

Natural Deictic Communication with Humanoid Robots

Osamu Sugiyama, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro and Norihiro Hagita

Abstract— A simple view of deictic communication only includes the indication process and recognition process: a person points at an object and says something about it such as “look at this,” and then the other person recognizes the pointing gesture and pays attention to the indicated object. However, this simple view lacks three important processes: attention synchronization, context focus, and believability establishment. We refer to these three processes as “facilitation processes” and implement them in a humanoid robot with a motion capturing system. An experiment with 30 subjects revealed that the facilitation processes make deictic communication natural.

I. INTRODUCTION

The aim of our research is to develop “communication robots” that naturally interact with humans and support daily human activities based on advanced interaction capabilities with their human-like bodies. Since the target audience of a communication robot is ordinary people who do not have specialized computing and engineering knowledge, a conversational interface using both verbal and non-verbal expressions is becoming more important. For example, the precedent research works have shown the effectiveness of facial expressions [1], eye-gaze [2], and gestures [3].

This study focuses on deictic communication as shown in **Fig. 1**. In casual communication, people often use reference terms in combination with pointing, such as saying “look at *this*” while pointing at an object in order to draw others’ attention to the object. A simple view of deictic communication is the use of attention-drawing behavior based on pointing gestures with reference terms. People use pointing gestures as one of the mechanisms of joint attention [4]. Reference terms are important for quickly and naturally informing the listener of an indicated object’s location. It is important for a communication robot to be capable of both understanding people’s attention-drawing behavior and performing attention-drawing behavior for people [5]. This mutual attention-drawing activity is classified into an indication process and a recognition process in Fig. 1.

However, this simple view only encompasses the processes related to the interpretation of information, and it lacks a process for facilitating interaction. We believe that three major processes facilitate deictic communication. The first one is attention synchronization. When we point at an

object, we pay attention to whether the listener’s gaze is following our pointing gesture. Thus, if a communication robot is listening to a person using attention-drawing behavior, it needs to indicate that it is paying attention to the pointing gesture in order to facilitate the person’s attention-drawing behavior. The second process is context: when we use a reference term such as “look at *this* (white round box)” we omit the details such as “white round box.” This is relatively easy to understand if the context (in this example, the box) is established before the use of reference terms. The third process is believability. People sometimes feel that deictic communication is inaccurate [6]. Here, people need to believe the robot’s capability in terms of error-correction; otherwise, people may hesitate to use deictic gestures/utterances with a robot.

This paper addresses the facilitation processes in deictic communication, whereas the indication and recognition processes was reported in our previous work [5]. We implement the facilitation processes in a humanoid robot, Robovie, so that it can engage in the full range of deictic communication. Its effectiveness is demonstrated through an experiment.

II. RELATED WORKS

There have been many research works on deictic communication. These studies often focus on the recognition process: They proposed methods to recognize the object by using a pointing gesture, an utterance, and stored object information [7][8][9]. Moreover, some studies integrated the recognition process with the indication process [10] [11], but they could not dynamically handle environments where the locations of objects and people change; that is, they only performed pre-implemented gestures and utterances. Overall, these studies did not yet tackle facilitation processes.

Regarding facilitation processes, several previous studies in conversational agents revealed the importance of context and believability [12]. However, since these studies involved an agent on a computer screen, they did not address such issues in deictic communication, and no study has addressed the attention synchronization process. Many studies have utilized the body properties of a robot [1][2][3] to facilitate human-robot interaction in general; however, none of them addressed the facilitating processes in deictic communication. Actually, previous research in robotics has given scant attention to utilizing facilitating processes in deictic communication. A few studies have addressed synchrony in body movements in human-robot interaction. For example, Kanda and his colleagues found the importance of cooperative body movements [13] and

All the author is with ATR Intelligent Robotics and Communication Laboratory, 2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan (phone:+81-774-95-1492, e-mail: kanda@atr.jp).

Osamu Sugiyama and Michita Imai are also with graduate school, Keio University, Kouhokoku, Yokohama City, Kanagawa, Japan.

Hiroshi Ishiguro is with Faculty of Engineering, Osaka University, Suita City, Osaka, Japan.

