Noname manuscript No.
(will be inserted by the editor)

Field Trial of a Networked Robot at a Train Station

Masahiro Shiomi, Daisuke Sakamoto, Takayuki Kanda, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita,

Received: date / Accepted: date

Abstract We developed a networked robot system in which ubiquitous sensors support robot sensing and a human operator processes the robot's decisions during interaction. To achieve semi-autonomous operation for a communication robot functioning in real environments, we developed an operator-requesting mechanism that enables the robot to detect situations that it cannot handle autonomously. Therefore, a human operator helps by assuming control with minimum effort. The robot system consists of a humanoid robot, floor sensors, cameras, and a sound-level meter. For helping people in real environments, we implemented such basic communicative behaviors as greetings and route guidance in the robot and conducted a field trial at a train station to investigate the robot system's effectiveness. The results attest to the high acceptability of the robot system in a public space and also show that the operator-requesting mechanism correctly requested help in 84.7% of the necessary situations; the operator only had to control 25% of the experiment time in the semi-autonomous mode with a robot system that successfully guided 68% of the visitors.

Keywords Networked robot, Semi-autonomous communication robot, Field trial

1 Introduction

One of our goals is to develop a communication robot that is capable of natural human-robot interaction and can support human activities in real environments. For example, in the future, a communication robot at a train station might provide information about departure platforms, transfers, and nearby shops by effectively using both verbal and nonverbal expressions (Fig. 1). Since the target audience of a communication robot is the general public (people without specialized computing and engineering knowledge), a conversational interface that uses both verbal and nonverbal expressions is important. Past studies in robotics have emphasized the merits of robot embodiments that show the effectiveness of facial expressions [1], eye-gaze [2], and gestures [3].

However, it remains difficult to achieve robust verbal communication with communication robots. One major difficulty is the speech recognition of colloquial utterances in noisy environments. Such disturbances increase the dependence on the distance between the robot's microphone and the speaker. Recent speech recognition technology can only recognize formal utterances in noiseless environments. Although research continues in robot audition [4], the difficulties in real environments remain beyond the grasp of current technology. Therefore, robots continue to have difficulty handling natural language conversations as deftly as humans.

Another basic difficulty exists in the development process for communication robots in real fields. They must be placed in real situations; otherwise, we cannot reproduce similar situations in laboratories or elsewhere. Human behavior in real environments is too complicated to predict and becomes even more complex with a large number of various people, such as children,

Masahiro Shiomi, Takayuki Kanda, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita
ATR-IRC, Kyoto, 619-0288, Japan
Tel.: +81-774-95-1432
Fax: +81-774-95-1408
E-mail: m-shiomi@atr.jp

Daisuke Sakamoto
Japan Science and Technology Agency, ERATO Igarashi Design UI Project, 1-28-1-7F Koishikawa, Bunkyo, Tokyo, 112-0002, Japan

adults, and senior citizens, or if the environment is more intricate.

We must improve the sensing abilities of robots in such real environments so that they can work robustly. Some researchers are studying stand-alone robots that have complete sensing, decision making, and acting capabilities. On the other hand, if current technologies are used, such approaches have limitations to increase the abilities of robots in real environments.

To solve these problems, we have chosen a strategy known as a "network robot system"[5] that combines robots, ubiquitous sensors, and humans. In this strategy, a human operator supports the decisions of a robot during interaction. In other words, a robot behaves semi-autonomously [6]. Semi-autonomous communication robots can achieve useful tasks in real fields supported by a human operator.

Inspired by Norman's human model (p. 28 in [7]), we developed a model of a semi-autonomous communication robot that consists of three layers: visceral, behavioral, and reflective. The visceral layer corresponds to involuntary actions that can be done by such simple creatures as lizards. The behavioral layer corresponds to the subconscious behavior of mammals obtained through repeated training. The behavior from these two layers is unconsciously absorbed. Thus, humans can consider (reflective layer) how to behave even when taking actions with these two layers, such as walking, driving, and so forth.

The reflective layer is operated by humans (both operators and developers) in our model. In the beginning, most of the other parts will also be performed by operators, the reactive layer can be prepared from the beginning, and the behavioral parts will be gradually replaced with software modules by developers. Therefore, most language communication will be continually managed by human operators, and new behaviors will be continually supplied by human developers.

This paper reports a semi-autonomous robot system that makes two major contributions. First, it demonstrates how ubiquitous sensors and an operator contribute to achieve the semi-autonomy of a robot. In other words, if we permit human operators, vast potential exists for robotics technology to further contribute toward human-like communication services. Second, it demonstrates how people interact with guide robots in a train station, indicates how effectively communication robots support human activities in a real environment, and provides some design implications for such communication robots.

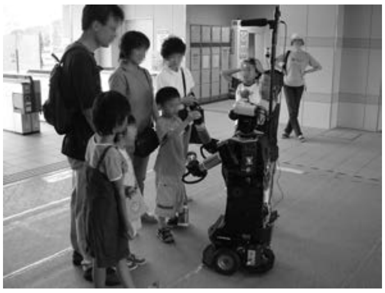


Fig. 1 Route guidance at a train station

2 Related Works

2.1 Autonomous Approach

Many past works focused on a robot that acts in everyday environments frequented by ordinary people [9–14]. For instance, Burgard et al. developed a tour guide robot [9] with robust navigational skills that behaved as a museum tool. Siegwart et al. developed a robot that guided people in large-scale environments [10]. Bauer et al. realized a robust navigating robot under an unknown urban environment using GPS data, cameras, laser range finders, and interactions with people [12]. Some researchers developed robots to support people in daily environments such as shops [13, 14]. These cases indicate that autonomous robots are already feasible for delivering pre-defined messages associated with locations, particularly at places with many people who have a special interest in robots, such as museums and world expos.

However, the inputs for these robots are limited; although buttons and tactile sensors were robustly used, these robots did not exploit natural language, which largely limited their capabilities.

2.2 Semi-Autonomous Approach

A human operator is often used to simulate the missing components of a system under development and to observe people's reactions to such a nascent system. This is known as the Wizard of Oz (WOZ) method [15, 16] in human-computer interfaces. Some studies have used WOZ techniques for human-robot interaction, although these prototypes demonstrated little autonomy. For example, Woods et al. used a tele-operated robot that approached people to observe their reactions [17]. Green et al. also used a tele-operated robot in a living room setting to determine what services people needed [18].

