An Approach to Integrating an Interactive Guide Robot with Ubiquitous Sensors

Yoshikazu Koide[1,2], Takayuki Kanda[1], Yasuyuki Sumi[1,2], Kiyoshi Kogure[1], and Hiroshi Ishiguro[1,3]

[1]ATR
Kyoto, Japan

[2]Kyoto University
Kyoto, Japan

[3]Osaka University
Osaka, Japan

E-mail: {yoshykaz, kanda, sumi, kogure, ishiguro}@atr.jp

Abstract— **An interactive exhibition guide robot has been developed, that is integrated with ubiquitous sensors. The robot utilizes its human-like body properties to perform human-like gestures such as eye-contact and pointing. The ubiquitous sensors are embedded in an exhibition room or attached to people in the room, and they provide information not only for the robot's interactive guide but also for the development of the robot. The results of a two-days exhibition experiment show that the robot attracted visitors and successfully guided them in the environment. Moreover, the obtained interaction statistics, including video from the ubiquitous sensors, prove to be useful for analyzing the interaction between the robot and visitors, and this implies that the use of such interaction data is a promising approach for developing interactive robots.**

Keywords: human-robot interaction; development methodology; ubiquitous sensor; sensor room environment;

I. INTRODUCTION

Over the past several years, many humanoid robots such as Honda's ASIMO have been developed. In the near future, humanoid robots are expected to have the capability of interacting with us in our daily life. Its human-like body enables us to intuitively understand its gestures and thus causes us to behave unconsciously as if we were communicating with humans [1]. That is, if a humanoid robot uses its body effectively, people will feel that their communication with it is natural. This could allow such a robot to perform communicative tasks in human society such as route guidance.

However, such a humanoid robot is difficult to develop for two major reasons. The first one is related to sensor capability, especially for communication. It is difficult to obtain information about its physical environment only by using sensors attached to its body. For example, a robot has difficulty in recognizing objects in the environment and correctly identify the person it is interacting with from among hundreds of candidates. Moreover, the current robotic technology is still far from possessing the capability of reasoning about others' interests and preferences, as well it is far from utilizing these properties to facilitate communication, a capability that humans take for granted.

The other difficulty is how to evaluate the communication between people and robots, which is closely related to decide the direction of robot development. Although several researchers have struggled to conduct the research on evaluating human-robot communication by utilizing subjective data [2, 3] and objective data such as alignment of bodily movement [4], the quantitative analysis approach is immature yet. It is so difficult to identify the inadequate points appearing during communication that there is no standard methodology for debugging and improving a robot's interactive behaviors.

Toward solving these problems, the emerging *ubiquitous sensor* technology is very promising. Thus we have developed a sensor room with ubiquitous sensors (i.e., both stationary and wearable) to identify persons and objects in the room [5, 6] and analyzed the interactions among people and robots. This environment makes it possible to gather a huge amount of data on the interactions among people and robots, which is structured into an *interaction corpus*. This corpus was successfully used to analyze the communication among people by Bono et al. [7]. This strongly suggests that this corpus is also useful for the analysis and evaluation of human-robot communication as well.

This paper proposes an approach for an interactive humanoid robot that facilitates a human visitor's experiences in an exhibition environment. With ubiquitous sensors installed in the environment, the robot can identify the visitors and recognize objects in the environment. Moreover, the environment provides the robot with information about each individual's histories of activities in the environment, which allows the robot to provide them personalized guidance. For example, each user's interest in an exhibition booth can be estimated from his/her consumption time at the booth. The robot can thus provide a visitor with guidance information best suit to the visitor without any question-answering dialogue, which is very difficult for human guides. While providing such information, moreover, the robot can perform human-like gestures such as pointing and eye-contact.

