Auton Robot
DOI 10.1007/s10514-006-9007-6

A humanoid robot that pretends to listen to route guidance from a human

Takayuki Kanda · Masayuki Kamasima · Michita Imai · Tetsuo Ono · Daisuke Sakamoto · Hiroshi Ishiguro · Yuichiro Anzai

Received: 24 August 2005 / Revised: 7 September 2006 / Accepted: 12 September 2006

Abstract This paper reports the findings for a humanoid robot that expresses its listening attitude and understanding to humans by effectively using its body properties in a route guidance situation. A human teaches a route to the robot, and the developed robot behaves similar to a human listener by utilizing both temporal and spatial cooperative behaviors to demonstrate that it is indeed listening to its human counterpart. The robot's software consists of many communicative units and rules for selecting appropriate communicative units. A communicative unit realizes a particular cooperative behavior such as eye-contact and nodding, found through previous research in HRI. The rules for selecting communicative units were retrieved through our preliminary experiments with a WOZ method. An experiment was conducted to verify the effectiveness of the robot, with the results revealing that a robot displaying cooperative behavior received the highest subjective evaluation, which is rather similar to a human listener. A detailed analysis showed that this evaluation was mainly due to body movements as well as utterances. On the other hand, subjects' utterance to the robot was encouraged by the robot's utterances but not by its body movements.

Q1

T. Kanda (✉) · M. Kamasima · M. Imai · T. Ono · D. Sakamoto · H. Ishiguro
ATR

M. Kamasima · M. Imai · Y. Anzai
Keio University

T. Ono · D. Sakamoto
Future University, Hakodate

H. Ishiguro
Osaka University

Keywords Human-robot interaction · Embodied communication · Cooperative body movement · Humanoid robot · Communication robot

1 Introduction

1.1 The communication robots

Over the past several years, many humanoid robots have been developed, and they can typically make sophisticated human-like expressions with their head and arms (Hirai et al., 1998; Sakagami et al., 2002). We believe that humanoid robots will be suitable for our research on "communication robots" that behave as peer-partners to support daily human activities based on advanced interaction capabilities. The human-like bodies of humanoid robots enable humans to intuitively understand their gestures and cause people to unconsciously behave as if they were communicating with humans. Thus, as well as providing physical support, these robots will supply communication support such as route-guidance (Ono et al., 2001) and education (Kanda et al., 2004a).

Recent research into HCI (human-computer interaction) has highlighted the importance of robots as a new interface. Reeves and Nass researched the role of computers as new interface media in the manner of TV and radio, and they proved that humans act toward computer interfaces (even a simple text-based interface) as if they were communicating with other humans (Reeves and Nass, 1996). Cassell et al. showed that anthropomorphic expressions, such as those by arms and heads on embodied agents, are important for effective communication with humans (Cassell et al., 1999; Nakano et al., 2003). Kidd and Breazeal compared a robot and a computer-graphic agent and found that subjects felt the robot to be more informative and credible than the computer-graphic agent for communication concerning real-world objects (that is, for manipulating colored objects on a table) (Kidd and Breazeal, 2004).

Previous works in robotics have emphasized the merits of robots' embodiment. For example, they have shown the effective usage of body properties in communication, such as facial expression, eye-gaze, and gestures (Breazeal and Scassellati, 1999; Nakadai et al., 2001). Moreover, mutual body movements have been investigated. The joint-attention mechanism is one typical mutual body movement, whereby humans utilize their eye-gaze and pointing gestures to mutually synchronize their attention. Scassellati developed a robot as a testbed for a joint-attention mechanism (Scassellati, 2000). In that work, the robot followed people's gaze in order to share attention. Imai and his colleagues used a robot's arms as well as eyes to establish joint attention and verified its effectiveness (Imai et al., 2003).

1.2 Importance of cooperative body movements

Furthermore, recent research works reported the importance of cooperative body movements. Ono and his colleagues verified the importance of eye contact, arm gestures, and appropriate positional relationships (orientation of body direction) in a route guide robot (Ono et al., 2001). In this research, it was found that body movements are not only used for visually understanding what the teacher (the robot that taught the route) says but also for synchronizing communication. That is, the body movements of the robot teacher made human listeners move their bodies in a similar way as the teacher did, such as an imitation of a pointing gesture (Fig. 1). Like this example, it is important to adjust the teacher's body movement appropriately, which causes the cooperative body movements of listeners, such as the imitation of pointing, and makes the interaction natural. The importance of cooperative body movements was also found in interaction between humans and an autonomous interactive robot. Kanda and his colleagues found that people caused cooperative body movements, such as eye contact and synchronized body movements, when the people evaluated the robot positively (Kanda et al., 2003). These research works highlighted the importance of cooperative body movements when robots played the role of a speaker while a human was a listener in an interaction.

Fig. 1 Embodied cooperative behaviors in human-human communication

On the contrary, few papers have reported cooperative behavior when a robot plays the role of a listener and a human is the speaker. Watanabe and his colleagues found the importance of temporal cooperativeness, and have developed a robot that is capable of giving responses to a speaking human (Ogawa and Watanabe, 2001). However, only temporal cooperativeness was considered in that case and little previous research has focused on the spatial cooperativeness of body movements of a robot listener.

Cooperative body movements were also utilized for developing an intelligent mechanism for robots based on imitation and learning. For example, interactive systems observe human behaviors for the purpose of synthesizing behaviors (Jebara and Pentland, 1999). One imitation mechanism for a robot was developed comprising a motion capturing system and a neural network (Billard and Mataric, 2001). However, these research approaches focused on the intelligent mechanism for generating a motion, and they did not reveal its effects on human-robot interaction, such as how effective cooperative behaviors make interaction more natural.

1.3 A communication robot that expresses listening attitude with cooperative body movements

In a route guidance situation, there are two roles: a teacher (mostly talking to explain the route) and a listener (mostly listening), and since the roles of teacher and listener can be clearly separated, there are two research directions:

(1) To develop a robot that teaches a route to a human (Ono et al., 2001)
(2) To develop a robot that listens to the route guidance instructions given by a human (this paper)

We believe that both directions are important, and these will be finally merged into an ideal communication robot that performs natural communication like humans do in any interaction scenes. Since we have already developed a robot for the teacher role (Ono et al., 2001), we are going to focus on the second direction in this paper.

The situation where a robot teaches a route to a person is apparently important, since communication robots are expected to perform the role of conveying information to people. Here, however, we also focus on a route guidance situation where a person teaches a route to a robot. We believe that it is a realistic situation for a communication robot, thus the function of expressing listening attitude needs to be developed. There are two examples of this situation. First, there is the case where a person asks a robot about some operation related to a place. Here, we believe that the most intuitive way to operate is to use utterances and gestures as humans do to each other. Thus, a communication robot should have a function to give response to the person to express its listening attitude and understanding. The second case is a situation of route guidance. Even when a robot explains a route to a person, the explained person will sometimes repeat the route explanation back to the robot to ensure his/her understanding is correct, such as saying “I see. That is, go straight, turn right, and then arrive at the destination. Is this right?” This often happens in inter-human conversation: After a person (A) explains a route to the other person (B) unilaterally, the role of speaker and listener switches, and person B confirms the route to person A by explaining it in his/her own words.