However, it is difficult for a human operator to be completely responsible for a robot's functionality. Our semi-autonomous approach's thrust resembles the WOZ method: a human operator with a prototype system gathers realistic data from users. On the other hand, an important difference is that we are trying to separate the two major components, the reflective layer and the behavioral/visceral layer, and to automate the latter as much as possible. Since we do not intend to immediately make our system autonomous, we assume that the system can request help from human operators. At the same time, we are trying to minimize operator support by focusing on autonomy in the non-language part, which will probably result in situations where a few operators can control hundreds of robots. In robotics, many studies have investigated the teleoperation of mobile robots, arm robots, pet-type robots [19], and even humanoids [20]. In particular, space exploration and similar domains require distant teleoperation that causes communication delay. Thus, several studies have utilized partial autonomy in teleoperation, such as obstacle avoidance and goal-directed locomotion [21–23].

Unfortunately, these studies did not focus on natural human-robot interaction, which apparently requires more complicated collaboration between an automated system and human operators.

3 System Configuration

Figure 2 shows a system overview that consists of ubiquitous sensors and three layers: reflective, behavioral, and reactive. The basic software design follows the architecture of a communication robot [8](reactive and behavioral layers). The system also has an operator-requesting mechanism that autonomously requests help from the human operator.

Environmental sensor data are sent to the robot by a 802.11 b/g wireless network. The robot uses this sensor information to move and interact with people. The robot sensor data and the environmental sensor data are sent to the operator to support the robot by the same network; basically the operator uses image data from the environmental cameras and sound information from the robot. The details of each system are described as follows.

3.1 Sensor and Actuator

3.1.1 Robovie

Figure 3 shows “Robovie,” our interactive humanoid robot that is characterized by its human-like physical expressions and its various sensors [8]. It has a head, two arms, a body, and a wheeled-type mobile base. Its height and weight are 120 cm and 40 kg. The robot has the following degrees of freedom (DOFs): two for its wheels, three for its neck, and four for each arm. Its lower mobile base is a Pioneer 3-DX (ActiveMedia). We used corpus-based speech synthesis [24] for generating speech. Robovie can work one hour without being recharged. To communicate with other sensors and an operator, it uses a 802.11b/g wireless network.

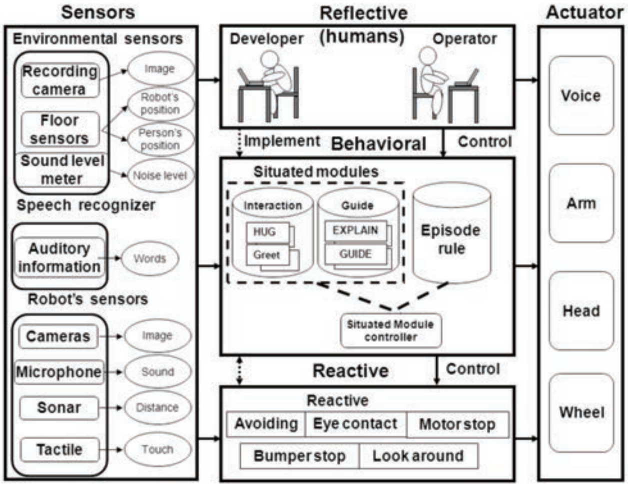


Fig. 2 System overview

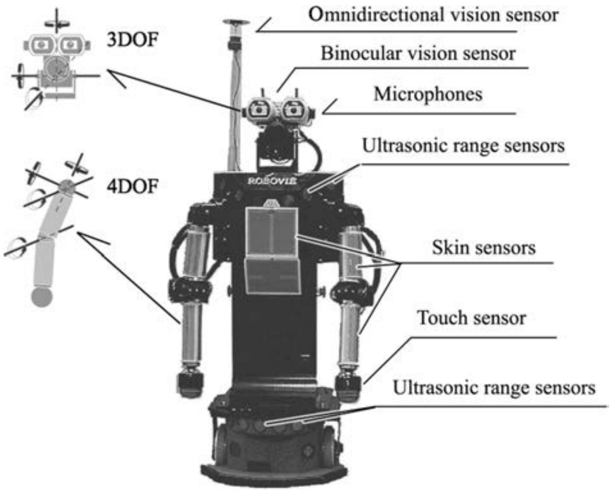


Fig. 3 Robovie

3.1.2 Floor Sensors

To detect the positions of people, we used an external remote PC and floor sensors because they can collect high-resolution data and are robust to occlusion. We installed 128 floor sensors units, VS-SS-F (Vstone Corporation, Osaka), around the robot that covered a 4 x 8 m area. Each sensor unit is 500 [mm^2] with a resolution of 100 [mm^2]. Their output is 1 or 0; the floor

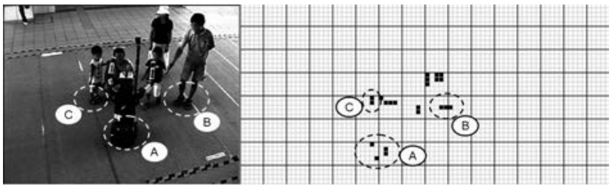


Fig. 4 Multiple people on floor sensors

sensor is either detecting pressure or it is not. Therefore, 25 binary data were acquired by one floor sensor. Floor sensors are connected with each other through an RS-232C interface at a sampling frequency of 5 Hz. Fig. 4 shows an interaction scene between the robot and people (left) and an example of outputs from the floor sensors (right). A black point indicates that a sensor detected pressure from a robot or a person.

We used a Markov Chain Monte Carlo method and a bipedal model to estimate people's positions on the floor sensors [25] since it is one robust sensing method for positions. It is free from occlusion, and the average position error is less than 21 cm. Thus, it is useful when a person interacts closely with the robot.

The floor sensors enable us to achieve three crucial functions for robot autonomy. The first is an approaching function with which the robot can approach a detected person to start an interaction. For this purpose, the floor sensor system estimates and sends the x-y coordinate position information of people who exist on the sensors using floor sensor data to the robot by network.

The second is a pointing gesture. The interacting person's position is used to generate a pointing gesture for route-guidance. The interacting person's distance is also used to point at the destination by the robot. For this purpose, the floor sensor system calculates the distance information between the robot and the nearest person using its position information.

The third function is the robot's position compensation. Although it has an odometer to estimate its position, since this devise is affected by wheel slip, it is not very accurate. The system continues to track the robot's position, and such information enables the robot system to maintain position accuracy. For this purpose, the robot sends speed and odometry information to the floor sensor system, which estimates the robot's position with this information and sends the estimated position to the robot by network.