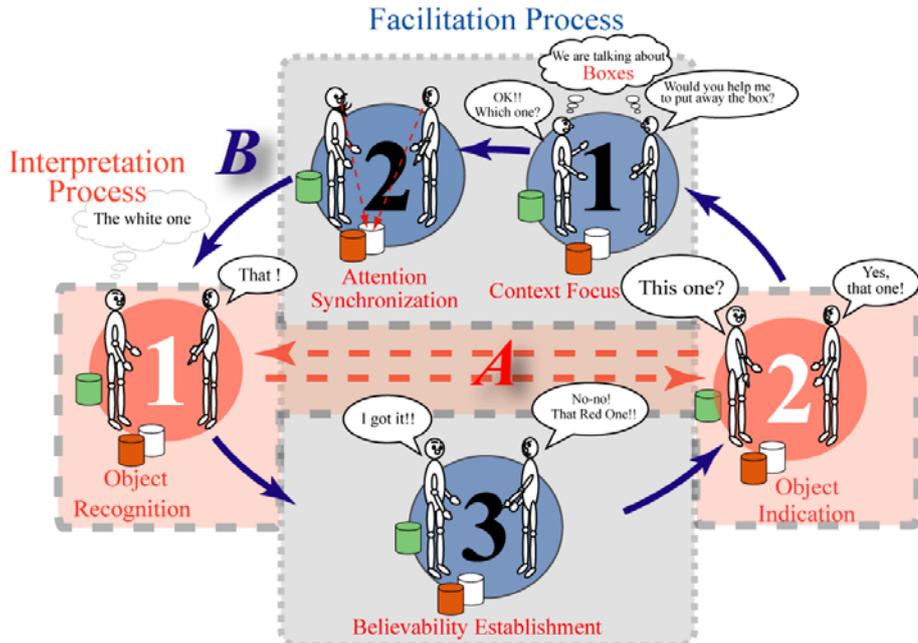


Fig.1: Interpretation and Facilitation Processes in Natural Deictic Communication

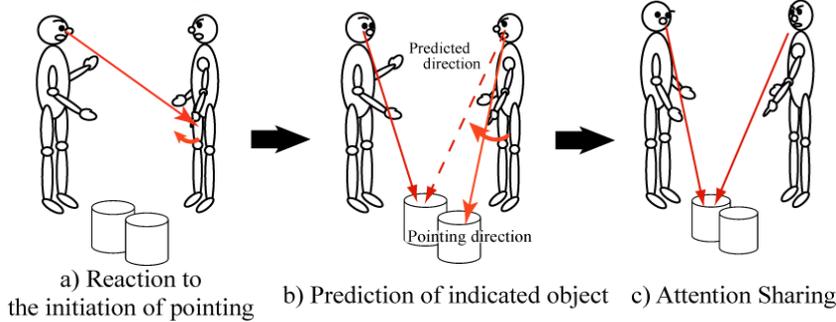


Fig.2: The model of Attention Synchronization

developed a robot that expresses cooperative body movements to establish reliability and sympathy in human-robot interaction [14]. 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 [15]. However, these previous studies were not concerned with deictic communication.

III. NATURALNESS IN DEICTIC COMMUNICATION

In this research, we outlined the following five continuous processes, as shown in Fig. 1, for achieving natural deictic communication between robots and humans.

1. **Context Focus** Speaker and listener share the objects of focus based on the context of the previous verbal communication.
2. **Attention Synchronization** Listener pays attention to the indicated object by synchronizing with speaker's attention-drawing behavior.
3. **Object Recognition** Listener recognizes the indicated object based on attention-drawing behavior.
4. **Believability Establishment** Speaker corrects a recognition error of listener when he notices the misunderstanding.

5. **Object Indication** Listener, at this stage changing his role to speaker, confirms his recognized object with attention-drawing behavior.

We categorized these five processes into an Interpretation Process and a Facilitation Process. The Interpretation Process involves interpreting an attention-drawing behavior. On the other hand, the Facilitation Process involves facilitating natural deictic communication between humans and robots. Detailed explanations are given as follows.

A. Interpretation Process

When we talk about the object in an environment, we point at the object of our consideration with a deictic pronoun such as the reference terms 'this' and 'that' and draw the listener's attention toward the object. We refer to this behavior utilizing pointing gestures as well as reference terms as attention-drawing behavior. The Interpretation Process consists of Object Recognition and Object Indication as shown in Flow A of Fig. 1.

Object Recognition

This is the process of interpreting an attention-drawing behavior based on its pointing posture as well as a reference term and object-property information in the utterance [5].

Object Indication

This is the process of generating a robot's attention-drawing behavior with pointing and a reference term in order to confirm to the recognized object with the interacting person.

B. Facilitation Process

However, it is not sufficient to achieve natural deictic communication between robots and humans only with interpretation processes. When we talk about the object in an environment with attention-drawing behavior, the speaker always monitors the listener's reaction toward his behavior and evaluates whether the listener understands this behavior. At the same time, the listener also picks up the intention of the speaker's attention-drawing behavior and synchronizes his gaze by following the behavior. That is, deictic communication is not a simple process where one person is pointing and another person is passively interpreting the pointing; it is a dynamic process in which the speaker and listener mutually engage in an interaction within a short time of a second or less. Without this dynamic process, deictic communication would merely be giving a command to a static machine, which is not as natural as inter-human communication at all.