Moreover, we can also expect this work to contribute to research on embodied communication where a robot performs cooperative behaviors in the role of a listener toward the speaking person. When a robot is in a speaker role, it is not necessary to adjust its behaviors to the human listener, since the speaker initiates utterances and gestures and it is the listener who performs cooperative behaviors toward the speaker; thus, it is a relatively difficult research issue to develop a robot that behaves cooperatively with a human speaker.

In this paper, we propose a mechanism for a communication robot that autonomously expresses its listening attitude and understanding to a speaker in the role of a listener in a route guidance situation. In other words, the robot pretends to listen to the speaker in conjunction with cooperative body movements. Since no speech-recognition function is used in this research, the robot does not linguistically understand what is said by humans. Concretely, our robot utilizes both body movements and utterances to give responses to a human speaker as a human does. It selects appropriate cooperative body movements from among 18 implemented behaviors such as eye contact and nodding, which are prepared in a bottom-up manner by referring to previous research works in robotics and cognitive science. The selection rules were implemented by retrieving knowledge from a human operator with a WOZ (Wizard of Oz) method. The evaluation experiment proves the effectiveness of the proposed method and identifies how the robot’s body movements and utterances affect subjective evaluation and behaviors of the robot. Through this research approach, we aim to identify an ideal mechanism for a communication robot with human-like body properties.

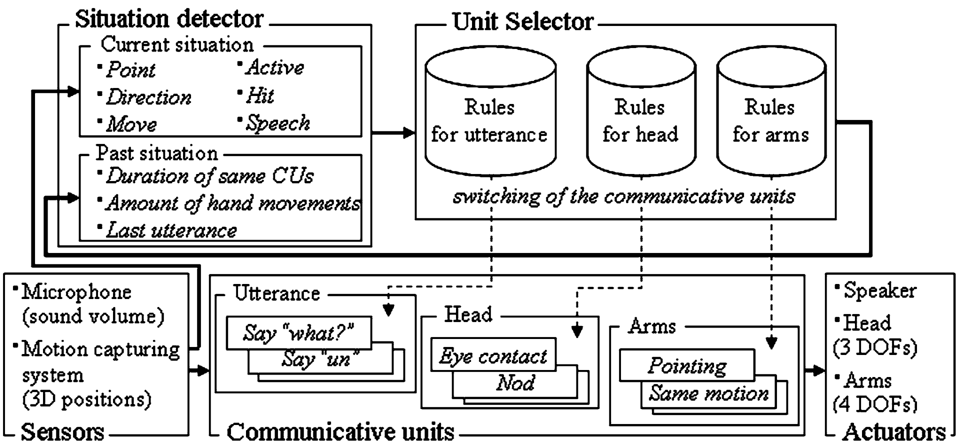
2 System configuration

We have developed a humanoid robot system that performs cooperative behaviors with a human in a route guidance situation, the purpose of which is to naturally communicate with humans. Concretely, when a human explains a route to the robot, it expresses cooperative body movements and utterances to express its listening attitude and understanding, or to *pretend* to listen, to the explanation. Figure 2 shows an overview of the developed system. The following subsections describe the design policy, details about the system’s components, and preliminary experiments to set up the system’s rules and parameters.

2.1 Design policy

The system is designed to realize an ideal listener robot that expresses responsive behaviors to a speaker as if it were a human listener in an inter-human conversation. The essential components of the system consist of both cooperative body movements, which have been identified to be important such as eye contact and imitation of pointing, and simple utterances to give responses. We named the components “communicative units,” and developed the “*pretending* listening behaviors” by controlling the use of the communicative units along with the current state (posture and whether or not speaking) of a speaking person and the past state (posture and utterance) of the robot.

Fig. 2 System configuration



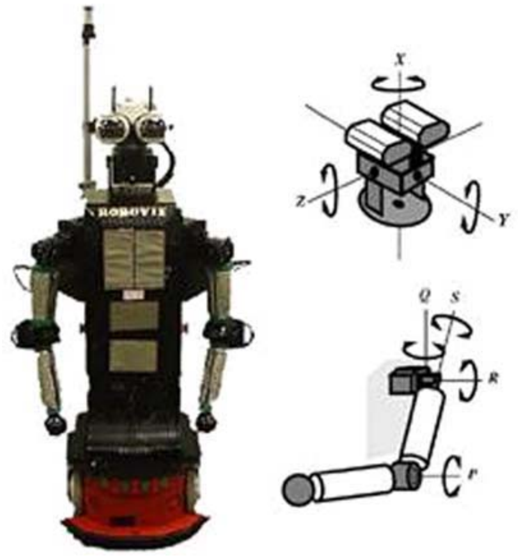


Fig. 3 Humanoid robot "Robovie"

Through this development, our purpose is to prove the validity of our framework for utilizing cooperative behaviors for listener behavior. Thus, we focused on the minimum essential components for body movements and utterances and did not include redundant behaviors or subtle expressions, such as facial emotions and slight movements. For example, the utterances "un" "un un" and "a ha" would be redundant. We only included important body movements mainly reported in previous research on HRI (human robot interaction); as a result, we ignored less important body movements. Of course, humans are doing more various behaviors than what the developed robot does; so if our framework is proved to be valid, we believe that we can further improve the performance of the system by adding other behaviors.

Our hypothesis behind the implementation was that we can perform appropriate body movement and vocal backchannel without the semantics from speakers' utterances. Of course, it will be difficult in general; but, during the route-guidance, the listener can also get information through the teaching person's body movements. The results of the WOZ experiment seem to indicate that this is a valid hypothesis.

2.2 Hardware

Figure 3 shows the humanoid robot "Robovie" (Kanda et al., 2004b). It is capable of human-like expression and recognizes individuals by using various actuators and sensors. Its body possesses highly articulated arms, eyes, and a head, which were designed to produce sufficient gestures for communicating effectively with humans. The sensory equipment includes auditory, tactile, ultrasonic, and vision sensors, which allow the robot to behave autonomously and to interact with humans. All processing and control systems, such as the computer and motor control hardware, are located inside the robot's body. The height is 1.2 m and its radius is 0.5 m.

We adopted a microphone and a motion capturing system as the system's sensors. The microphone is attached to the robot, which acquires the utterance volume of a human. The motion capturing system acquires three-dimensional numerical data on the human body movements. It consists of 12 sets of infrared cameras with an infrared irradiation function and markers that reflect infrared rays. The motion capturing system calculates the three-dimensional position of each marker based on the two-dimensional positions on all of the cameras' pictures. The system's time resolution is 60 Hz and spatial resolution is about 1 mm in the experimental environment. The attaching position of each marker is shown in Fig. 4. There is an approximately 50 milliseconds delay to calculate the three-dimentional position of markers with these settings.

2.3 "Communicative units" for cooperative behaviors

The effectiveness of temporal-cooperative behaviors was already verified in previous research. Watanabe et al. found that nodding behavior of a robot makes human-robot communication as natural as human-human communication

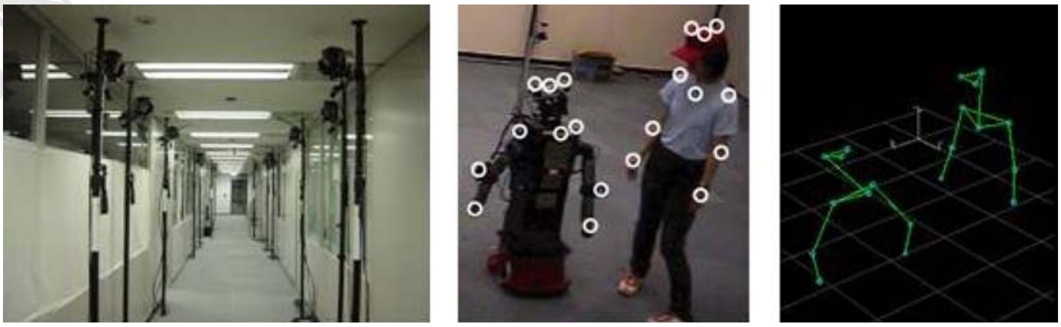


Fig. 4 The motion capturing system (left), attached markers (center), and obtained 3-D numerical position data of body movement (right)

(Ogawa and Watanabe, 2001). We define such nodding as *temporal-cooperative* behavior, because it expresses a reacting attitude to the partner's action with appropriate timing. Similarly, backchannel utterances such as "un" and "un un" are also temporal-cooperative behaviors.

