of musical conductor behaviors, which induced human synchronized body movements. Other highly rated modules also produce attractive behaviors, such as asking and calling, which induce human reactions. We believe that the entrainment scores provide a wealth of information for developing interactive behaviors of robots that communicate with humans.

IV. DISCUSSIONS

A. Evaluation of the Developed Robot

In our approach, the robot is proactive in initiating interactions and entices humans to respond adaptively to their actions. Our previous experiment [10] suggests that this proactive behavior pattern has more merit in inducing human behaviors than a passive behavior pattern. The robot’s embodiment (head, eyes, arms, etc.) also helps actively entrain humans during interaction. According to the humans’ active and reactive behaviors, the robot selects suitable interactive behaviors to promote the interaction.

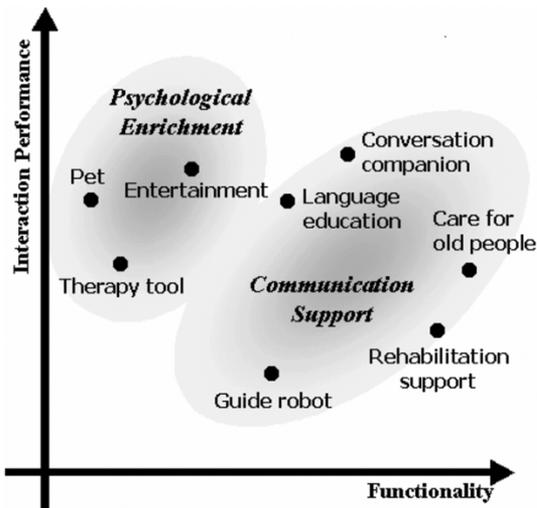


Fig. 16. Possible applications of partner robots.

The experiment results indicate the high performance of the developed robot in the interaction. The subjects behave with the robot as if they were interacting with a human child. On average, they kept eye contact with the robot through more than half of the experiment's duration. Some of them joined in physical interaction such as exercise (synchronously imitated the robot's arm movements). We believe that the human-like appearance and complex interactive behaviors of the robot caused these interhuman behaviors.

B. Perspective for Partner Robots

If we can achieve enough interaction functions of the robots, various applications of interactive humanoid robots will emerge, and they will participate in our daily life. Fig. 16 indicates the possible applications of partner robots. Of course, these robots are capable of direct *psychological enrichment* tasks, such as mental therapy and entertainment, in the same way as animal-like pet robots. In addition to these fundamental functions, the human-like appearance will enable the humanoid robot to smoothly perform *communication support* tasks, such as conversation companionship and language education, as well as physical support tasks in the human daily environment.

Meanwhile, since research on partner robots has just started, various research topics remain in this field. In particular, *evaluation and analysis methods* is a unique and important topic for partner robots, since we need to measure the psychological effect as well as the physical effect of partner robots, while traditional systems can be analyzed by physical measures only. We believe that establishing evaluation methods will promote the development of partner robots. Regarding development, there are many elemental technologies needed for partner robots, such as speech recognition, vision-based recognition, control of complex hardware systems, learning in interaction, and adaptation to individuals.

C. Toward Evaluation Methods of Partner Robots

The experiment reveals the correlation between humans' subjective evaluations and body movements. If a human eval-

uates the robot highly, then the human behaves cooperatively with it, which will further improve its evaluation. That is, once they establish cooperative relationships with the robot, they interact well with the robot and evaluate the robot favorably. Regarding evaluation of the implemented behaviors, the modules that entrain humans into interaction were highly evaluated, such as asking something that induces a human's answer, and producing cheerful body movements like exercising to let humans join and mimic the movements. We believe that the entrainment can help us to establish cooperative relationships between humans and robots.