The interaction corpus gathered in the environment helps us to improve the robot's guiding behaviors. We have conducted a two-day experiment in a research exhibition, where the humanoid robot performed a guiding task for more than 170 visitors. After the experiment, we analyzed the reaction of the visitors accumulated as the interaction corpus. The analysis tool, called *corpus viewer*, provided us with videos from various angles, which were indexed precisely by robot's behaviors and other interaction events. As a result, this made it easier for us to

---

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

observe the visitor’s reaction to a specific behavior of the robots during the guidance and to identify the visitor’s objects of attention from the camera attached to them. This shows that the interaction corpus strongly supports our approach to develop an interactive guide humanoid robot that naturally communicates with humans.

II. SYSTEM CONFIGURATION

Figure 1 shows the concept of the developed guide robot; the *ubiquitous sensor room* captures visitors’ experience and records it to the database, called the *interaction corpus.* The interactive humanoid robot *“Robovie”* retrieves the recorded information from the *interaction corpus* to guide visitors to preferred exhibited contents.

A. Ubiquitous sensor room

Figure 2 is an exhibition scene in the ubiquitous sensor room where some of the humans wore the *wearable sensors* (also shown in **Figure 3**) and *stationary sensors* were embedded. This room is a rectangular room measuring 7.5 m by 10 m. The ubiquitous sensors (both wearable and stationary) include IR-trackers, LED-tags, video cameras and microphones. Each of these components is attached to the visitors, guides, the robots, and potential focal points such as posters in the room. By using the LED-based tag system (IR-tracker and LED-tags), the system identifies the objects that each camera captures. In other words, it enables the system to identify the focus of attention of each visitor.

An IR-tracker consists of a CMOS camera and an IR pass filter. It identifies an LED-tag, which is accomplished by receiving an infrared LED from the tag that illuminates its ID periodically. Humans in the room wear the tags shown in Figure 3. There are two types of wearable sensors: the wearable sensor set, which consists of an LED-tag, IR-tracker, camera and microphone, and the LED-tag badge-type is simplified so that it only consists of an LED-tag.

There are IR-trackers embedded in the environment. When an IR-tracker attached to an object receives the signal of an LED-tag attached to a visitor, the visitor’s focus of attention is interpreted to be on that object. For example, when a visitor is looking at a poster, the IR-tracker attached to the poster receives the signal of the visitor’s LED-tag. Similarly, when humans *A* and *B* are communicating, the system is capable of recognizing it as an IR-tracker attached to a human *A* receiving a signal of the other human *B* and vice versa. The sensor information from microphones supports the decision of whether they are talking interactively or a human is explaining something to the others rather unilaterally, such as guiding the exhibition [5]. These human experiences are recorded into the *interaction corpus* with meta-data such as *watching, and talking*. The videos are also captured and recorded into the *interaction corpus.*