Meanwhile, the occurrence of *spatial-cooperative* behaviors was found in a situation where a robot taught a route to a human (Ono et al., 2001). For example, in their research, eight out of ten subjects performed imitation of pointing with arms. We can find similar spatial-cooperative behaviors in joint-attention mechanism, where a listener looks or points in the direction that a speaker is looking or pointing at to share attention about objects or directions (Moore and Dunham, 1995).

Here, we adopt these temporal or spatial cooperative behaviors found in previous research. There are certain components to realize cooperative behaviors, which are called *communicative units*. By continuously controlling the use of communicative units, the developed system controls each of the head, right arm, left arm and utterance of the humanoid robot to express its listening attitude. We have already proposed the notion of communicative units for an autonomous interactive robot (Kanda et al., 2004b), where the communicative units realize basic motion for general communication, such as eye contact and pointing. We believe that future communication robots will be equipped with a basic library of body movements so that developers can easily configure high-level communication by combining them. Through this research, we would like to also establish a fundamental set of communicative units and the method to appropriately use them; we believe that it will have great merits on various future communication robots.

Table 1 shows all implemented communicative units. Only one communicative unit can be active within a part of body (right arm, left arm, head, and utterance), and each communicative unit for a part can run in parallel; thus, multiple communicative units can be active in the robot. In this research, each communicative unit refers to an output from a motion capturing system to obtain human positions. Regarding the communicative units related to the head and arms, they calculate the destination angle of each joint of the robot's head and arms based on numerically obtained data of human body movements. For instance, the calculations in *Hec* (eye contact) and *Rsr* (synchronized arm movement) are described as follows (Henceforth, each communicative unit is described with its name identifier, such as *Rsr*):

Hec: This calculates both the robot's head direction vector and the human's head direction vector and then calculates the desirable angle of the robot's head so that these two vectors exactly indicate the opposite direction on a certain line.

Rsr: This calculates the angle of a human's right shoulder and elbow and then reconfigures these angles into the angle of the robot's right arm so that the robot seems to show the same motion as the human does. (The same angles do not seem to show the same motion. Thus, we need to adjust the angles between the robot and humans with a simple look-up table prepared in advance.)

Some communicative units such as nodding (*Hnd*) do not refer to the input from the motion capturing system. For example, *Hnd* changes the head's orientation from the current one to a relatively lower one for a while.

In addition, we prepared a parameter "response-delay time (*d* sec)" to make communicative units more natural. Because the robot can react faster than what humans do due to the fast calculation of the motion capturing system, we have observed unnaturalness in the robot's cooperative behaviors when the delay *d* was not present. That is, it was rather reflecting human motion rather than reacting to human action. This response-delay time *d* was simply realized by letting the robot's system refer to the *d* sec older data obtained from the motion capturing system. Our system implements

Table 1 Implemented communicative units

Right arm	Left arm
Rsr: Same motion as human's right hand	*Lsl*: Same motion as human's left hand
Rsl: Same motion as human's left hand	*Lsr*: Same motion as human's right hand
Rpr: Points in the direction indicated with right hand	*Lpr*: Points in the direction indicated with right hand
Rpl: Points in the direction indicated with left hand	*Lpl*: Points in the direction indicated with left hand
Rno: Do nothing	*Lno*: Do nothing
Head	Utterance
Hec: Eye contact	*Seh*: Says "*eh?* (what?)"
Hrp: Turn the head in the direction indicated with right hand so that it seems to look in that direction	*Sun*: Says "*un*."
Hlp: Turn the head in the direction indicated with left hand so that it seems to look in that direction	*Suu*: Says "*un un*."
Hnd: Nod	*Ssd*: Says "*sorede* (so what?)."

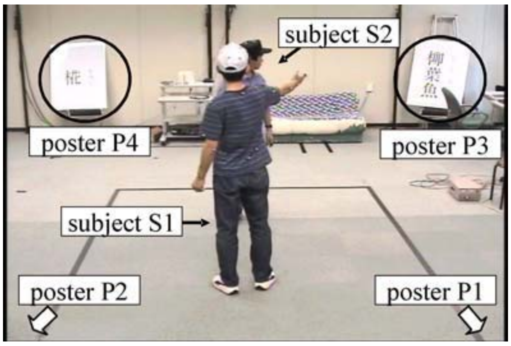


Fig. 5 Scene of the experiment for measuring humans' response delay

response-delay time *d* in the units *Rsr*, *Rsl*, *Rpr*, *Rpl*, *Lsl*, *Lsr*, *Lpr*, *Lpl*, *Hec*, *Hrp*, and *Hlp*.

2.4 Preliminary experiment for measuring response-delay time

We conducted a preliminary experiment to choose the appropriate response-delay time *d* sec (explained in the previous subsection) where we measured the delay time of humans during a pointing conversation.

Method

We employed 25 pairs of university students (23 men, 27 women) for the preliminary experiment. They were asked to participate in "experiments to talk with a humanoid robot." employed them in first-come-first-employed manner. There were no special request for subjects' capability except for being fluent in Japanese and no specific selection was conducted to choose the subjects.

We placed four posters, P1, P2, P3, and P4, in each corner of a room measuring 8 m × 15 m. The setup of the experiment is shown in Fig. 5. The posters described difficult Kanji characters (since each Kanji character is associated with a semantics and has a multiple way of readings, even Japanese adults usually do not know the readings of very difficult Kanji). Two subjects S1 and S2 were face-to-face in the center of the room. S1 pointed at a poster and spoke the reading of the Kanji to teach the reading to S2. S1 repeated this for posters P1 (right rear), P2 (left rear), P3 (right front), and P4 (left front). The task (teaching the reading of the Kanji) was a pseudo task so that the subjects would not be nervous about their body movements. The true purpose was to measure the delay of the movements from the start of S1's to that of S2's, which were measured by using a motion capturing system.