We believe that the robot should simultaneously react to human attention-drawing behavior as a listener while at the same time setting up an environment where the human as a listener feels comfortable in accepting the robot's attention-drawing behavior.

From this point of view, we added three new processes: *Context Focus*, *Attention Synchronization* and *Believability Establishment*. Flow B in Fig. 1 outlines the five processes to achieve natural deictic communication.

Attention Synchronization

This process is to provide the interacting person with a feeling that the robot is attending to his or her pointing behavior. The flow of attention synchronization is shown in Fig. 2. This consists of three sub-processes:

(a) Reaction to the initiation of pointing

When the listener notices the initiation of a speaker's attention-drawing behavior, the listener immediately starts to follow it with his gaze.

(b) Prediction of indicated object

Next, the listener estimates the intention of the speaker's attention-drawing behavior and predicts the object that the speaker is going to indicate. The listener often starts looking at the object before the pointing motion is finished.

(c) Attention sharing

Finally, when the speaker finished moving his hand in pointing, the speaker fixes the pointing gesture toward the indicated object and looks at it. At this time, the listener is also looking at the object. Thus, they are sharing their attention toward the object when the pointing motion is completed.

We took these actions as a reflex movement and not noticed its importance in inter-human deictic communication.

However, attention synchronization must be one of the essential processes needed to facilitate natural deictic communication. Since such attention synchronization occurs in human communication, we believe that the robot should have this capability; otherwise, the speaker would have misgivings about the listener's understanding of the indicated object. In addition, we assume that attention synchronization makes the interaction quicker, since the speaker becomes more certain about the interaction.

Context Focus

This process sets up a presupposition toward the deictic communication by utilizing attention-drawing behavior. In inter-human communication, the speaker and the listener share the objects of focus based on the context of the previous verbal communication before the deictic communication. Without this process, there would be a flood of indication candidates in the living environment, resulting in the human feeling uncomfortable toward deictic communication with a robot. By focusing on a certain kind of object, the human naturally immerses himself in deictic communication with a robot.

Believability Establishment

This process provides human believability toward a robot that is capable of interacting while using attention-drawing behavior. Communication having ambiguous expressions, such as those in deictic communication, is built upon the believability of the partner also being capable of managing an ambiguous expression; otherwise, the human would hesitate to interact with the robot in a deictic way. In this research, we construct this believability by creating a robot that is capable of error correction. In inter-human deictic communication, the speaker can correct the listener's error immediately whenever he or she notices the listener's misunderstanding. It is very clear that a human has this ability, but it is not clear that a robot is capable of handling error correction. Therefore, it is essential to show that the robot has the ability to correct errors when it misunderstands a human's indication. Developing this ability in the robot is considered key to making humans possess believability toward the deictic communication with the robot. Accordingly, this is one of the essential processes in achieving deictic communication that is as natural as inter-human communication.

IV. DEVELOPMENT OF A COMMUNICATION ROBOT CAPABLE OF NATURAL DEICTIC COMMUNICATION

Based on the processes introduced in Section III, we developed a communication robot that can engage in natural deictic communication with a human. This section describes the details of this robot system.



Fig. 3: Vicon motion-capturing system

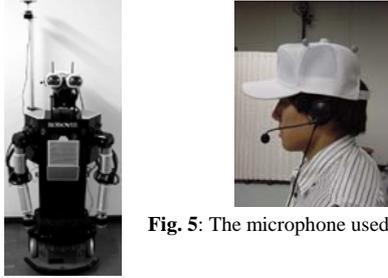


Fig. 4: Robovie

Fig. 5: The microphone used in system

A. Hardware Configuration

A robot system was developed using a communication robot, “Robovie” [16], a motion-capturing system, and a microphone. “Robovie” (Fig. 4) is 1.2 m tall with a 0.5 m radius and a human-like upper body designed for communicating with humans. It has a head (3 DOF), eyes (2*2 DOF), and arms (4*2 DOF). With a speaker in its head, it can produce sound output. With its 4-DOF arms, it can point with a gesture similar to that of humans. Using its motion-capturing system (Fig. 3), we can obtain 3D positions from markers attached to the person, Robovie, and objects; accordingly, by using an Ethernet, the system obtains their 3D positions as input and then calculates their locations. In addition, the speaker wears a microphone as shown in Fig. 5 to avoid difficulty with speech recognition.