B. Interactive humanoid robot “Robovie”

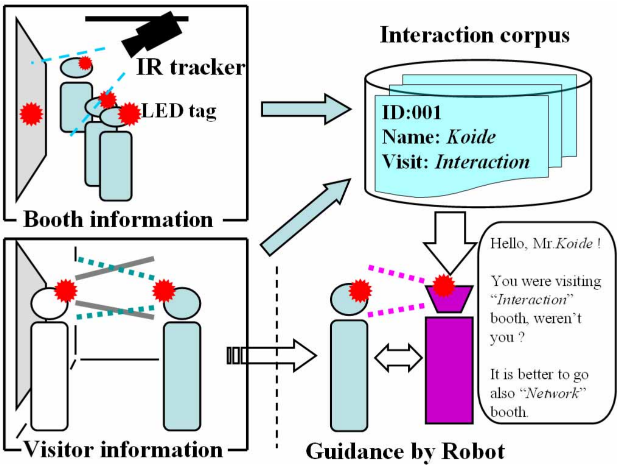


Figure 1: Concept of guide robot with ubiquitous sensors

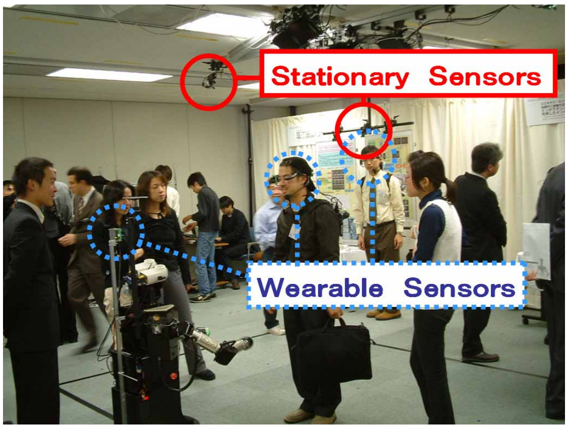
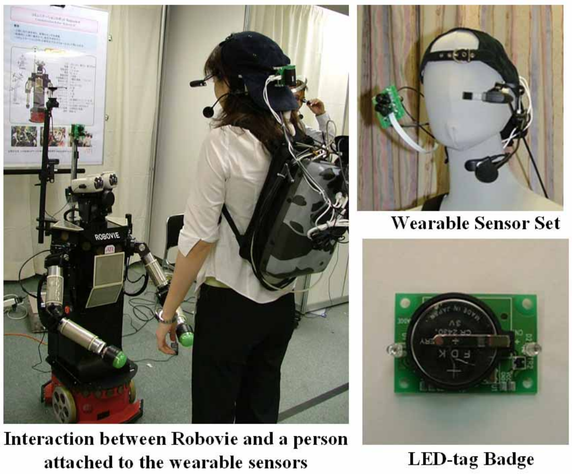


Figure 2: Ubiquitous sensor room


Interaction between Robovie and a person attached to the wearable sensors

Wearable Sensor Set

LED-tag Badge

Figure 3: Wearable sensors and Robovie

“Robovie” is characterized by its human-like body expression and various sensors (**Figure 3**: left). The human-like body consists of eyes, a head and arms. These components generate the complex body movements required for communication. In order not to give humans guard, the height of the robot is only 120 cm. The diameter is 40 cm. The robot has two arms (4*2 DOFs), a head (3 DOFs), and a mobile platform. It has a speaker for talking in a synthesized voice.

The sensory equipment includes auditory, tactile, ultrasonic, and vision sensors, which allow the robot to behave autonomously and to interact with humans. All the processing and control systems, such as the computer and motor control hardware, are located inside the robot's body. The sensory information from these sensors is also recorded in the *interaction corpus*.

Robovie has a software mechanism for performing consistent interactive behaviors based on *situated modules* [8]. A situated module realizes an interactive behavior such as shaking hands, greeting, briefly conversing, and pointing to an object in the surroundings, which is controlled by a simple rule set named *episode rules*.

C. Corpus viewer: a tool for analyzing interactions

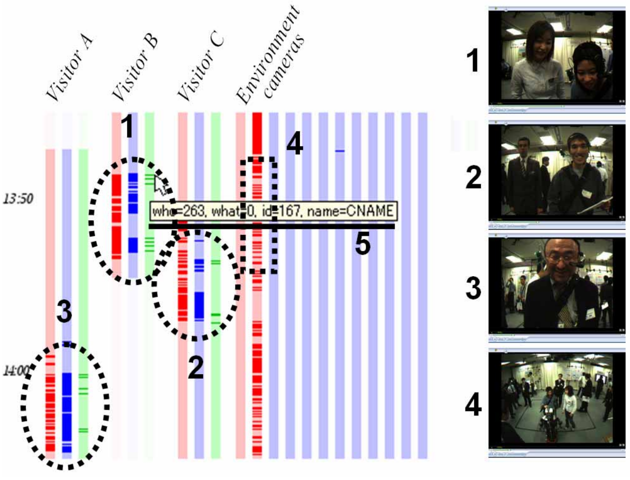
Corpus viewer is a tool to help us visualize the *interaction corpus* and to analyze the interaction by observing the videos. **Figure 5** is a snapshot of the *corpus viewer*. It enables us to easily retrieve the video of the interaction from various angles such as the camera attached to the interacting human (Fig. 5-II-a), robot (Fig. 5-I-1,2,3, 5-II-b), and the environment (Fig. 5-I-4). In the figure, each column indicates the interaction of infrared signals among them. The viewer allows us to choose our preferred duration of the interaction for watching the video. The numbered circles correspond to scenes of the numbered video scene shown in the right-side.