3.1.3 Speech Recognizer

For the speech recognition function, we prepared an external remote PC to which the robot sends audio input from its directional microphone to achieve fast speech recognition using a software application for automatic speech recognition robust to noisy environments and speaker variability (age and gender) [26]. In the front-end block, environmental noise is first reduced by a feature-space noise suppression method using clean speech Gaussian mixture models and Wiener filters. The speech recognition block is comprised of two parallel decoders that correspond to adult and child speech, and each decoder uses two phoneme acoustic models for male and female speech. Four different levels of signal-to-noise ratios were also implemented in the acoustic models to improve robustness against noise. For the speech recognition block's language model, we implemented a dictionary of about 100 words, including place names and greetings, and a simple grammar that imposed rules for connecting words in the dictionary. Finally, the speech recognition results were sent to the robot by network.

3.1.4 Sound-Level Meter

We installed a sound-level meter with an external PC to measure the environment's noise level and to send it to the robot to autonomously change its volume. For example, the robot increases its volume when the noise level exceeds 70 [db] and decreases it when the noise level is less than 65 [db]. These thresholds were decided beforehand based on the environment's noise level.

3.1.5 Tele-operation Interface

We developed an operator interface for controlling the robot's behaviors (Fig. 5) that consists of two windows: controlling robot and informing the operator. The left window is an interface for choosing all the needed situated modules of the robot, such as greetings, offering route guidance, explaining several destinations, and saying goodbye. The right window uses a sphere image to inform the operator that he/she needs to assume control. This interface changes to red when the robot requests the operator's help. With the software the operator can also control such low-level functions as wheel speed, the position of each joint, and utterances; but due to the time delay caused by excessive input, the operator rarely uses them.

The operator uses vision information from six cameras through cables and the auditory information from the robot's microphone transmitted through the wireless network. This information is available in real time without delay.

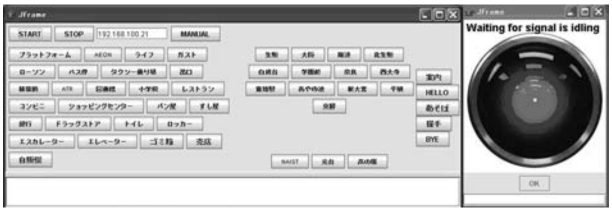


Fig. 5 Control software for human operator

3.2 Reactive Layer

The conceptual purpose of the reactive layer is to achieve safe interaction with lifelike behavior. For lifelike behaviors, the robot controls eye movements based on output from its touch sensors to exhibit lifelikeness. For safety, the robot's locomotion and motors stop when an object contacts its bumper or overload of any motor is detected. These reactive behaviors were prepared for a general environment. Therefore, we only implemented simple mechanisms in the reactive layer that work correctly and do not require software updating.

3.3 Behavioral Layer

The conceptual purpose of the behavioral layer is to achieve task/environment dependent behaviors. The behavioral layer consists of situated modules, a situated module controller, and episode rules. The situated modules allow the robot's interactive behavior with situation dependent sensory data processing to recognize reactions from humans. Because each module works in a particular situation, developers can easily implement situated modules by only considering the particular limited situation. A situated module is implemented by directly coupling the communicative sensory-motor units with others to supplement such sensory-motor units as utterances and gestures.

Episode rules describe the state transition rules among the situated modules. The robot can autonomously interact with people with the behavioral layer. We implemented two types of situated modules: guidance behavior and greeting and free-play behaviors.

3.3.1 Situated Modules

In the beginning, the robot approaches a person detected by the floor sensors and initiates interaction by greeting and offering a handshake. If the person requests directions, the robot immediately starts guidance behaviors if it correctly recognized the utterance. Adults often seek such information. The robot is also capable of free-play behavior that is popular with children. The robot sometimes triggers tactile interaction with a child by saying, "Let's play a game called touch." After the child reacts, it continues performing free-play behaviors as long as the child responds and initiates brief small talk, such as "Where are you from?" It also offers children a hug. The robot also offers such information around the station as, "There is a new shopping center close to the station." After it exhibits several free-play behaviors, it initiates guiding behavior. At the end of the interaction, the robot exhibits goodbye behavior.

For the situated modules for guiding, our robot can guide visitors to 38 nearby places by asking, "Where are you going?" If an interacting person responds, the robot starts to provide directions. For example, when guiding a visitor to the bus stop, the robot points toward the exit and says, "Please go out this exit and turn right. You'll immediately see the bus stop." When the robot explains the route, it utilizes a pointing gesture as well as such reference terms as "this" and "that." Since Japanese has three types of reference terms associated with positional relationships between two interacting persons and the object being discussed, we installed a "three-layer attention-drawing model" for these reference terms and pointing gestures [3]. Thus, the robot autonomously generates behavior to guide visitors to these destinations using appropriately chosen reference terms and gestures. It also has a map for these locations. If the interacting person cannot directly see the destination, such as a place outside the station, the robot points to a visible place, such as the exit, and verbally supplements the remaining directions.

3.3.2 Episode Rules

The relationships among behaviors are implemented as rule-governing execution orders called episode rules to maintain a consistent context for communication. Their basic structure consists of previous behaviors (e.g., successfully finished greeting behaviors) and subsequent behaviors (e.g., offering route-guidance behavior). The situated module controller selects a situated module based on the 1311 implemented episode rules. As described above, episode rules are designed to achieve the following six kinds of behaviors: approaching a visitor, extending a greeting, offering small talk, providing information around the station and free-play, offering route guidance, and saying goodbye. Moreover, an event-driven transition was described so that when a passenger initiates a route-guidance conversation, the robot begins to offer route guidance.