B. Software Configuration

The configuration of the developed robot system is shown in Fig. 6. The interpretation processes in Figure 1 are achieved by the white-colored components. Meanwhile, the facilitation processes are achieved by the orange-colored components. Each process gets its input from either/both speech recognition or/and gesture recognition, and it controls Robovie’s behavior as its output. For speech recognition input, we used a speech recognition system [17] and attached a microphone to a human in order to avoid noise (we appreciate that there are many studies on noise-robust speech recognition that do not require microphones attached to people). The speech recognition system can recognize reference terms, object-color information, and error-alerting words in human utterances. On the other hand, gesture recognition is done with a motion-capturing system. In gesture recognition, the system handles two types of processes: pointing posture recognition and pointing motion detection. The following describes the details of implementing each process in Fig. 1.

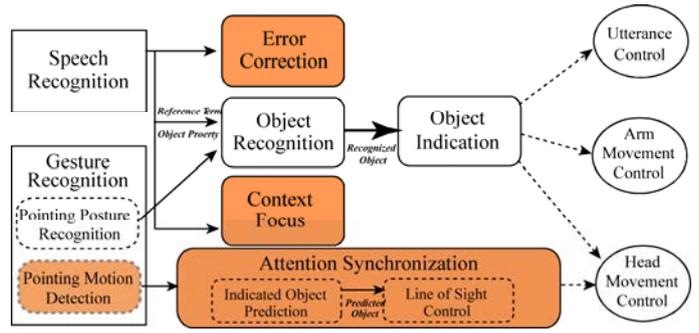


Fig 6: System Configuration

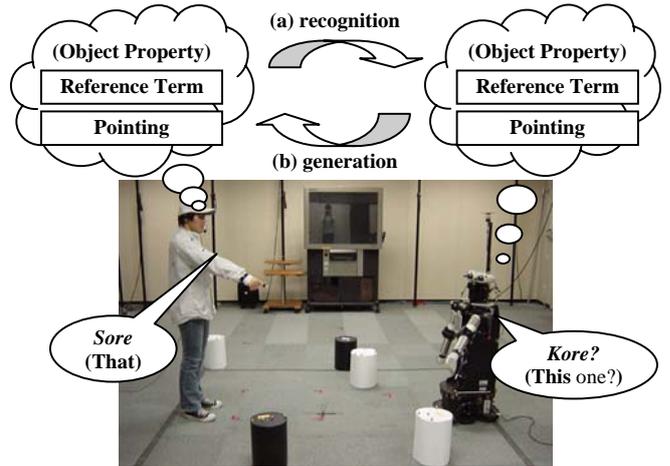


Fig. 7: Attention-drawing behavior

C. Interpretation Processes

The Interpretation Processes are developed based on the three-layer model previously developed in [5]. In Object Recognition, the system receives a reference term and object-property information from the speech recognition module and pointing gesture information from gesture recognition in order to identify the indicated object (See Fig.7 (a)). Meanwhile, in Object Indication, the system generates an appropriate attention-drawing behavior enabling the humanoid robot to confirm the human-indicated object with the same expression (Fig.7 (b)). The details are described in [5].

D. Facilitation Processes

This section describes the implementation of Facilitation Processes designed to facilitate natural deictic communication between humanoid robots and humans. The following describes the details of the Facilitation Processes.

Context Focus

The Context Focus process is implemented as voice communication with the human. In the first stage of the interaction in our experiment, the robot (R) and the human (H) have a short conversation:

R: May I help you?

H: Yes, please help me to put away these *round boxes*.

R: OK. Would you tell me the order of the *boxes* to put away?

In this case, the robot and the human can focus on the round boxes in the environment based on the context described above. It is very important for both robots and humans to be aware of specific objects indicated by the context. Otherwise, they would have to consider objects from a wide variety of candidates when they engage in deictic communication. In this research, the robot could only handle the round boxes. Thus, this conversation is only for the human to mentally prepare for the objects in the context. Enabling robots to distinguish objects by context remains a future work.

Attention Synchronization

The process of Attention Synchronization is implemented by three sub-processes: Pointing Motion Detection, Indicated Object Prediction and Line of Sight Control. The important requirement for attention synchronization is that the robot starts looking at the indicated object quickly before the person finishes the pointing motion. Thus, the robot has to predict the indicated object from the pointing motion. The following describes each sub-process in order.