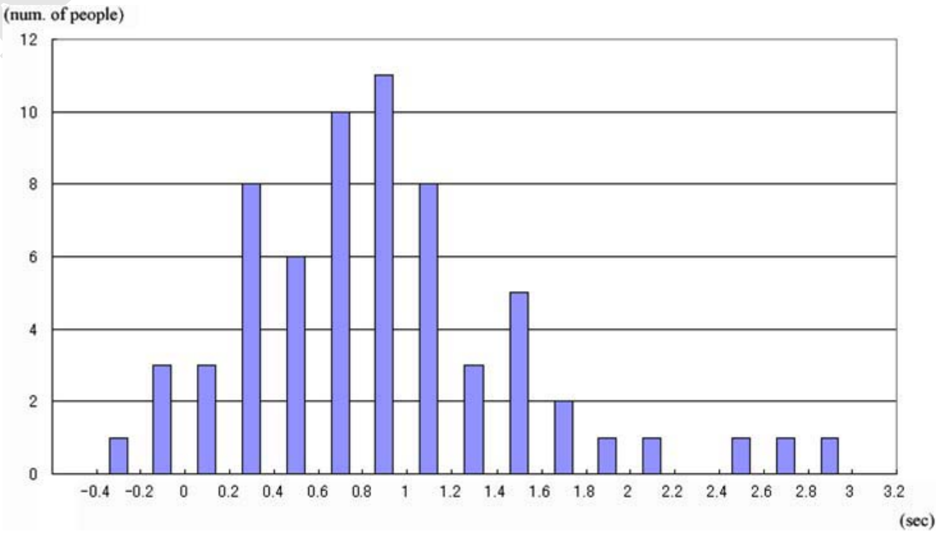
Measurement of delay time

By using the numerically obtained body movement data, we determined the start time of S1's movement (*t1*) to be the earlier of the following two movements: the time when S1 started to move his/her arm (the start of pointing) and the time when S1 started to move his/her head (the start of eye gaze). Similarly, the start time of S2 (*t2*) was defined as the time when S2 started to move his/her head (the start of the looking motion). The response-delay time of the reaction is retrieved as *t2*–*t1*.

Result

Figure 6 displays the response-delay times for the four pointing behaviors for all subjects (data from 17 pairs was used while that of 8 pairs was omitted due to data collection errors with the motion capturing system). The average delay time was 0.89 s (standard deviation 0.63). We utilized this parameter in the developed system so that the response-delay time *d* was 0.89 s.

Fig. 6 Results for the humans' response delay



2.5 Preliminary experiment with WOZ settings for retrieving rules to control the use of communicative units

Another preliminary experiment was conducted to retrieve the control rules for communicative units. Although we have already reported the findings from this experiment, which verified the effectiveness of communicative units (Sakamoto et al., 2005), here we briefly explain them because they are closely related to the implementation of control rules.

Settings

The subjects for the experiments were 50 university students (23 male, 27 female) who also participated in the other preliminary experiment described in the previous subsection. After learning a route by walking, they were asked to teach it to the robot. For each teaching of a route, we prepared two experiment conditions:

Rc condition: the robot expresses its listening attitudes with communicative units, chosen by human operators to be appropriate to each situation.
Rs condition: the robot stayed stationary.

In addition, the subjects were paired and one subject in each pair explained the route to the other (*H condition*).

Here, human operators chose communicative units (denoted in Table 1) preferable for the current situation as shown in Fig. 7. Two specific persons who were well trained to operate the system (one of whom is the co-author of the paper) always served as the operators. There were markers of the motion capturing system attached to both subjects and the robot (Fig. 4). The human operators continuously assigned which communicative units should be used. Those communicative units were then executed by the robot based on output from the motion capturing system. As a result, subjects reported better impressions for the Rc condition than the Rs condition, which seems to indicate a positive perspective of a robot that exhibits those behaviors, as reported in Sakamoto et al. (2005).

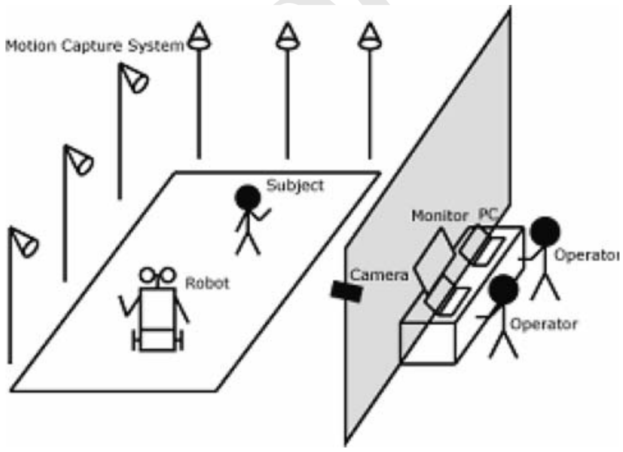


Fig. 7 Settings for WOZ experiment

Analysis of operator's selection

In the experiment, two operators controlled the robot's behavior. There was no script prepared in advance for the operator, because we were not sure what behaviors would be appropriate. We asked the operators to establish a consistent manner of operation so that the behaviors would be consistent between different subjects. The operators used some test subjects within the laboratory and tried to make the robot behaviors appropriate from their subjective view.

They only controlled the selection of the communicative units, and did not directly control head orientation or arm gestures. Thus, the system controlled spatial cooperative behaviors of the robot, while the human operators decided the communicative units to be executed with appropriate timing. We recorded the operation of choosing communicative units along with video of the experiments, output from the motion capturing system, and utterance information obtained from the microphones.

We believe that this is one of the important points of the research. The operators' decisions were recorded at the symbol level, but not at the raw sensory-motor level. If we were to allow the operators to directly control the motors of the robot, their operation (such as, moving the robot's head in a horizontal direction) might have multiple meanings (such as, for nodding, facing its head in the indicated direction, just making its pose as default, etc.); thus, the mapping, required for later implementation, between sensory input and robot's behavior would be more complicated, due to such complex decision-making behind the motor control of the operators. That is why we implemented sensory-motor mapping (communicative units) first, and tried to retrieve operators' behavior through symbolic operations.

After the experiment, we analyzed the operation records in order to retrieve the if-then rules for selecting communicative units. We assigned the reason why the operator chose each of the communicative units that appeared in the operation records (such as, "because the robot's left hand was so close to the subject that it would have get contact with him/her, its right hand was used", or "there were no specific action needed for its head so the eye-contact module was chosen"). Then, we added if-then rules that could be implemented with its sensors until most of the operations could be reproduced by the rules. As a result, the following rules were retrieved.

- "Eye contact" and "the same arm movement" are usually selected.
- When a subject points in a certain direction by lifting one of his/her hands, the robot points in the same direction

and turns its head in the pointed direction so that the robot appears to look in that direction.

- If the robot did not conduct eye contact for a while, perform eye contact
- While the subject is moving his/her hand, perform an imitating gesture with the same-side hand
- If the subject is so close that the robot's hand might get contact with him/her, use the other hand instead.
- When the robot tries to perform an imitating gesture and the subject is facing it, perform a mirrored imitating gesture instead.
- Backchannel feedback is given in response to the subject's explanation (after a certain blank period following humans' utterances).

(The experiments were conducted with Japanese subjects. There is a cultural characteristic in giving response behaviors: Maynard reported that the backchannel (giving a response) frequency of Japanese is higher than that of Americans (with brief utterances: 165 times for Japanese and 35 times for Americans among each 36-min data set; with head movement: 104 times for Japanese and 5 times for Americans). Even though the role of the backchannel is the same in both languages (Maynard, 1986).

2.6 Situation detector and unit selector

We analyzed the human operators' decisions to retrieve the rules for selecting communicative units, as described in the previous section, and implemented them into the system. The system consists of two parts: a situation detector and a unit selector.

Situation detector

The situation detector detects 6 current characteristics and 5 past characteristics of the situation. The current characteristics are about the subject's posture with respect to the robot posture (such as *Hit* and *Direction* characteristics, described below) and whether or not the subject is speaking. The situation detector identifies them by referring to the input from the motion capturing system and a microphone, and also remembering short-term past situations. These are the six characteristics:

Point: Whether he/she is using the right (left) hand for pointing?