Meanwhile, the multiple linear regression explains 59% of the subjective evaluation. This is remarkable because it is performed without regard to the contents or context of language communication. With speech recognition, the robot can talk with humans, although its ability is similar to that of a little child. Some of the subjects spoke to the robot. Often, there were requests for the robot to present particular behaviors (especially behaviors it had performed just previously), to which it sometimes responded correctly and sometimes incorrectly. To analyze this, we could use several analytical methods such as conversation analysis. However, these methods are rather subjective. On the other hand, our evaluation employed objective measures only *numerically obtained body movements without context*, which means there could be a lot of potential usages.

Furthermore, this objectivity would provide powerful means for interactive robots. If these robots can estimate how interacting humans evaluate them, they can adaptively adjust their behaviors according to the estimation. In other words, interactive robots could learn and adjust their behavior in the interaction with humans. In addition, we believe this method would be applicable to different subjects (age, culture, etc.), different agents (physical-virtual, body shape, behaviors, etc.), and interhuman communication.

V. CONCLUSION

This paper reported our exploratory approach for developing and evaluating interactive humanoid robots. The developed robot Robovie has a human-like appearance and various sensors for interacting with humans. Regarding the development, we adopted a *constructive approach* (implement as many interactive behaviors as possible with a bottom-up design) to realize the autonomy and complexity of the interactive robot, which behaves like a human child and attempts daily communication with humans.

We performed an experiment to evaluate the developed robot and analyzed the interaction between the robot and humans. In the experiment, the humans behave as if they were interacting with a human. They kept eye contact with the robot and imitated gestures of the robot. These entrainments of body movements indicate the high performance of the developed robot in the interaction. As the result of further analysis on body movements, positive correlations between cooperative body movements and subjective evaluations were discovered. Furthermore, the multiple linear regression explains 59% of the subjective evaluation without regard to language communication.

Recently, interactive humanoid robots are being put to practical use for educational tasks, mentally healing tasks, and so forth. We believe that the numerical analysis of body movements, especially estimation of momentary evaluation, will make interactive robots more adaptive in interacting with humans. Also, our approach of numerical analysis of body movements is widely applicable to various tasks in embodied communication, such as designing the behaviors of robots and analyzing human communication.

REFERENCES

- [1] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent ASIMO: System overview and integration," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2002, pp. 2478–2483.
- [2] M. Fujita, "AIBO: Toward the era of digital creatures," *Int. J. Robot. Res.*, vol. 20, no. 10, pp. 781–794, 2001.
- [3] K. Wada, T. Shibata, T. Saito, and K. Tanie, "Analysis of factors that bring mental effects to elderly people in robot assisted activity," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, vol. 2, 2002, pp. 1152–1157.
- [4] Personal robot PaPeRo. NEC Co. [Online]. Available: http://www.incx.nec.co.jp/robot/PaPeRo/english/p_index.html
- [5] T. Ono, M. Imai, and H. Ishiguro, "A model of embodied communications with gestures between humans and robots," in *Proc. 23rd Annu. Meeting Cognitive Science Soc.*, 2001, pp. 732–737.
- [6] C. Breazeal and B. Scassellati, "A context-dependent attention system for a social robot," in *Proc. Int. Joint Conf. Artificial Intelligence*, 1999, pp. 1146–1151.
- [7] Y. Matsusaka *et al.*, "Multi-person conversation robot using multimodal interface," in *Proc. World Multiconf. Systems, Cybernetics and Informatics*, vol. 7, 1999, pp. 450–455.
- [8] K. Nakadai, K. Hidai, H. Mizoguchi, H. G. Okuno, and H. Kitano, "Real-time auditory and visual multiple-object tracking for robots," in *Proc. Int. Joint Conf. Artificial Intelligence*, 2001, pp. 1425–1432.
- [9] H. Ishiguro, T. Ono, M. Imai, and T. Kanda, "Development of an interactive humanoid robot "Robovie"—An interdisciplinary approach," in *Robotics Research*, R. A. Jarvis and A. Zelinsky, Eds. New York: Springer-Verlag, 2003, pp. 179–191.
- [10] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, "Development and evaluation of an interactive humanoid robot Robovie," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2002, pp. 1848–1855.
- [11] E. Hall, *The Hidden Dimension*. New York: Anchor, 1969.
- [12] B. Reeves and C. Nass, *The Media Equation*. Stanford, CA: CSLI, 1996.
- [13] T. Kanda, H. Ishiguro, M. Imai, T. Ono, and K. Mase, "A constructive approach for developing interactive humanoid robots," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2002, pp. 1265–1270.
- [14] N. Roy, J. Pineau, and S. Thrun, "Spoken dialogue management using probabilistic reasoning," presented at the Conf. Association for Computational Linguistics (ACL), Hong Kong, China, 2000.
- [15] R. A. Brooks, "A robust layered control system for a mobile robot," *IEEE J. Robot. Automat.*, vol. RA-2, pp. 14–23, Mar. 1986.