(a) Pointing Motion Detection

This process monitors human arm motion and determines whether the present human motion is a pointing gesture. Human pointing gesture has rich patterns: some people twist the body to point at a left-side object with the right hand, and others first raise a finger near the head and then point at the object. On the other hand, humans also exhibit various movements when they are not pointing but just standing and talking. In order to detect a human pointing gesture in various situations, we utilize three different parameters for pointing gestures:

- i) *The speed of fingers* v_p ;
- ii) *The distance between the human's finger and body* d_p ;
- iii) *The angle between the human's vertical axis and his/her pointing direction* θ_p .

If the human is in the process of performing a pointing gesture, one of these parameters should be increased. We analyze various pointing motions and set up a threshold for each parameter. The evaluation equation is given as follows.

$$f(t) = \begin{cases} 1 & (v_p > 100_{mm/s} \parallel d_p > 400mm \parallel \theta_p > 15^\circ) \\ 0 & (v_p \leq 100_{mm/s} \parallel d_p \leq 400mm \parallel \theta_p \leq 15^\circ) \end{cases}$$

where t is a frame of the motion-capturing data, and '1' means detection of pointing while '0' means no detection.

(b) Pointing Object Prediction

This process predicts the object that a human is going to indicate in the first stage of pointing motion so that the robot can synchronize its gaze to the human's pointing motion. In order to predict the object that the human is going to point, the system predicts a future human finger position based on the present speed vector of the fingers. Utilizing the position of fingers, it searches for the indicated object.

The difficulty is in deciding the finger position at how many seconds later the system should use in predicting an indicated object. If this time interval is too long, the robot's line of sight goes past the indicated object and sees a farther object. If it is too short, the robot's line of sight stays on a

nearer object. In order to solve this trade-off problem, we conducted an easy experiment to evaluate how far in the seconds later a finger position should be for use in appropriately predicting an indicated object. The results led us to set a prediction time of 0.3 seconds later. The system predicts the human's finger position at 0.1–0.3 seconds later, and based on these finger positions, it searches for the object and decides on the object nearest the calculated position temporally and spatially.

(c) Line of Sight Control

This process controls Robovie's line of sight expressed by its head sliding toward the indicated object predicted by the process of Pointing Object Prediction. The process controls Robovie's line of sight as follows;

1. If pointing gesture is not detected, the process controls the line of sight to see human's face.
2. If pointing gesture is detected and indicated object is predicted, the process controls the line of sight to see the indicated object.
3. If pointing gesture is detected but there are no predicted objects, the process controls the line of sight to see the former predicted object (if there are no predicted object in the past, see the human's face.).

Error Correction

The process of error correction is to establish believability toward the robot. It is implemented as voice communication with the human. If the process of speech recognition detects error-alerting words in a human utterance, this process is called by the system. In this process, the robot asks the human to repeat his or her indication and deletes the latest recognized object from its memory. By successfully using this process on the robot system, the human can feel that it is possible to correct the indication if the robot initially misunderstands it. Consequently, this improves the human's sense of believability toward deictic communication with the robot.

V. EXPERIMENT

The purpose of the experiment is to verify the effectiveness of the facilitation process. We particularly focused on the subjective impression, since the facilitation process is supposed to mentally affect the interacting person. Because facilitation processes are considered to be mutually affected by each other, we evaluated the total impression of the facilitation processes. We also compared the developed system of deictic communication with an alternative way of communication in order to demonstrate the performance of the developed system. In the experiment, subjects were asked to teach the robot the order of five boxes that the robot was to bring.

A. Method

Experimental Environment The experiment was conducted in a rectangular room 7.5 m by 10.0 m (Fig. 8 and 9). Due to the limitations imposed by the motion-capturing system, we only used an area 3.5 m by 3.5 m in the center of the room. There are five cylindrical boxes (radius 110 mm, height 300 mm). At the beginning of each session, each



Fig. 8: A scene of the experiment

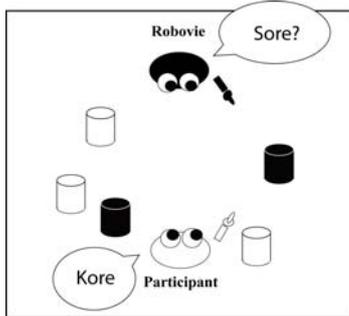


Fig. 9: Experimental Environment

subject was asked to place these five boxes freely in the area except for places to where Robovie could not point. The subject stood in front of the robot, usually at distance of 0.5 m to 1.5 m from the robot.

Subjects Thirty university students (16 men, 14 women) participated in the experiment.

Conditions We adopted 2 (facilitation process) x 2 (method of instruction) conditions for the experiments. The definition of each factor is as follows.