Direction: In which direction is he/she pointing, to the right or to the left side of the robot?

Move: Is the right (left) hand moving? (Does the speed of the hand exceed a certain threshold?)

Active: Is the right (left) hand used for guiding gestures (pointing and the movement between pointing)?

Hit: Is he/she so close to the robot that it might hit him/her with its right (left) hand if the robot moves it?

Speech: Is he/she speaking?

The remaining five characteristics are metrics based on the robot's most recent actions:

- How long has the same communicative unit with the head been in progress?
- How long has *Hrp* or *Hlp* (facing its head in an indicated direction) been in progress?
- How long has *Rpr, Rpl, Lpr, or Lpl* (pointing in a direction) been in progress?
- How much did it move its hand during a past certain number of seconds?
- What did it say in its last utterance?

Unit selector

The unit selector consists of a set of rules for selecting appropriate communicative units for each of the head and both arms. Figures 8 and 9 describe all implemented rules related to the arms and the head and utterances. These rules are based on the analysis of the operator, described in the previous section. The rules are implemented as a combination of if-then rules referring to the six current situations and five past situations detected by the situation detector.

For example in Fig. 8, if the last behavior module is not *Rpr*, *Rpl*, *Lpr*, or *Lpl*, a human is pointing with the right hand (*Point* = *right hand*), the human is using the right hand for route guidance (*Active* = *right hand*), the human is not in the region where either of the robot's hands might hit him or her (*Hit* = *nothing*), and the human is pointing to the robot's left side (*Direction* = *left*), then *Lpr* is selected.

3 Experiment

We conducted an experiment to verify the significance of the developed system. The hypothesis for the experiment was "if a robot performs embodied cooperative behaviors corresponding to the interacting human based on the developed system, then the human will perceive the communication with the robot during the route guidance is smooth."

3.1 Method

A human teacher (denoted as *Teacher*) taught a route to a destination to the developed robot or a human learner (denoted as *Learner*). The following presents the details of the experimental procedure.

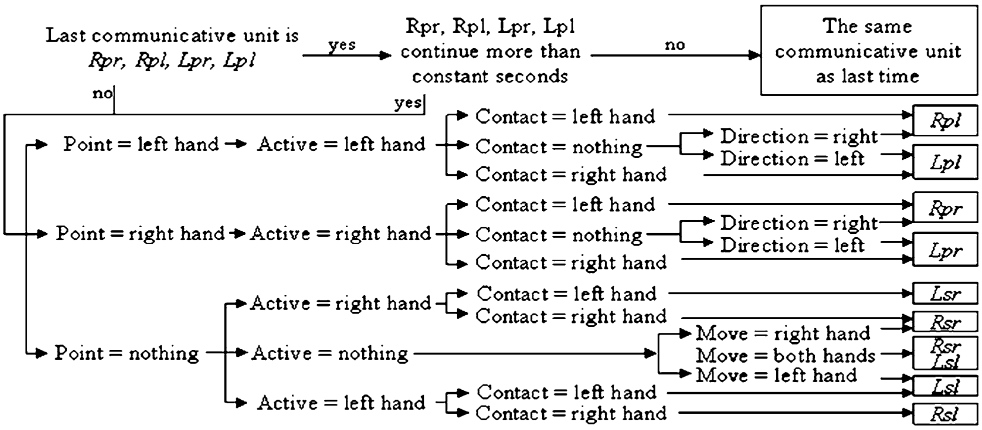
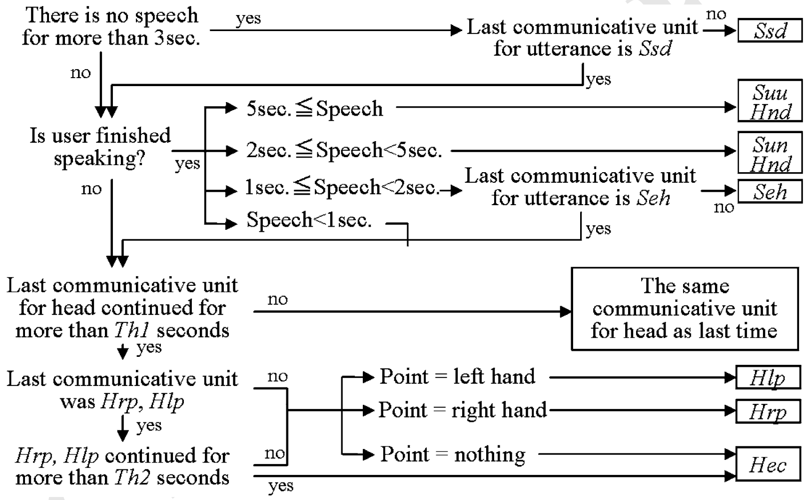


Fig. 8 Illustration of the rules for selecting communicative units for arms

Fig. 9 Illustration of the rules for selecting communicative units for head and utterances



Subjects

We employed 81 university students as subjects in the experiment (36 men, 45 women). They were asked to participate in "experiments to talk with a humanoid robot." We employed them on a first-come-first-employed basis. There were no special requests for subjects' capabilities except for being fluent in Japanese, and no specific selection was conducted to choose the subjects. They had never visited this environment before, so they did not know the route that they would teach or be taught. None of them had participated in the previous experiment described in Section 2.

Conditions

We investigated the effect of the *Learner*'s embodied cooperative behaviors on the *Teacher*. We set five *Learner* conditions as follows:

Human condition (H condition)
The *Teacher* teaches a human the route.

Robot cooperative condition (Rc condition)
The *Teacher* teaches the robot that performs embodied cooperative behaviors.

Robot body move condition (Rb condition)
The *Teacher* teaches the robot that performs embodied cooperative behaviors without utterances (only body movements).

Robot voice condition (Rv condition)
The *Teacher* teaches the robot that performs embodied cooperative behaviors without body movements (only utterances).

Robot static condition (Rs condition)
The *Teacher* teaches the robot that remains stationary (without body movements and utterances).

(We chose to keep the robot stationary for the control condition because it more naturally falls within human social norms than other reactions, such as random movement, would. It would not be unnatural, for example, for an unfriendly person to remain nearly stationary while listening to route guidance.)

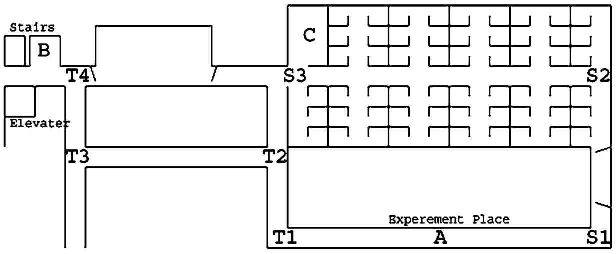


Fig. 10 Environment for the route-guidance experiment

We defined Robot condition (R condition) as the set of Rc, Rb, Rv, and Rs conditions.

Environment

Figure 10 shows the experimental environment. The *Teacher* told the route to the *Learner* at A, and the destination that the *Teacher* taught is one of two lobbies (B or C).