For example, the circle numbered *1* is an interaction between the robot and a visitor *B*, which is also shown as the detailed view (Fig. 5-II). The leftmost column (a) of figure (II) indicates whether the IR-tracker attached to the robot captured the signal from visitor *B* (marked as a darker color, originally red). Column (b) indicates that the LED-tag of the robot was found by the others (originally colored blue). If more than one IR-tracker captures the signal from the robot's LED-tag, the other columns will also appears, corresponding to the other IR-trackers. Column (c) indicates the behavior of the robot. When the user of the tool selects a behavior (marked as darker bar, originally green), it shows the contents of the behavior such as "who=263, what=0, id=167, name=CNAME" (Fig. 5-I-5, 5-II). CNAME (or *Call name*) is the name of the behavior the robot performed, and the other information is associated with the behavior.

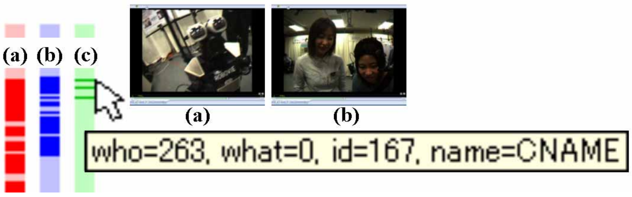
Moreover, we can switch the main view according to the preferred target for analysis, such as a human, the robot, or a certain executed behavior of the robot. Thus, it is easy to find the scene of the interaction. For instance, it provides the scenes when the robot is pointing at an object to guide a human so that we can easily watch the response of visitors during the behavior.

III. IMPLEMENTATION OF GUIDING BEHAVIORS

The ubiquitous sensors provide us the information of each visitor's history of activity and the current state in the room. Based on these data, the robot selects a topic and talks about this topic with the visitor. Each topic has been implemented as a *situated module* that follows implemented *episode rules.*



(I) Main view



(II) Detailed view

Figure 5: *Corpus viewer*
(tool for analyzing captured *interaction corpus*)

A. Topics that the robot talks about

The following describes the topics that Robovie talks about based on information of the ubiquitous sensors. In the actual experiment, the guide robot spoke about the topics in Japanese. Each italic string within the parenthesis represents the ID of the situated module that corresponds to the topic.

Call name

Robovie calls the name of visitors (*Call name*), such as "Hello, Mr. Koide," "Mr. Koide, you've come here again, haven't you?" (for the person who started to interact with the robot again).

Booth recomendation

Robovie talks about the booth where the visitor has visited (*Visited booth*), the booth which the visitor seems to be interested in (*Interesting booth*), the booth where the system wants to recommend for a visit (*Booth recomendation*), the booth where there are many visitors at the time (*Crowded booth*) or that is popular for the visitors (*Popular booth*). For instance, it says comments such as "You have visited *X* booth," and "I recommend the *X* booth," and "There are many visitors at *X* booth." Each user's interest about an exhibition booth is estimated from his/her consumption time at the booth.

About the other visitors

Robovie talks about the other visitor who have talked with the interacting visitor (*Talked person*), the visiting histories of the other visitor (*Others' experience*), and the other visitor who has interests similar to those of the interacting visitor (*Similar interest person*). For instance, it

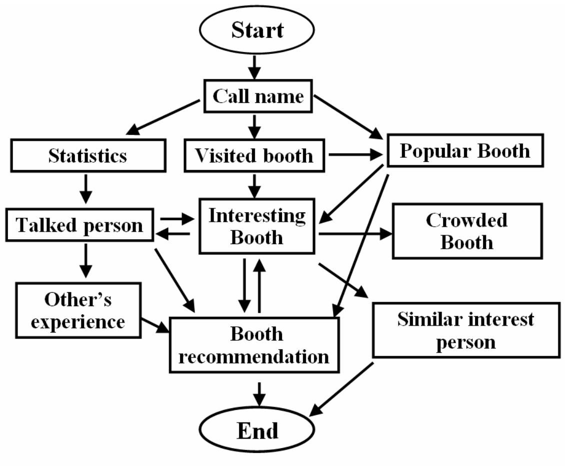


Figure 7: Basic flow of topics

says ,"You have met Mr. Koide," "Mr. Koide has visited *X* booth, *Y* booth, and …," and "Mr. Koide is as interested in *X* booth as you are."