Facilitation process factors

a) With facilitation process

Before the session, subjects were instructed how to correct errors of the robot, and they could do so freely as they gave orders (error correction). They were also asked to start giving an order while explicitly mentioning the context. For example, they might say, “please bring the **boxes**” (context). During the session, the robot controlled its gazing direction to the direction in which the subjects were going to point, which is based on the developed mechanism of attention synchronization (attention synchronization). Note that the robot also looked at boxes at the confirmation phase.

a') Without facilitation process

The subjects were given none of the instructions mentioned above. When subjects gave the object order, the robot did not control its gazing direction but just looked in the subject’s direction. Subjects were not allowed to correct errors of the robot. Note that the robot looked at boxes at the confirmation phase.

Method of instruction factor

b) Deictic method (pointing + reference term)

Subjects were asked to use pointing gesture, reference terms and object color information when they give orders to the robot.

b') Symbolic method

Subjects were asked to read the numbers on the boxes when they give orders to the robot. Two-digit IDs were attached to the box in 14-point font, which is readable from a 2-m distance.

This setup is intended to simulate a situation where there is some difficulty in finding the symbol used to identify the object. In our daily situations, for example, we might say “please look at the book entitled *communication robots*”; here, we need to see the characters on the object and read them. This load was simulated with the two-digit ID with a font size of 14 points.

The experiment was a within-subject design, and the order of experimental trials was counterbalanced.

B. Procedures

Before the experiment, subjects were instructed on how to interact with the robot, by both the deictic method (reference terms and pointing) and the symbolic method. After the instructions, subjects experienced four sessions for all four conditions. In each session, they conducted the following procedures three times:

1. They freely place the five boxes.
2. Under (a) condition, they talk about context.
3. They decide the order of five boxes.
4. They indicate the five boxes one by one. The method of instruction is either (b) or (b’), depending on the experimental condition. For example, under (b) condition, a subject might point at the first box and say “this” and then at the second box and say “that white one.” The subject continues this until the fifth object.
5. After the subjects indicate the fifth box, Robovie repeats the given order. This time, the robot looks at each object with its utterance for confirmation. For example, under (b) condition, Robovie uses reference terms while pointing at each of the five boxes.

After the three repetitions, subjects fill out a questionnaire to give their impressions of the interaction with the robot. We designed the interaction to include the quickness of indicating objects; for this purpose, subjects continuously indicate a box five times. Since the impression of the interaction could depend on the placement of the boxes, we gave subjects three opportunities to experience the interaction under each condition.

C. Measurement

We expected that the facilitation process would make the deictic communication of the robot more natural. In addition, we expected that the attention-synchronization mechanism would give a better impression of sharing spatial information of the boxes. Thus, we measured the following impressions with 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.

Naturalness: Your feelings of naturalness of the conversation

Sharing information: Your feelings of sharing information with the robot

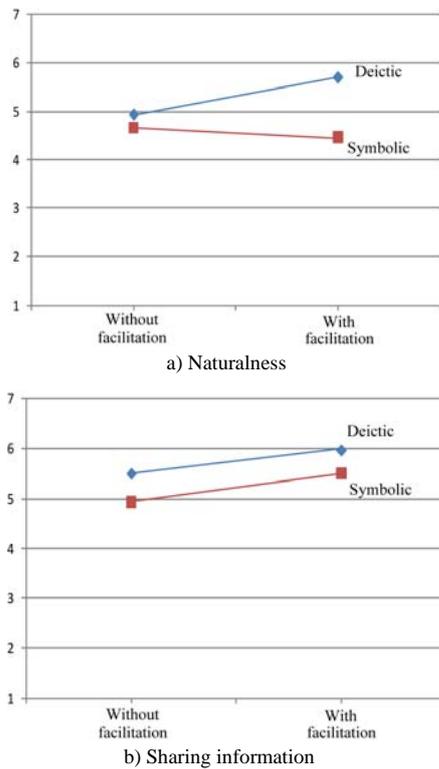


Fig.10-1: Results for facilitation process

Regarding the comparison between the deictic method and the symbolic method, we are interested in quickness and correctness [6]. Thus, we also measured the following impressions in the questionnaire:

Quickness: Quickness of your indication of the boxes

Correctness: Correctness of the robot's identifying your indication of the boxes

Understanding: The robot's understanding of the indication
We also measured the system's performance as the accuracy of recognizing indication in both the deictic method and symbolic method. The main cause of error comes from failure in speech recognition. Since speech recognition sometimes fails whereas the deictic method utilizes multimodal input, we expected the deictic method to provide better performance in reality.