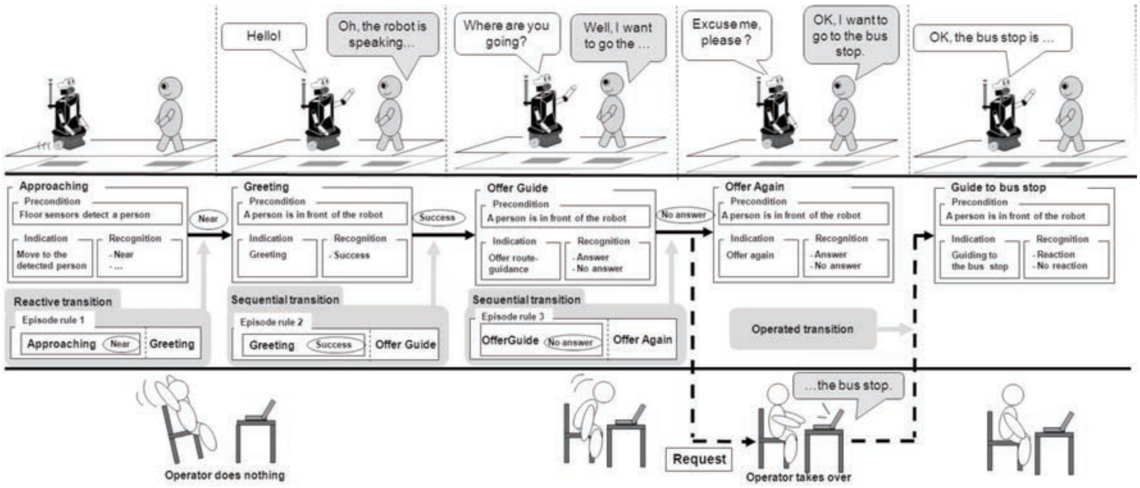


Fig. 6 Illustration of interaction scenes with operator

Figure 6 shows one example of the episode rules and interaction scenes between a robot and a person. After the floor sensors detected the person's position, the robot approached the person with the approaching situated module. This caused a reactive transition governed by episode rule 1. The robot greets the person by executing the situated module Greeting. After the Greeting, the robot provides route guidance by executing the Offer Guide. In this example, because the visitor does not respond or the speech recognizer fails to detect what the visitor said, the Offer Guide results in a No answer, and the robot asks the visitor again using Offer Again. At the same time, the operator-requesting mechanism (described below in 3.4) fires so that the operator is asked to assume control. The visitor might answer the robot with such a response as, " I'd like to go to the bus stop, "which is heard by the operator. As a result of the operator's control, Guide to the bus stop is finally selected.

3.4 Reflective Layer

The conceptual purpose of the reflective layer is to integrate an autonomous robot system with humans so that the system as a whole can process natural language, think deeply, and improve with human support. The robot can autonomously operate without the reflective layer. In addition, using the reflective layer, the system requests help from the human operator and starts to work autonomously when an interaction between the robot and the target visitor is finished. In this system, two types of information are used for the mechanism: the sub-system's report and behavior transition.

3.4.1 Reports from Sensors and Actuators

Each sensor and actuator can individually report problems to request operator assistance. For example, in the reactive layer, the robot stops its body movements and locomotion when an actuator detects a motor overload. The robot also stops its body movements and locomotion when a tactile sensor is continuously reacting more than five seconds. Another mechanism detects interaction level errors. The speech-recognition module monitors several negative phrases, such as "I don't understand," "That's not right," and "That isn't what I asked." Such statements indicate a problem at the level of human-robot interaction.

3.4.2 Behavior Transition

The second mechanism detects interaction-level problems. For example, if the interacting people find that the robot does not answer correctly, most of the time they simply leave without complaining. Such interaction-level problems reflect the service contents. Therefore, in this study, we focused on situations in which a robot guides people to develop this mechanism.

Episode rules also monitor behavior transitions, and the robot system requests the operator's help when a transition pattern matches pre-defined situations whose details for trapping apparent errors are described below.

a) When the robot cannot hear anything in particular from a visitor after twice offering route guidance

This episode rule refers to the following situations: (a) an interacting person does not speak because he/she is not interested in route guidance, which is outside the range of the implemented interaction model, or (b) the speech recognition module fails. If the route-guiding module twice gets a no-answer, the system requests the operator's help.

b) When the robot continues the interactive behavior more than ten times without performing the route-guiding module

The robot usually exhibits the route-guiding module if people follow the ordinary interaction flows. On the other hand, people can cause a different interaction flow, for example, by continuously ignoring the robot's handshake request and initiating interaction by touching its shoulder.

c) When the robot continuously offers route guidance three or more times

The robot exhibits route-guidance behavior when it recognizes such a spoken request as "please give me some directions." This episode rule refers to the following situations: (a) an interacting person is greatly interested in route-guidance behavior or (b) the robot's route guidance continuously fails, which results in continuous requests for route guidance.

4 Field Trial at a Train Station

4.1 Environment

The six-day experiment was conducted at a terminal station of a railway line that connects residential districts with the city center with four to seven trains per hour. Fig. 7 shows the experiment's environment. Most users descend the stairs from the platform after exiting trains. We set the robot and the sensors in front of the right stairway (Fig. 7) and informed the visitors that the robot can provide directions. As shown in Fig. 7, we placed floor sensors in the center of a 4 * 8 [m] floor area and six cameras on the ceiling. The sound-level meter with an external PC was installed to the right under camera 'F'. The robot moved on the floor sensors.

We recorded all sensory data including the data from the floor sensors, video images from the ceiling cameras and the robot's camera, and the auditory information from the robot's microphone. We used these data to investigate the effectiveness of our robot system. We also received permission to record these video and auditory data from the train station authorities and placed posters in the station to inform visitors.

4.2 Participants

The station visitors were mainly commuters and students, and on weekends families visited the station to see the robot. The visitors could freely interact with our robot. We asked those who interacted with the robot to complete questionnaires after their interaction. Children were asked to fill out questionnaires if they understood their purpose.

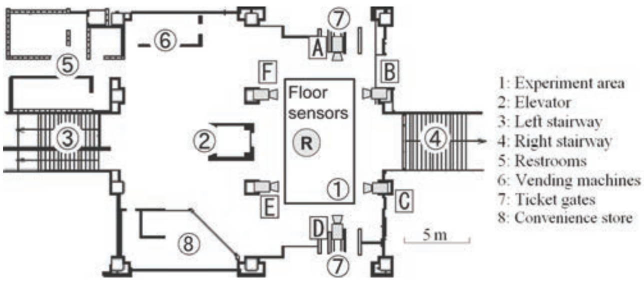


Fig. 7 Station map

4.3 Conditions

We prepared an autonomous condition to reveal how much the robot can do without human help to investigate how completely the autonomous robot can support people in a real environment. We did not prepare a full-operated condition because this study's main purpose is to investigate how the developed system supports autonomous robots by minimizing the operator's load. In the experiments, we prepared several time slots and counterbalanced their order.