Procedure

Since a human *Teacher* taught a route to a human *Listener* in the human condition, we needed to pair two subjects and operate the paired subjects simultaneously. Each subject participated in both the H condition and the R condition. As for the R condition, one from among the Rc condition, Rb condition, Rv condition, and Rs condition was chosen randomly. In the H condition, each subject behaved as both *Teacher* and *Learner*. In addition, an experimenter guided the *Teacher* along the route that he/she would teach to the *Learner* before the experiment. The order of the two experiments (R and H conditions) was counter-balanced. (For half of all subjects, we conducted the experiments in the H-R order, while the R-H order was used for the rest.) The route guidance destination (lobby B or C) was randomly assigned within paired subjects so that each of the subjects was taught the route he/she did not know. (For example, supposing there are a paired subject *X* and *Y*, subject *X* teaches a route to lobby B and subject *Y* teaches a route to lobby C).

First, the *Teacher* is taught a route to the lobby (B or C); he/she will guide by actually walking to the destination. After that, the *Teacher* is given the instruction at a point close to point A that: "There is a person (*Learner*) who gets lost. He/she will ask you the route to the lobby, so please explain the route. At first, please point to the first corner, and start with "from this corner" to teach the route." The *Learner* is given an instruction to wait at point A and ask the *Teacher* for the route when the *Teacher* comes. The experiment starts when the *Teacher* arrives at point A, where the *Learner* is waiting. To control the R and H conditions, we instructed the *Learner* not to ask for the route repeatedly. The experiment was finished when the *Teacher* finished the route guidance, and neither the robot nor the human *Listener* was designed to follow the route after the guidance. Instead, the experimenter came and picked up the *Teacher* in order to let the *Teacher* answer the questionnaire.

Evaluation

We administered a questionnaire to obtain subjective evaluations of when the subjects behaved as *Teacher* and also analyzed their behavior toward the *Learner*. In the questionnaire, we investigated the influence of robot's behaviors that affect communication. Specifically, we investigated aspects of conveying the information, reliable communication, and sympathetic interaction, where the last two aspects are the ones related to human-like natural communication. Concretely, the following six questions were used in the questionnaire. The subjects answered each question on a 1-to-7 scale, where 1 stands for the lowest evaluation and 7 stands for the highest.

- Aspects of conveying information

  Q. 1 Time to recall the route
  Q. 2 Easiness of teaching the route to the partner

- Aspects of reliable communication

  Q. 3 The partner's listening to the guidance
  Q. 4 The partner's understanding of the guidance

- Aspects of sympathetic interaction

  Q. 5 Your feelings of sharing information with the partner
  Q. 6 Your empathy with the partner.

Regarding the *Teacher*'s behavior, the following factors were recorded and analyzed.

- Total duration of utterance
- Total amount of arm gesture (sum of both hands' movements per second)

3.2 Results

First, we compared the subjective impressions for the H condition, the Rc condition (a robot with the cooperative embodied behavior), and the Rs condition (a static robot) to verify the significance of the developed system.

Significance of the developed system

Table 2 shows the average, the standard deviation, and the result of analysis of variance (ANOVA) among the H, Rs, and Rc conditions of the six items on the questionnaire. In the table, standard deviation is given in parentheses after the average value. The comparison is also illustrated in Fig. 11.

Table 2 Comparison among the H, Rc, and Rs conditions

	Q. 1(Recallability)	Q. 2 (Easiness)	Q. 3 (Listening)	Q. 4 (Understanding)	Q. 5 (Sharedness)	Q. 6 (Empathy)
H condition (40 subjects)	5.38 (1.48)	4.68 (1.44)	6.28 (1.06)	5.18 (1.39)	5.10 (1.32)	4.78 (1.23)
Rc condition (20 subjects)	4.75 (1.80)	3.75 (1.71)	5.50 (1.60)	5.05 (1.61)	4.40 (1.57)	3.65 (1.42)
Rs condition (20 subjects)	4.45 (1.67)	3.25 (1.25)	3.95 (1.50)	4.05 (1.28)	2.80 (1.24)	2.35 (1.27)
Result of ANOVA ($F(2,77)$)	$p = .090$ (+) $F = 2.48$	$p < .01$ (**) $F = 6.97$	$p < .01$ (**) $F = 21.36$	$p = .015$ (*) $F = 4.41$	$p < .01$ (**) $F = 18.93$	$p < .01$ (**) $F = 24.02$
Multiple comparison	(H > Rs) $p < .05$	H > Rc, H > Rs $p < .05$	H > Rc, Rc > Rs $p < .05$	H > Rs, Rc > Rs $p < .05$	H > Rs, Rc > Rs $p < .05$	H > Rc, Rc > Rs $p < .05$

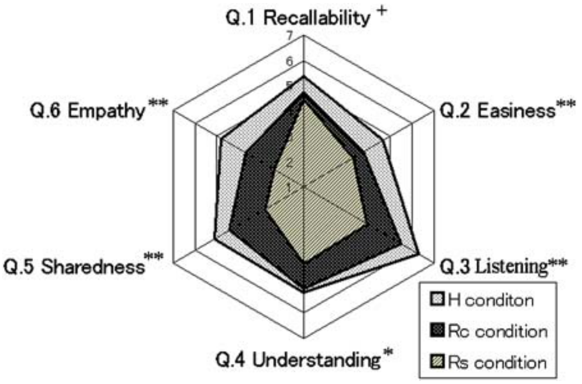


Fig. 11 Comparison of subjective evaluation between the Human (H) condition, the Robot cooperative (Rc) condition, and the Robot static (Rs) condition

The number of subjects was 40 in the H condition, 20 in the Rc condition, and 20 in the Rs condition.

The ANOVA (analysis of variance) result revealed significant differences in Q. 2, Q. 3, Q. 4, Q. 5, and Q. 6, and an almost significant difference in Q. 1. For each of the significant items, an LSD (least significance difference) method provided a multiple comparison among the H, Rs, and Rc conditions. As a result, there was a significant difference for the Rc condition > Rs condition in Q. 3 (listening), Q. 4 (understanding), Q. 5 (sharedness), and Q. 6 (empathy). These results proved that a subjective evaluation of the *Teacher* for the robot with embodied cooperative behaviors (Rc) is higher among the aspects of reliability and sympathy compared with the robot without body movements and voice (Rs). We believe that this proves the significance of the developed system.

Meanwhile, there was no significant difference between the Rc and Rs conditions among the aspects of conveying information; Q. 1 (recallability) and Q. 2 (easiness). Thus, the developed system had no effect on the aspects of conveying information. Moreover, the subjective evaluation for Rc was lower than the H condition in Q. 2 (Easiness) Q. 3 (Listening) and Q. 6 (Empathy), which suggests that the realized natural communication by the developed system is still far from that of inter-human communication; therefore, there are some things we can improve in the system for more naturalness.

Analysis of the effect of robot's body movements and utterances

We performed a detailed analysis on the robot's body movements and utterances by comparing the Rc, Rb, Rv, and Rs conditions. Table 3 shows the average and standard deviation of the six questionnaire items. It also describes the results of two-way factorial ANOVA among the conditions, where the two factors are "body movements" and "voice." The Rc condition has both factors, but the Rb condition has only the factor of body movements, the Rv condition has only the factor of voice, and the Rs condition has neither factor. The number of subjects was 20 in the Rc condition, 21 in the Rb condition, 20 in the Rv condition, and 20 in the Rs condition.