Performance: The rate that the robot system correctly identified the object indicated by the subjects

D. Results

Table I shows the system's performance. Figures 10-1 and 10-2 show the results of the questionnaire. In order to analyze the results of impressions, we conducted repeated measures ANCOVA (Analysis of covariance), which is an extension of repeated measures ANOVA (Analysis of variance) to include covariance. There are two within-subject factors, *facilitation process* (with facilitation process or without facilitation process) and *method of instruction* (deictic method or symbolic method). In addition, we used the *performance* (the rate that the robot system correctly identified the object) for the covariance, since some of the impressions, such as the impression of the

Table I: Performance

	w/o facilitation	with facilitation
Deictic	98.4%	98.8%
Symbolic	95.9%	95.0%

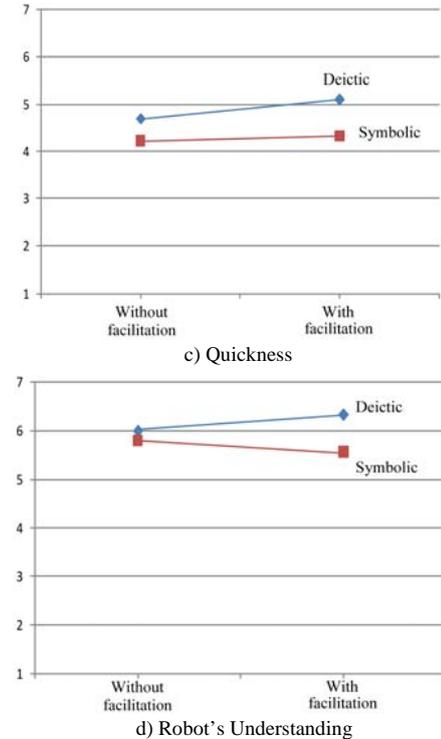


Fig. 10-2: Results for method of instruction

robot's understanding, could be affected by the real performance rather than by the conditions.

Naturalness (Fig. 10-1 (a)) There were a significant difference in the method of instruction ($F(1,86)=8.142$, $p<.01$) and the interaction between the two factors ($F(1,86)=4.988$, $p<.05$). Analysis of the interaction resulted in a significant difference in the method of instruction under the with-facilitation process condition ($p<.01$) as well as a significant difference between the facilitation processes under the deictic communication condition ($p<.01$). Thus, the facilitation process contributed to the feeling of natural interaction when subjects interacted with the robot in deictic communication.

Sharing information (Fig. 10-1 (b)) There were significant differences in the facilitation process ($F(1,86)=7.121$, $p<.01$). There were nearly significant differences in the method of instruction ($F(1,86)=3.782$, $p=.055$), and no significant difference in the interaction between the two factors. Thus, subjects felt that the robot with a facilitation process is better at sharing information than a robot without it. In addition, it was suggested that deictic communication could also contribute to the feeling of sharing information better than could symbolic communication.

Quickness (Fig. 10-2 (c)) There were significant differences related to the instruction method ($F(1,86)=8.100$, $p<.01$). There was no significant difference in the facilitation process, interaction between the two factors, and

performance. Thus, subjects felt that the deictic method is quicker than the symbolic method.

Correctness There was no significant difference in facilitation process, instruction method, and interaction between the two. There was nearly a significant difference in the performance ($F(1,86)=2.978$, $p=.088$). This suggests that subjects might attribute the robot's performance as the correctness impression, whereas they did not feel the difference in either facilitation process or instruction method.

Understanding (Fig. 10-2 (d)) There was significant difference in the performance ($F(1,86)=6.453$, $p<.05$), and no significant difference in the other factors. Thus, subjects distinguished the real performance and evaluated it as the robot's understanding.

Performance (Table I) For performance, we conducted within-subject design ANOVA, which indicates a significant difference in the method of instruction ($F(1,86)=16.300$, $p<.01$). There was no significant difference in the facilitation process. Thus, the multimodal input used in the deictic method positively affected the system performance.

E. Discussion

Effect of facilitation process

The experimental results revealed that the facilitation process has a positive effect on the impression of naturalness and sharing information. We also obtained free-form descriptions of this effect:

- "The head motion (in the facilitation process) is human-like."
- "The robot reacted to my motion, which gave me a lifelike impression."
- "I felt it great that the robot is capable of error-correction."
- "Although it is a robot, I felt it unnatural since the robot did not respond to me."

Of course, the experimental environment is not realizing the human conversation which is more ambiguous and polysemous, still we believe that these findings demonstrate the importance of the facilitation process for deictic communication.