Autonomous condition: The robot was completely autonomous and did not use the functions in the reflective layer; the operator never assumed control.

Semi-autonomous condition: The robot used all the implemented layers: reactive, behavioral, and reflective. It usually operated autonomously when there were no visitors, so the operator just monitored the situation without taking control of the robot.

As described in Section 3.5.1, the operator only controlled the robot's behavior and worked with the speech recognition function when the operator-requesting mechanism detected a need for operator help. The operator did not monitor interactions unless the robot asked for help so that we could observe fairly well whether this semi-autonomous mechanism works. The operator finished its control, and the system started to work autonomously when interaction between the robot and the target visitor finished. The operator received more than two hours of training for controlling the robot.

5 Results

This section reports the results from two major viewpoints: technical achievements and attitudes toward the robot. The former consists of system performance, the success rate of the route guidance, the operating time, the performance of the operator-requesting mechanism, and the success rate of speech recognition. The latter consists of how visitors interacted with the robot and questionnaires of subjective impressions.

5.1 Technical Achievements

5.1.1 System Performance

The robot system worked quite well under both conditions (Fig. 8). Based on the position information from the floor sensors, it autonomously approached and interacted with 168 people during the trials: 77 people in the autonomous condition and 91 people in the semi-autonomous condition. When the robot provided route guidance, it correctly pointed with gestures calculated by the position relationships between the visitor and itself.

Note that the sensors robustly estimated the positions of people and the robots in crowded situations; for example, even though many occlusions were caused by the presence of over 15 people, no collisions occurred (Fig. 9). The robots never caused a dangerous situation (e.g., with small children or senior citizens).

In the semi-autonomous condition, the operator also worked well as a speech recognition function for the robot in the trials. For example, during crowded and noisy situations, the robot smoothly talked with the visitors with operator support.

5.1.2 Task Performance

For each visitor who responded to the robot's offer of route-guidance or asked about a route, we assume that the robot successfully gave directions if it correctly offered one or more route-guidance directions. Thus, even when a visitor asked for routes to more than one place, the guidance was judged a success.

In the autonomous condition, the success rate was 29.9% among 77 visitors. The main cause of failure was speech recognition error. In the semi-autonomous condition, the success rate was 68.1% among 91 visitors. 31.9% failure remained, mainly due to visitors who stopped interacting with the robot before the operator assumed control. In addition, speech recognition failure sometimes caused a breakdown of the operator-request actions. If the speech recognizer simply failed to detect the speech or the recognition result was rejected because it did not match the pre-assumed model (this often happened), the operator was successfully requested. The problem was when the speech recognizer picked up a false positive result, which resulted in mistaken guidance even though the system had not detected the situation as problematic.

5.1.3 Operating Time

The experimental time was 45,900 seconds, the overall interaction time was 20,078 seconds, and the overall idling time was 25,822 seconds. The robot autonomously interacted with people for 8,551 seconds; the operating time (when the operator controlled the robot) was 11,527 seconds. Thus, the operator controlled the robot 25% (11,527/45,900) of the experimental time in the semi-autonomous condition.

A tradeoff exists between task performance and operating time. That is, higher operating time results in better task performance, but it also requires more elaborate operator control. In our case, we designed the system to minimize operator time; if the operator controlled the robot from the beginning of the interaction, task performance would increase.

5.1.4 Performance of Operator-requesting Mechanism

"Operator needed" situations

First, we evaluated the performance of the operator-requesting mechanism when the " operator is needed, " defined as a situation where a person silently looked at the robot for 10 seconds after it talked to the person or where a person asked the robot twice for a route-guidance. We measured such cases: the operator-requesting mechanism called the operator within 10 seconds, 20 seconds, or the end of interaction when a " operator is needed " situation happened.

The reason for the definition of an " operator is needed "situation depends on the observed interactions between the robot and visitors, as described in Section 5.2.1. We often observed people repeatedly for route-guidance in the autonomous condition; they observed the robot and talked to it for more than 20 seconds, even though the robot did not react to them. In addition, some children silently looked at the robot more than 10 seconds after the robot addressed them.

"Operator is needed" situations happened 85 times within interactions with 91 subjects. The operator-requesting mechanism called the operator 32 times (37.6%) within 10 seconds, 61 times (72.9%) within 20 seconds, and 72 times (84.7%) before the end of the interaction. Therefore, the operator-requesting mechanism detected almost all situations that the robot could not handle by itself. We believe that the mechanism improved the success rate of the route guidance.

"Operator is not needed" situations

We also evaluated the performance of the operator-requesting mechanism in “operator is not needed” situations, defined as situations where an “operator is needed” did not happen in an interaction and where nobody was on the floor sensors more than three seconds.

“Operator is not needed” situations happened 490 times in the semi-autonomous condition. The operator-requesting mechanism mistakenly called the operator 23 times (4.7%). The cause of the mistakes is based on the noise information of the floor or the tactile sensors and speech recognition errors. Therefore, the operator-requesting mechanism detected almost all of the “operator is not needed” situations. We believe that the mechanism reduced the operating times.

5.1.5 Success Rate of Speech Recognition

We evaluated the speech recognizer’s performance, which is critically relevant to task performance in the autonomous condition, by calculating the correct answers per speech utterances recognized by the robot. If the robot system failed to detect speech, such as a voice that is too low, it was not counted. A correct answer is defined as speech where the speech recognizer outputs a recognized word whose meaning matches the visitor’s speech.

The system detected 1,571 sentences during the field trials, 334 of which were correctly recognized. This is a success rate of only 21.3%, although in the laboratory we achieved word accuracies that exceeded 90% with 70 dBA of background noise [26]. This contradiction indicates the inefficiency of current technologies in real-world situations.

Speech recognition failed for several reasons: mismatching with the prepared language model, inadequate vocabulary, excessively low speech volume, excessively loud voices (mainly from children), and non-constant environmental noise. In the station, the noise level usually ranged between 65 and 70 dBA, which is not quiet but is still a possible level for the speech recognizer.

5.2 Attitudes toward the robot

5.2.1 Visitor Interactions

Visitors freely interacted with the robot, especially many who seemed curious about such interaction. When a robot offered route guidance, some people repeatedly asked for such simple destinations as a vending machine or the toilet. They seemed fascinated by the robot and continued to interact with it even though it failed to react to their speech due to speech recognition errors.