Takayuki Kanda (Member, IEEE) received the B. Eng, M. Eng, and Ph. D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively.

From 2000 to 2003, he was an Intern Researcher at ATR Media Information Science Laboratories. He is currently a Researcher at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. His current research interests include intelligent robotics, human–robot interaction, and vision-based mobile robots.

Dr. Kanda is a Member of the Robotics Society of Japanese, the Information Processing Society of Japan, and the Japanese Society for Artificial Intelligence.



Hiroshi Ishiguro (Member, IEEE) received the D.Eng. degree from Osaka University, Osaka, Japan, in 1991.

In 1991, he started working as a Research Assistant of the Department of Electrical Engineering and Computer Science, Yamanashi University, Yamanashi, Japan. Then he moved to the Department of Systems Engineering, Osaka University, as a Research Assistant in 1992. In 1994, he was an Associate Professor, Department of Information Science, Kyoto University, Kyoto, Japan, and started research on distributed vision using omnidirectional cameras. From 1998 to 1999, he worked in the Department of Electrical and Computer Engineering, University of California, San Diego, as a Visiting Scholar. In 2000, he moved to the Department of Computer and Communication Sciences, University, Wakayama, Japan, as an Associate Professor and became a Professor in 2001. He is now a Professor of the Department of Adaptive Machine Systems, Osaka University, and a Group Leader at ATR Intelligent Robotics and Communication Laboratories, Kyoto. Since 1999, he has also been a Visiting Researcher at ATR Media Information Science Laboratories, Kyoto, and has developed the interactive humanoid robot, Robovie.



Michita Imai (Member, IEEE) received his Ph.D. degree in computer science from Keio University, Tokyo, Japan, in 2002.

In 1994, he joined NTT Human Interface Laboratories. He joined ATR Media Integration and Communications Research Laboratories in 1997. He is currently Assistant Professor, Faculty of Science and Technology, Keio University, and a Researcher at ATR Intelligence Robotics and Communications Laboratories, Kyoto.

Dr Imai's research interests include autonomous robots, human–robot interaction, speech dialogue systems, humanoids, and spontaneous behaviors. He is a Member of the Information and Communication Engineers Japan (IEICE-J), the Information Processing Society of Japan, the Japanese Cognitive Science Society, the Japanese Society for Artificial Intelligence, the Human Interface Society, and the Association of Computing Machinery.



Tetsuo Ono was born in Hokkaido, Japan. He received the Ph.D. degree in information science from Japan Advanced Institute of Science and Technology (JAIST), Ishikawa, Japan, in 1997.

He was a Researcher at ATR Media Integration and Communications Research Laboratories between 1997 and 2001. He is currently an Associate Professor of Media Architecture at the Future University Hakodate, Hokkaido, Japan, and also a Visiting Researcher at ATR Intelligent Robotics and Communication Laboratories,

Kyoto, Japan. His research interests include a computational mechanism of emotion, evolution of languages, human–robot communications, and interactive systems.

Dr. Ono is a Member of the Information Processing Society of Japan, the Japanese Cognitive Science Society, and the Japanese Society for Artificial Intelligence.