Deictic communication vs. Symbolic communication

Since the deictic method integrates the input from gesture recognition and speech recognition [5], it is robust to an error in recognition. For example, if it misunderstands an utterance of "this" as "that," the system could mostly identify the correct object from the pointing gesture. This benefit was clearly demonstrated in the system's performance as well as the impressions of correctness and robot's understanding. Moreover, deictic communication was evaluated as quicker than symbolic communication, and possibly better in the impression of sharing information. These findings support the assumption that deictic communication is useful in human-robot interaction.

VI. CONCLUSION

We proposed three facilitation processes in deictic communication: attention synchronization, context focus and believability establishment. We implemented these processes as well as indication-recognition processes in a

humanoid robot, Robovie, with a motion-capturing system. We conducted an experiment with 30 subjects to clarify the effectiveness of the facilitation processes. This revealed that the facilitation processes make deictic communication natural. Subjects also reported that they felt the feeling of sharing information with the robot that used the facilitation processes. In addition, the comparison with a symbolic-command condition revealed that deictic communication is better at recognizing indicated objects due to multimodal processing.

ACKNOWLEDGEMENT

This research was supported by the Ministry of Internal Affairs and Communications of Japan.

REFERENCE

- [1] C. Breazeal and B. Scassellati: "Infant-like social interactions between a robot and a human caretaker," *Adaptive Behavior*, 8(1), 2000.
- [2] K. Nakadai, K. Hidai, H. Mizoguchi, H. G. Okuno, and H. Kitano: "Real-Time Auditory and Visual Multiple-Object Tracking for Robots," *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI'01)*, pp. 1425-1432, 2001.
- [3] T. Ogata and S. Sugano, Emotional Communication Between Humans and the Autonomous Robot Which Has the Emotion Model, *Int. Conf. on Robotics and Automation (ICRA'99)*, pp. 3177-3182, 1999.
- [4] C. Moore and Philip J. Dunham eds: *Joint Attention: Its Origins and Role in Development*, Lawrence Erlbaum Associates, 1995.
- [5] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, "Three-Layer Model for Generation and Recognition of Attention-Drawing Behavior", *Int. Conf. on Intelligent Robots and Systems (IROS2006)*, 2006.
- [6] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita: "Three-layered Draw-Attention Model for Humanoid Robots with Gestures and Verbal Cues," *Int. Conf. on Intelligent Robots and Systems (IROS2005)*, pp. 2140-2145, 2005.
- [7] A. Haasch, N. Hofemann, J. Fritsch, and G. Sagerer: "A multi-modal object attention system for a mobile robot," *Int. Conf. on Intelligent Robots and Systems (IROS2005)*, pp. 1499-1504, 2005.
- [8] Z. M. Hanafiah, C. Yamazaki, A. Nakamura, and Y. Kuno: "Understanding Inexplicit Utterances Using Vision for Helper Robots," *Int. Conf. on Pattern Recognition*, 2004.
- [9] T. Inamura, M. Inaba, and H. Inoue: "PEXIS: Probabilistic Experience Representation Based Adaptive Interaction System for Personal Robots," *Systems and Computers in Japan*, Vol. 35, No. 6, pp. 98--109, 2004.
- [10] B. Scassellati: "Investigating Models of Social Development Using a Humanoid Robot," *Biorobotics*, MIT Press, 2000.
- [11] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin: "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," *Int. Conf. on Intelligent Robots and Systems (IROS2005)*, pp. 383-388, 2005.
- [12] J. Cassell, *Conversational Agents*, The MIT Press, 2000.
- [13] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, Body movement analysis of human-robot interaction, *Int. Joint Conf. on Artificial Intelligence (IJCAI2003)*, pp. 177-182, 2003.
- [14] M. Kamashima, T. Kanda, M. Imai, T. Ono, D. Sakamoto, H. Ishiguro, and Y. Anzai: "Embodied Cooperative Behaviors by an Autonomous Humanoid Robot," *Int. Conf. on Intelligent Robots and Systems (IROS2004)*, pp. 2506-2513, 2004.
- [15] H. Ogawa and T. Watanabe, InterRobot: speech-driven embodied interaction robot, *Advanced Robotics*, Vol. 15, No. 3, pp. 371-377, 2001.
- [16] T. Kanda, H. Ishiguro, M. Imai, and T. Ono: "Development and Evaluation of Interactive Humanoid Robots," *Proceedings of the IEEE*, Vol. 92, No. 11, pp. 1839-1850, 2004.
- [17] C. T. Ishi, S. Matsuda, T. Kanda, T. Jitsuhiro, H. Ishiguro, S. Nakamura, and N. Hagita, Robust speech recognition system for communication robots in real environments, *Int. Conf. on Humanoid Robots (Humanoids2006)*, 2006.