Statistics of the exhibition

Robovie talks about the statistics of the exhibition, such as "Now, there are 17 people in this room," and "50 visitors have visited this room today."

B. Episode rules to generate natural flow of topics

To let humans feel that the communication is natural, there are two approaches to decide the order of the topics provided by the robot. One is to imitate a human guide's behavior, to make a model of their communication, and to implement it. This ethological approach has worked successfully for pet robots such as Aibo. However, humans are too complicated to implement an interactive robot using ethological approach; on the other hand, the developed robot has some superior abilities that humans do not have.

Instead of the above strategy, we choose the engineering approach in a trial-and-error manner. The episode rules are simple rules that can handle more complicated transitions than a simple state transition flow [8]. **Figure 7** represents the principal transitions among the topics (*situated modules*). Some episode rules are designed to avoid repetition of similar topics for the same visitor. There are also episode rules to avoid showing the behaviors that it performed recently, because the current interacting visitor might have observed the behavior already. However, *Call name* is an exception to such limitations, since every visitor seems to prefer it. These rules have been established during development with trial-and-error iterations, including preliminary experiments.

IV. EXPERIMENT

We conducted a two-day experiment at a research exhibition to evaluate the developed robot. The interaction between the robot and visitors were captured as the *interaction corpus* and analyzed with the *corpus viewer* later.

Table 1: Exhibited behaviors during the experiment

Name	No. of exec.
Call name	267
Statistics	67
Talked person	69
Other's experience	27
Visited booth	138
Interesting booth	64
Recommended booth	96
Popular booth	13
Crowded booth	13
Similar interest person	18

A. Experiment settings

The experiment was performed during our research institute's annual research exhibition for two days. There were five research exhibition booths in the room, and the robot was one of them. At each of the booths, one researcher was on duty to give guidance on his/her research. At the robot booth, Robovie talked with the visitors by using the topics described in the previous section, such as explaining the other booths. Meanwhile, a human explained the robot and this research to visitors aside the robot interaction. The researchers at the five booths, including robot booth, prepared a questionnaire and asked the visitors to fill it out when exiting the room.

B. Statistics

More than 300 visitors came to our ubiquitous sensor room during the two days. The exhibition was free without any restriction so that various people came, such as researchers, reporters, university students, and residents near the institute. There were 171 participants who wore the wearable sensor set or LED-tag badge. In detail, there were 105 wearable sensor set users and 78 LED-tag badge users, while some of them used both the wearable sensor set and the LED-tag badge.

Robovie interacted with 108 unique visitors who wore the sensors, and it started the interaction 268 times in total. **Table 1** indicates the number of behaviors it exhibited. An analysis of the detailed interaction is reported in the next section.

There were 79 of the 171 participants answering the questionnaire. Of these 32 visitors (40.5%) chose the robot booth as the most interesting at the exhibition, and 26 visitors (32.9%) indicated that they made the longest stay at the robot booth. They also rated the familiarity of the robot on a 1-to-7 scale (7 is the highest), which resulted in an average of 4.3.

C. Analysis of interaction by using the Corpus viewer

We analyzed the interaction between the robot and visitors by using the *Corpus viewer*. In particular, we focused on the interaction sequence from the *Call name* behavior through to the *Booth recommendation* behavior. As shown in **Table 1**, *Booth recommendation* behavior were executed 96 times during the experiment; however, the majority of the execution was for the human guide

(a) before (b) during robot's utterance (c) after

Figure 8: A visitor's reaction for *Call name* behavior

(a) *Call name* (b) *Statistics* (c)*Booth recommendation*

Figure 9: Comparison of reactions among topics

Table 2: Reactions for *Booth recommendation*

Type of reaction	No. of visitors
Went to the recommended booth	4
Looked at the direction, but did not go	5
No reaction for the recommendation	4

(researchers who belong to the same institute as the authors) or for nobody (the interacting visitor went away before the behavior sequence ended), so we excluded these samples from the analysis. As a result, 13 video sequences of the interaction that includes both *Call name* behavior and *Booth recommendation* behavior were analyzed in detail.