The two-way factorial ANOVA revealed that there were significant simple main effects for the body movement factor in Q. 3, Q. 5, and Q. 6, and an almost significant effect in Q. 4. For the voice factor, there was a significant simple main effect in Q. 5. Furthermore, there were significant statistical interactions between the body movement factor and the voice factor in Q. 3 and Q. 5. These results indicate that both the body movement factor and utterance factor affected on the reliability (Q. 3, 4) and sympathy (Q. 5, 6) aspects (since there are simple main effects of both factors or the interaction), and the body movement factor was relatively more dominant than the utterance factor because some of the questionnaire items were only affected by the body move-

Table 3 Comparison of the effect of the body movement factor and the utterance factor

	Q. 1 (Recallability)	Q. 2 (Easiness)	Q. 3 (Listening)	Q. 4 (Understanding)	Q. 5 (Sharedness)	Q. 6 (Empathy)
Rc condition (20 subjects)	4.75 (1.80)	3.75 (1.71)	5.50 (1.60)	5.05 (1.61)	4.40 (1.57)	3.65 (1.42)
Rb condition (21 subjects)	4.43 (1.80)	3.67 (1.74)	5.76 (0.89)	4.57 (1.54)	4.38 (1.50)	3.76 (1.51)
Rv condition (20 subjects)	5.05 (1.54)	3.35 (1.42)	5.15 (1.27)	4.45 (1.32)	4.40 (1.23)	3.00 (1.26)
Rs condition (20 subjects)	4.45 (1.67)	3.25 (1.25)	3.95 (1.50)	4.05 (1.28)	2.80 (1.24)	2.35 (1.27)
Factor of body movements ($F(1,77)$)	$p = .681$ (n.s.) $F = 0.17$	$p = .239$ (n.s.) $F = 1.41$	$p < .01$ ($**$) $F = 13.76$	$p = .084$ ($+$) $F = 3.06$	$p = .013$ ($*$) $F = 6.50$	$p < .01$ ($**$) $F = 11.43$
Factor of utterance ($F(1,77)$)	$p = .229$ (n.s.) $F = 1.47$	$p = .792$ (n.s.) $F = 0.07$	$p = .112$ (n.s.) $F = 2.59$	$p = .174$ (n.s.) $F = 1.88$	$p = .011$ ($*$) $F = 6.82$	$p = .383$ (n.s.) $F = 0.77$
Interaction ($F(1,77)$)	$p = .719$ (n.s.) $F = 0.13$	$p = 1.00$ (n.s.) $F = 0.00$	$p = .014$ ($*$) $F = 6.29$	$p = .921$ (n.s.) $F = 0.01$	$p = .013$ ($*$) $F = 6.50$	$p = .215$ (n.s.) $F = 1.56$

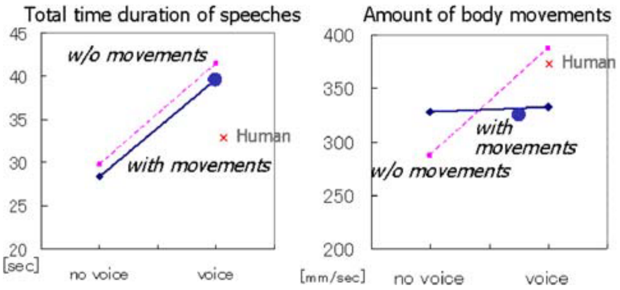


Fig. 12 Illustration of comparison of subjects' behavior toward the robot

ment factor. Regarding the effects for conveying information aspects, there was no significant effect caused by either the body movement factor or the voice factor.

Analysis of the effect of subjects' behavior toward Listener

Figure 12 shows the result of the analysis on the *Teacher*'s behavior toward the *Learner* in the Rc, Rb, Rv, and Rs conditions. There were a few subjects' data excluded from the analysis due to the failure of recoding of the motion capturing system. We analyzed the total duration of utterance and the total amount of arm gesture.

The left figure in Fig. 12 shows a comparison of the total duration of utterance. The two-way factorial ANOVA proved that only the utterance factor increased the utterance of the *Teacher* (utterance factor: $p < .01$, body movement factor: $p = .577$) The right figure refers to a comparison of the body movements. The two-way factorial ANOVA showed no significant difference, though it might be affected by the utterance factor (utterance factor: $p = .131$, body movement factor: $p = .846$). Thus, the utterance of the *Listener* robot had an effect on promoting a human *Teacher* in a route guidance situation, but its body movement did not affect the *Teacher* behavior. That is, the listener's vocal response promoted the speech of the speaker, which fits with a previous report in psychology on inter-human communication of Japanese (Tsukahara et al., 1997).

4 Discussions

Summary of the result

The experimental result demonstrated the significance of the developed robot system that reacts to the *Teacher*'s route guidance with both spatial and temporal cooperative behaviors. In addition to the subjective impressions, subjects provided free-form comments after the experiment, such as "with the robot's arm movements, nodding, and giving vocal responses, I could recognize that it comprehended what I was saying." We believe that the significance of the robot system with cooperative behaviors is proved. We focused on the "pretending listening" in this research, because our focus was on the embodied communication between a robot and people. This fundamental result suggests that, if we add a recognition function at the language level into the robot, we will be able to develop a robot that "understands and show its understanding" to people speaking to it.

Concerning the comparison of the body movement and utterance factors, both factors affect how the robot exhibits its listening behavior to the *Teacher*. Particularly, for sharedness (Q. 5), it seems that each factor sufficiently affected the subjective evaluation to make the effect of their mixture seem little bigger than that of each of them individually. We believe that the subjects received adequate signals of sharing the information from the robot merely by voice or body movements. Regarding listening (Q. 3) and empathy (Q. 6), only the body movement factor affected the subjective impression. To summarize, we believe that both factors affected the impression, and the body movement factor was more dominant than the utterance factor on the impressions.

On the contrary, the robot's utterance promoted the *Teacher*'s utterances, but its body movements did not have such an effect. This finding suggests that a robot can elicit a more elaborative explanation from a speaking person by reacting to the utterances, which may have a merit in speech recognition by the robot. That is, both body movements and utterances are important reaction for the robot to give better impression to and retrieve enough information from a speaking person.

This result matches the findings in HCI. Whittaker and O' Conaill analyzed inter-human communication through a video-conferencing system and found that the task achievement is mainly through vocal channels and that emotional information is mainly conveyed through visual channels (Whittaker and O'Conaill, 1997). In our research, since the task was closely related to the three-dimensional real world, the body movement factor seemed to affect also the task-achievement side, such as the *Teacher's* impression of listening to the *Listener*, as well as the emotional aspect of sympathy. This implies that the effect of visual channels, such as the body movements of robots, has higher power in a real-world task than the ones previous HCI have treated, such as on-screen and virtual world communication.

Effect of embodied cooperative behavior on aspects for conveying information

Ono et al. reported that gestures from a robot causes cooperative body movements in a human listener, such as pointing in the same direction as the robot, which promotes understanding by the listener about guidance along a route (Ono et al., 2001). One hypothesis we intended to prove was that the listener's cooperative body movements might promote the *Teacher's* gestures in teaching a route and support the teacher in recalling information about the route.

It seems, however, that the comparison of the subjective impression on the aspects for conveying information (Q. 1, 2) did not show a significant difference. Thus, even if there had been any effect on the aspects, it would have been smaller than the effects on other aspects. Moreover, the H condition received a better impression for the aspect than the Rs condition. This indicates a disadvantage in having a static robot compared to the human listener. Also, we found a significant difference in the H condition $>$ Rc condition. These results seem to suggest that the robot is not yet as good a listener as a human, probably due to the robot's appearance and lack of social expectation. For example, subjects reported on the difference of their behavior to the robot with the one to humans, such as "I spoke to the robot as if I were talking to a child," "I used simple landmarks when I explained directions to the robot," "I talked slowly and loudly to the robot," "I explained the route in detail to the robot," and "I did not give detailed explanations to the robot."