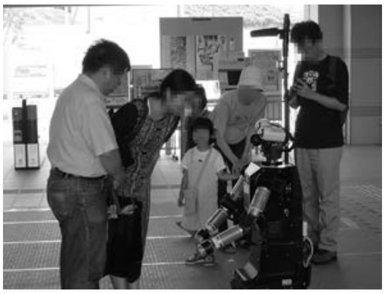


Fig. 8 Interaction between robot and visitors

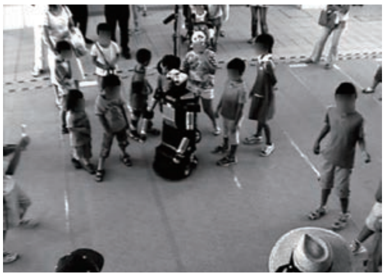


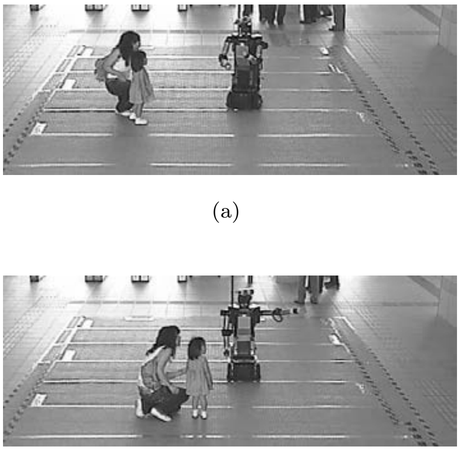
Fig. 9 Many people interacting with robot

Moreover, bystanders often observed conversations between the robot and other visitors, particularly parents of children who were interacting with the robot.

Perhaps, more interesting interactions reflect the smaller numbers of people who actually used the robots. Some visitors asked for information about a place that they really seemed to want to find, such as the nearest bus stop or shopping mall. These people appeared satisfied with the information from the robot, said thank you, and left after getting route guidance (Fig. 8).

Moreover, interactions between the robot and families were also interesting. Some parents encouraged their children to interact with the robot when it addressed them. Most children showed some anxiety because the robot was completely new to them; they often did not say anything for more than 10 seconds (Fig. 10-(a)). In such situations in the semi-autonomous condition, the operator-requesting mechanism sometimes called the operator because the robot could not hear anything for a long time or because interactive behaviors were repeated over ten times. Then the operator supported the robot by selecting such behaviors as offering route guidance again (Fig. 10-(b)).

Other interesting scenes involved multiple groups of people around the robot, as shown in Fig. 9. In such situations, different groups simultaneously interacted with the robot. They did not interact with each other directly, but they interacted with the robot in turn. Moreover, some parents ordered to their children to take turns with other children.



(a)

(b)

Fig. 10 Family interacting with robot in semi-autonomous condition

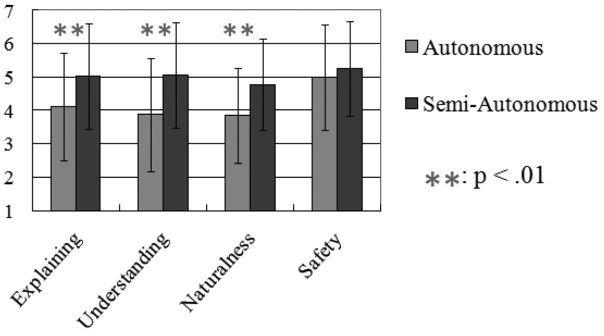


Fig. 11 Subjective impressions

5.2.2 Subjective Impressions

We asked all the visitors who interacted with the robot to answer a questionnaire in which they rated items on a scale of 1 to 7, where 7 is the most positive. We gathered 77 questionnaires in the autonomous condition and 91 in the semi-autonomous condition. The following items were used:

- Explanation: degree to which you understood the robot's explanations

- Understanding: degree to which the robot understood you

- Naturalness: degree of naturalness of robot's behavior

- Safety: degree to which you think the robot is safe

Figure 11 shows the questionnaire results. We verified the differences between the semi-autonomous and autonomous conditions with an Analysis of Variance (ANOVA) that revealed significant differences between the conditions for the impressions of Explaining, Understanding, and Naturalness (p<.01). For these impressions, the subjects evaluated the semi-autonomous robot more highly than the autonomous robot. In other words, these results indicate higher acceptability of the semi-autonomous robot than the autonomous robot; we note that all items on the autonomous condition are around the middle. We think the results also indicate that visitor basically accepted the autonomous robot. Furthermore, there was no significant difference in the Safety impression.

6 Discussion

6.1 Design Implications

This study also offers design implications. In this section, we describe them based on observations of the interactions between the robot and ordinary people in the train station.

6.1.1 Bystanders who overheard interaction

In the trials, we often observed bystanders who were just looking at the conversations between the robot and other visitors. Unfortunately, most bystanders were standing beyond the floor sensors, so the robot could not approach them. If the environmental system covered a wider area in the station, the robot could interact with more people.

One of the possibilities for covering a wider area is to use laser range finders. Recently, Dylan et al. developed a robust human-tracking system with multiple laser range finders [27] that might enable a robot to move by using the robust position information of the robot and visitors. Satake et al. also proposed a method for a mobile robot to approach visitors more naturally by considering their trajectories [28]. From another perspective, Shiomi et al. investigated how human-robot interaction changes when the robot moves forward or backward to encourage people to listen to a guide robot [11]. These approaches are also useful to increase the number of interacting people.

6.1.2 Response timing of the robot in conversations

The robot sometimes could not respond to visitors because it was waiting for the speech recognition result from the speech recognizer or the operator when the robot was talking with visitors. Long waiting times might give negative impressions to the interacting visitors; in fact, some visitors reported such feelings in their free-answer comments.

We believe that a conversational filler behavior is useful to reduce such negative feelings [29]. In Japanese “etto” is used to buy time and resembles “well...” or “uh...” in English. When visitors ask the robot something, it should use such words to buy time while it is waiting for the speech recognition results from the speech recognizer or the operator.

6.1.3 Interaction with multiple groups of people

Multiple groups of people often simultaneously interacted with the robot, as shown in Fig. 9. At first, we assumed that each group of visitors such as a family might prefer to interact with the robot. Thus, the interaction ways in real environments were quite different from our estimation.