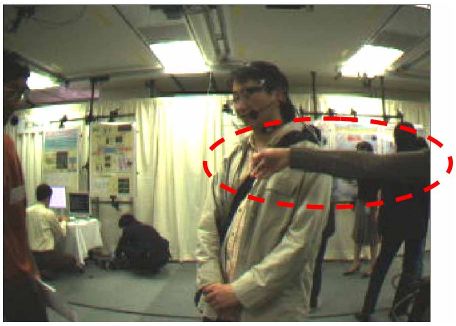
Reaction for Call name behavior

Figure 8 is an example of visitors' reactions for *Call name* behavior. The leftmost picture (retrieved from the video data in the *interaction corpus*) is a scene of the interaction before it called his name, the middle one is a scene while it called his name, and the right one is a scene just after the behavior. We found that all of 13 visitors gave some reaction, such as gazing with its eye while smiling and talking in response. Moreover, we roughly compared visitors' reaction for each topic. **Figure 9** is a typical sequence of reactions. Although most of them gave a positive reaction to *Call name* behavior, they seemed to just hear the robot's talk and guidance for the other topics without giving explicit responses, except for the *Booth recommendation* behavior.

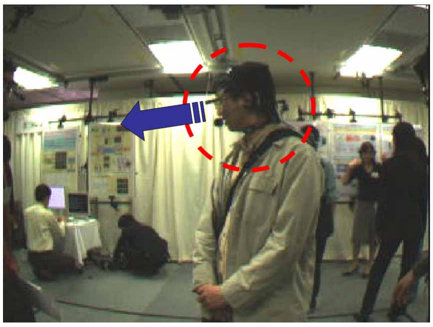
Reaction for Booth recommendation behavior

We also found many reactions for *Booth recommendation* behavior. The robot utilizes its arm to point in the direction of the recommended booth during only the *Booth recommendation* behavior; contrary, it just talked and gazed at the human face during the other behaviors. **Table 2** shows the reactions of visitors. Four visitors went to the recommended booth immediately or soon after the interaction with the robot. In addition, five visitors instantly looked at the direction of the recommended booth, but did not go to the booth. They did

(a) Interacting visitor (right) and other visitor observing the interaction (left)



(b) Interacting visitor also pointed in the direction the robot pointed in. (*Booth recommendation* behavior)



(c) Observing visitor also reacted.

(d) Interacting visitor went to recommended booth.

Figure 10: Interaction with multiple visitors.

not give such a reaction to the other behaviors. This result shows that the robot's embodied action caused such a difference in reactions.

Furthermore, the *Corpus viewer* enabled us to observe interesting scenes where the robot was involved in inter-human communication. **Figure 10** is a scene of the interaction with multiple visitors. Since the robot was designed to interact with one person, it identified multiple IDs but only spoke to the right-side visitor in the picture. When the robot pointed at the booth (second picture in the figure), the interacting visitor also pointed in the same directions in response to it. As shown in the third picture, the observing visitor gazed in the pointed direction by both the robot and the interacting visitor, which caused another inter-human communication as shown in the fourth picture.

D. Discussions

The experimental results revealed that the *Call name* behavior made possible by the ubiquitous sensors attracted visitors' attention. It seems appropriate to start the interaction with this behavior. Furthermore, in the *Booth recommendation* behavior, the robot successfully utilized its body properties to guide humans. Moreover, this example of interaction with multiple visitors seems to suggest the potential ability of the guide robot. That is, the robot may promote inter-human communication around it. It is one of our important future tasks to verify the effect of interactive behaviors that explicitly encourage such an inter-human communication. Although the implemented task of the guide robot was simple, the experimental results verified that this approach is promising for developing an interactive humanoid guide robot.