Generality of findings and Limitations

Since this "pretending listening" behavior does not depend on the appearance of Robovie, which has a less sophisticated design than other humanoid robots such as Asimo (Sakagami et al., 2002), we believe that the developed system and the experimental results are applicable for other humanoid robots that have a similarly simple or better appearance.

The experimental result showed that a robot with cooperative behavior affected for natural communication with humans to some degree, but not as much as inter-human communication. Our implementation includes fundamental cooperative behaviors with large movements, but it is apparently not perfect. We believe that its performance depends on our implementation yet. On the contrary, since some of the human *Listeners* in the experiment did not seem to be such good listeners, such as their just listening without exhibiting responses to the *Teacher*. Thus, the ideal robot might be able to realize natural communication as average humans do if we could implement further body movements and utterances, or add other hardware devices for subtle expressions such as facial expressions or degrees of freedom to the waist (Miyashita et al., 2004).

The findings also depend on the task. For example, we can expect that effects for the body movements might be stronger if a task requires significantly more spatial precision.

This research was conducted with the global perception of a motion capturing system, which could potentially cause a negative effect to the naturalness of the interaction of the robot. For example, the pointing behavior of an instructor is biased based on whether or not the listener can perceive the object of attention (Trafton et al., 2005). This type of information is difficult to account for using the global perception of a motion capture system. However, since the robot's reactions were limited to simple ones, such as nodding and synchronized arm movements (when it is facing its head in the direction, the speaker's motion is within the possible sight of the robot's eye) in the route guidance situation, we believe that the global perception did not cause a negative effect. Of course, we should be aware that this point will be more important when the robot will behave in different situations with global perception, which will affect whether the developed technique will be applicable for a robot without global perception. Since the presence of the humanoid robot is very strong, usually people (subjects) seem to interact primarily with the robot, rather than with the motion capturing system. Thus, if we appropriately design the system so that, for example, the robot does not react to visual stimuli that are out of the sight from the robot's eye, we can exploit the global perception in order to develop interaction mechanisms for the humanoid robot.

5 Conclusion

This paper reported the development of an autonomous interactive humanoid robot that is capable of "pretending listening behavior" based on embodied cooperative behaviors such as the eye contact and synchronization of arm movements seen in inter-human communication. We conducted an experiment in a route guidance situation where a human *teacher* taught a route to the robot. The results revealed that the developed robot has a positive effect on the teacher's impression about reliability and sympathy. Moreover, the detailed analysis indicated that both body movements and utterances contributed to the impression, though the body movement factor was the more dominant one. In contrast, the robot's utterances encouraged the human teacher's utterances to the robot, but the body movements did not. Thus, the importance of both utterances and body movements was demonstrated. To summarize, we developed "pretending listening" behaviors for a humanoid robot by reactively controlling its head, arms and utterances to the speaking person, which is a fundamental technique for a humanoid robot that is able to naturally communicate with people as humans do.

Acknowledgments This research was supported by the National Institute of Information and Communications Technology of Japan.

References

Billard, A. and Mataric, M. 2001. Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture. *Robotics and Autonomous Systems*, 37(2–3):145–160.

Breazeal, C. and Scassellati, B. 1999. A context-dependent attention system for a social robot. In *Proc. Int. Joint Conf. on Artificial Intelligence*, pp. 1146–1151.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., and Yan, H. 1999. Embodiment in conversational interfaces: Rea. In *Proceeding of the CHI' 99 Conference on Human Factors in Computing Systems*, pp. 520–527.

Fujita, M. 2001. AIBO: Towards the era of digital creatures. *International Journal of Robotics Research*, 20:781–794.

Hirai, K., Hirose, M., Haikawa, Y., and Takenaka, T. 1998. The development of the Honda humanoid robot. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1321–1326.

Imai, M., Ono, T., and Ishiguro, H. 2003. Physical relation and expression: Joint attention for human-robot interaction. *IEEE Transaction on Industrial Electronics*, 50(4):636–643.

Jebara, T. and Pentland, A. 1999. Action reaction learning: Automatic visual analysis and synthesis of interactive behaviour. In *Proceedings of International Conference on Computer Vision Systems*.

Kanda, T., Ishiguro, H., Imai, M., and Ono, T. 2003. Body movement analysis of human-robot interaction, In *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 177–182.

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. 2004a. Interactive robots as social partners and peer tutors for children: A field trial, *Human Computer Interaction*, 19(1–2):61–84.

Kanda, T., Ishiguro, H., Imai, M., and Ono, T. 2004b. Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE*, 92(11):1839–1850.

Kidd, C. and Breazeal, C. 2004. Effect of a robot on user perceptions. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2004)*, pp. 3559–3564.

Maynard, S. 1986. On back-channel behavior in Japanese and English casual conversation. *Linguistics*, 24:1079–1108.

Miyashita, T. and Ishiguro, H. 2004. Human-like natural behavior generation based on involuntary motions for humanoid robots. *Robotics and Autonomous Systems*, 48:203–212.

Moore, C. and Dunham, P.J. (eds.). 1995. Joint attention: Its origins and role in development. Lawrence Erlbaum Associates.

Nakano, Y., Reinstein, G., Stocky, T., and Cassell, J. 2003. Towards a model of face-to-face grounding. In *Proc. Association for Computational Linguistics*, pp. 553–561.

Nakadai, K., Hidai, K., Mizoguchi, H., Okuno, H. G., and Kitano, H. 2001. Real-time auditory and visual multiple-object tracking for robots. *International Joint Conference on Artificial Intelligence*, pp. 1425–1432.

Ono, T., Imai, M., and Ishiguro, H. 2001. A model of embodied communications with gestures between humans and robots, In *Proceedings of Annual Meeting of the Cognitive Science Society*, pp. 732–737.

Ogawa, H. and Watanabe, T. 2001. InterRobot: Speech-driven embodied interaction robot. *Advanced Robotics*, 15(3):371–377.

Reeves, B. and Nass, C. 1996. *The Media Equation.*

Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., and Fujimura, K. 2002. The intelligent ASIMO; System overview and integration. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'02)*, pp. 2478–2483.

Sakamoto, D., Kanda, T., Ono, T., Kamashima, M., Imai, M., and Ishiguro, H. 2005. Cooperative embodied communication emerged by interactive humanoid robots. *International Journal of Human-Computer Studies*, 62:247–265.

Scassellati, B., 2000. Investigating models of social development using a humanoid robot. *Biorobotics.*

Trafton, J.G., Cassimatis, N.L., Bugajska, M.D., Brock, D.P., Mintz, F.E., and Schulz, A.C. 2005. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Trans. on Systems, man and Cybernetics, Part A: Systems and Humans*, 35(4):460–470.

Tsukahara, S. and Hanazawa, S. 1997. Effects of back-channel in inter-human communication. In *proceedings of 61st Annual Conference of the Japanese Psychology Society*, p. 134.

Whittaker, S. and O'Conaill, B. 1997. The role of vision in face-to-face and mediated communication. In K. Finn, A. Sellen, and S. Wilbur, (eds.) *Video Mediated Communication*, Lawrence Erlbaum Associates, pp. 23–49.