To more naturally interact with multiple groups of people, the robot should consider their purposes toward the robot. If they conflict, the situations will be chaotic. To prevent such situations, the robot should control social situations and explicitly indicate the contexts to unify everyone's purposes toward the robot. [30]. To interact with a group simultaneously, the robot should estimate whether a group's state is suitable for the robot's intended task [31].

6.1.4 Effects of operator's existence

In the semi-autonomous conditions, we did not explicitly admit the existence of the operator. If we had confessed the operator's existence, would the interactions change? Yamaoka et al. investigated how people feel when they are interacting with the robot itself or a human behind it [32].

They reported that two-thirds of the participants of the experiments felt that they were interacting with the robot itself even if they were informed about the operator's existence. Their enjoyment was unaffected by the knowledge of whether the robot was controlled by a program or a human, although their impression of robot intelligence indicated that they distinguished between these conditions. Therefore, we think that the operator's existence would not change interactions dramatically.

On the other hand, visitors must be informed of the operator's existence if the robot's services are closely related to such privacy issues as using personal information. Our study was conducted as an academic trial to investigate the effectiveness of a semi-autonomous robot system in a real environment with a route-guidance service; no such problems happened in our study. We note that a semi-autonomous approach is a powerful way to realize an actual working robot in a real environment, but ethical issues must be considered carefully.

6.2 Perspective for Development Methodology

In this section, we report the prospects of completely exploiting the gathered realistic user data as a reflective layer. These data enable us to implement greater autonomy in the system, even though we only improved a few of the robot's functions based on this approach.

6.2.1 Finding Interaction Flows

In this study, people's main interest was testing the robot's capability, not receiving information. Even after receiving the information, some lingered around the robot and continued to interact with it. In fact, some people asked the robot about route guidance more than three times.

At first, we assumed that visitors might prefer to finish the interaction after receiving the route guidance. Thus, the interaction flows in real environments were quite different from our estimation. Therefore, in autonomous condition, the robot sometimes failed to interact with people well.

On the other hands, in the semi-autonomous condition, the operator was able to supply a flexible interaction flow in such situatios. We expect that analysis of the operation logs will enable us to improve the interaction flow by retrieving a typical interaction flow made by the operator.

6.2.2 Incremental Developing Behaviors

One of the difficulties in developing such a real field system is that predicting all the behaviors of people is very difficult. In the experiment, we prepared several destinations for guidance about facilities without including such simple destinations that visitors could see because they were quite near the robot.

However, the main interest of visitors was testing the robot's capability; visitors often asked for such places as the vending machines in the station, even though they were visible just a few meters away. Thus, we implemented such guidance behavior during the experiment. Because the provided service was very simple, we did not have another chance to incrementally implement behaviors. We believe that the need for incremental development will increase if the robot's task becomes more complex.

6.2.3 Decreasing operator load

One important future work in semi-autonomous robot systems is decreasing the operator load. This can be achieved in two ways: increasing the robot's autonomy and developing more useful interfaces for operations.

The former is related to developing a recognition system about the environment around the robot and the interacting people. For example, robust position estimation and localization within a wide area are important functions to increase the robot's autonomy [27, 33]. Another perspective for the former is to learn the operator's decision or interaction logs with sensor information because such log information provides powerful learning data for robot behavior [34,35].

Related to the latter, some researches enable one operator to control multiple robots by reducing the operator load [6,36]. One problem in teleoperation with multiple conversational robots is the conflict of using the operator's resources. The operator can only deal with speech recognition for one robot at a time, even though multiple robots simultaneously need the resource. Conversational interactions tend to follow patterns that sometimes make it possible to anticipate the need for the operator. Dylan et al. scheduled behaviors to avoid conflicts about operator resources [6]. These approaches will decrease the load of operators who control communication robots.

6.2.4 Improving Speech Recognition Performance

In our experiments, even though the prepared speech recognition system achieved 92.5% word accuracy in an indoor, 75 dBA noise environment, it only resulted in 21.3 accuracy in the real environment.

Currently, the most critical failure is caused by speech recognition, which mainly reflects utterances that do not fit the implemented language model. For example, even though " Would you tell me the route to Kyoto? " is included in the model, it has difficulty recognizing the keywords in such similar utterances as "Tell me how to go to Kyoto" (grammar mismatch) and "Would you tell me the route to Osaka?" (vocabulary mismatch). This is critical because daily conversation has various ways of expressing ideas.

We analyzed how many utterances spoken by visitors matched the implemented language model, which resulted in a rate of 51.3%. Speech recognition often fails when mismatches exist between utterances and the model. One of the difficulties of improving the performance of speech recognition in a real environment is the producibility of the situations; however, such real data must be gathered and used to increase system performance and to develop a new system that works in real environments. Therefore, we plan to improve the performance for the remaining 48.7% of the utterances by adding mismatched utterances to the language model.

6.3 Limitations

Since we only conducted tests with a particular robot and sensors, an operator, and in the specific environment of a train station, the generality of our findings is limited from the viewpoint of reproducibility. However, such a situation is difficult to avoid in human-robot interaction because using two or more different robots is too expensive; generalizing findings by preparing a large number of operators with knowledge of robotics and an understanding of the system in different environments is also difficult. Yet we believe such field trials are critical to investigate how the current technologies work in real situations and how to improve them by realistic data. We believe that our findings are applicable to other robots with similar appearance and interaction complexity.

7 Conclusion

We implemented a networked robot system that consists of a semi-autonomous communication robot, floor sensors, cameras, and a sound-level meter. Moreover, we implemented an operator-requesting mechanism that autonomously detects situations that the robot cannot handle by itself and requests that a human operator assume control. This mechanism is an important function for semi-autonomous robots. Such basic communicative behaviors as greetings and route guidance are implemented for the robot, which autonomously approaches visitors and interacts based on the position information estimated by the floor sensors. The robot autonomously controls the interaction flows based on sensory information.

We confirmed that the robot system worked correctly in a real environment through a field trial at a train station where the robot was given a route-guidance task. The results suggest the promising potential of the robot system for serving people. The mechanism correctly requested operator's help in 84.7% of the necessary situations. The operator only controlled 25% of the experiment time and mainly operated such high-level functionality as the transition of behaviors. The robot system successfully guided 68% of the visitors whose subjective impressions were also good and indicated high acceptability of our robot in the public space. Observed interaction scenes between the robot and visitors also provided design implications. Important future work includes establishing a methodology that utilizes the gathered data to improve the robot's performance.