Contrary to these positive findings, the detailed guidance information such as *statistics* and *talked person* did not make a positive impression on the visitors. One difficulty seems to be the limited size of the exhibition. Since there were only five exhibition booths, the visitors probably did not need to acquire the information about the other booths from the robot, although it was sufficient for an experimental trial. They seemed interested in the novel robotics devices rather than the topics that it provided. Another difficulty is the recognition of social relationships. There was no mechanism to identify specific relationships, so the robot just talked about other humans who seemed to have similar interests; however, the visitor usually could not identify the person whom the robot talked about. Since recent ubiquitous sensor technology is gradually making it possible for a robot to recognize human's social role in communication [9] and friendly relationships [10], we believe that this problem will be solved in the not-so-distant future.

V. CONCLUSION

This paper reported an approach to integrating an interactive guide robot with ubiquitous sensors. The ubiquitous sensor provides the robot with sufficient sensory information to personalize guidance. At the same time, it enables us to analyze the human-robot communication in detail. That is, the ubiquitous sensors support both development and evaluation, which are mutually indispensable. We conducted a two-day experiment where the robot guided over 170 visitors. The captured interaction corpus reveals that the robot successfully attracts the visitors' attention with its human-like body properties. The findings from the experiment verify the appropriateness of our approach for an interactive humanoid guide robot. Our important future work will involve applying this fundamental development methodology to more general and complicated tasks of interactive humanoid robots.

ACKNOWLEDGEMENT

We wish to thank Yusuke Murakami of Kyoto University for his support in system development. We also wish to thank Dr. Norihiro Hagita, Dr. Kenji Mase, and Dr. Yasuhiro Katagiri at ATR for their helpful advice.

REFERENCES

[1] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, Development and Evaluation of an Interactive Humanoid Robot "Robovie", *Int. Conf. on Robotics and Automation (ICRA 2002),* pp. 1848-1855, 2002.

[2] T. Kanda, H. Ishiguro, and T. Ishida, Psychological analysis on human-robot interaction, *Int. Conf. on Robotics and Automation (ICRA 2001)*, pp.4166-4173, 2001.

[3] T. Ogata and S. Sugano, Emotional Communication Between Humans and the Autonomous Robot Which Has the Emotion Model, *IEEE Int. Conf. on Robotics and Automation (ICRA'99),* pp.3177-3182, 1999.

[4] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, Body Movement Analysis of Human-Robot Interaction, *Int. Joint Conf. on Artificial Intelligence (IJCAI 2003)*, pp.177-182, 2003.

[5] Y. Sumi, T. Matsuguchi, S. Ito, S. Fels, and K. Mase, Collaborative capturing of interactions by multiple sensors, *Int. Conf. on Ubiquitous Computing (Ubicomp 2003)*, pp. 193-194, 2003.

[6] N. Hagita, K. Kogure, K. Mase, and Y. Sumi, Collaborative capturing of experiences with ubiquitous sensors and communication robots, *Int. Conf. on Robotics and Automation (ICRA 2003)*, September, 2003.

[7] M. Bono, N. Suzuki, and Y. Katagiri, An analysis of participation structure in conversation based on Interaction Corpus of ubiquitous sensor data, *Int. Conf. on Human-Computer Interaction (INTERACT 03)*, pp. 713-716, 2003.

[8] T. Kanda, H. Ishiguro, M. Imai, T. Ono, and K. Mase, A constructive approach for developing interactive humanoid robots, *Int. Conf. on Intelligent Robots and Systems (IROS2002), 2002*

[9] M. Bono, N. Suzuki, and Y. Katagiri, An analysis of non-verbal cues for turn-taking through observation of speaker behaviors, *Joint Int. Conf. on Cognitive Science (CogSci)*, 2003.

[10] T. Kanda and H. Ishiguro, Reading human relationships from their interaction with an interactive humanoid robot, *Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE),* 2004. (to appear)