Acknowledgements We wish to thank the staff of the Kinki Nippon Railway Co., Ltd. for their cooperation. We also thank the following ATR members for their helpful suggestions and cooperation: Satoshi Koizumi, Toshihiko Shibata, Osamu Sugiyama, Kenta Nohara, and Anton Zolotkov. This research was supported by the Ministry of Internal Affairs and Communications of Japan.

References

1. C. Breazeal and B. Scassellati: "Infant-like social interactions between a robot and a human caretaker," Adaptive Behavior, 8(1), 2000.
2. Bilge Mutlu, Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita, Nonverbal Leakage in Robots: Communication of Intentions through Seemingly Unintentional Behavior, 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI2009), pp. 69-76, 2009.
3. Hideaki Kuzuoka et al. "Reconfiguring spatial formation arrangement by robot body orientation," 5th ACM/IEEE International Conference on Human-Robot Interaction(HRI2010), pp. 285-292, 2010.
4. Toru Takahashi, Kazuhiro Nakadai, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno: Missing-Feature-Theory-based Robust Simultaneous Speech Recognition System with Non-clean Speech Acoustic Model. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2009)
5. A. Sanfeliu, N. Hagita, and A. Saffiotti, "Network Robot Systems," Special Issue: Network Robot Systems, Robotics and Autonomous Systems, 2008.
6. Dylan F. Glas et al.: "Field Trial for Simultaneous Teleoperation of Mobile Social Robots," HRI2009, to appear.
7. D. A. Norman: Emotional design, Basic Books, 2003.
8. 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.
9. W. Burgard et al.: "The interactive museum tour-guide robot," National Conference on Artificial Intelligence, pp. 11-18, 1998.
10. R. Siegwart et al.: "Robox at Expo.02: A Large Scale Installation of Personal Robots." Robotics and Autonomous Systems, 42, 203-222, 2003.
11. M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita: "A Larger Audience, Please! - Encouraging people to listen to a guide robot -," 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI2010), 2010.
12. Andrea Bauer et al., "The Autonomous City Explorer: Towards Natural Human-Robot Interaction in Urban Environments," Int J Social Robotics, vol. 1, no. 2, Springer, 2009.
13. M.K. Lee, J. Forlizzi, P. E. Rybski, F. Crabbe, W. Chung, J. Finkle, E. Glaser, and S. Kiesler, "The Snackbot: Documenting the design of a robot for long-term human-robot interaction," Proceedings of HRI'09, 7-14.
14. H.-M. Gross, H.-J. Böhme, C. Schröter, S. Müller, A. König, C. Martin, M. Merten, and A. Bley, "ShopBot: Progress in Developing an Interactive. Mobile Shopping Assistant for Everyday Use," in International Conference on Systems, Man and Cybernetics, 2008
15. D. Dahlback, A. Jonsson, and L. Ahrenberg: "Wizard of Oz studies - why and how, Knowledge-based systems," Vol. 6, No. 4, pp. 258-266, 1993.
16. S. Dow, et al.: "Wizard of Oz Support throughout an Iterative Design Process," Pervasive computing, Vol. 4, No. 4, 2005.
17. S. Woods et al.: "Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods," Towards a Novel Methodological Approach, Int. Workshop on Advanced Motion Control, 2006.
18. A. Green et al.: "Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration," Proc. IEEE Int. Workshop on Robot and Human Interactive Communication, 2004.
19. Lee, Jun Ki, Stiehl, Walter Dan, Toscano, Robert Lopez, and Breazeal, Cynthia. "Semi-Autonomous Robot Avatar as a Medium for Family Communication and Education," Advanced Robotics, vol. 23, no. 14, pp. 1925-1949(25), Brill, 2009
20. N. E. Sian et al.: "Whole Body Teleoperation of a Humanoid Robot Integrating Operator's Intention and Robot's Autonomy," IROS2003, pp. 1651-1656, 2003.
21. B. P. Sellner et al.: "Attaining Situational Awareness for Sliding Autonomy," HRI2006, pp. 80-87, 2006.
22. Andrew Correa, Matthew R. Walter, Luke Fletcher, Jim Glass, Seth Teller, and Randall Davis, Multimodal Interaction with an Autonomous Forklift, in 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI) 2010, Osaka Japan, pp. 243-250, March 2010.
23. Tiffany L. Chen and Charles C. Kemp, Lead Me by the Hand: Evaluation of a Direct Physical Interface for Nursing Assistant Robots, in 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI) 2010, pp. 243-250, March 2010.
24. H. Kawai, T. Toda, J. Ni, M. Tsuzaki, and K. Tokuda, XIMERA: "A New TTS from ATR Based on Corpus-Based Technologies," Proc. of Fifth ISCA Workshop on Speech Synthesis (SSW5), pp. 179-184, 2004
25. T. Murakita et al.: "Human Tracking using Floor Sensors based on the Markov Chain Monte Carlo Method," Proc. Int. Conf. Pattern Recognition (ICPR04), pp. 917-920 (2004).
26. C. T. Ishi et al.: "Robust speech recognition system for communication robots in real environments," International Conference on Humanoid Robots, 2006.
27. D. Glas et al., "Laser Tracking of Human Body Motion Using Adaptive Shape Modeling," In Proc. Int. Conf. Intelligent Robots and Systems, pp. 602-608. 2007.
28. S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to Approach Humans?-Strategies for Social Robots to Initiate Interaction," HRI2009, pp. 109-116, 2009.
29. Toshiyuki Shiwa, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita, How Quickly Should a Communication Robot Respond?, International Journal of Social Robotics, 1(2), pp. 141-155, 2009.
30. M. Shiomi, T. Kanda, S. Koizumi, H. Ishiguro, and N. Hagita, "Group Attention Control for Communication Robots," International Journal of Humanoid Robotics (IJHR), 2008.
31. M. Shiomi, K. Nohara, T. Kanda, H. Ishiguro, and N. Hagita, "Estimating Group States for Interactive Humanoid Robots," IEEE International Conference on Humanoids, Dec. 2007.
32. Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita, Interacting with a Human or a Humanoid Robot?, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2007), pp. 2685-2691, 2007.
33. D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, Simultaneous People Tracking and Localization for Social Robots Using External Laser Range Finders, in Proc. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009), pp. 846-853, 2009.
34. M. Shiomi, T. Kanda, K. Nohara, H. Ishiguro, and N. Hagita, "Adaptive supervisory controls of a communication robot that approaches visitors," International symposium on Distributed Autonomous Robotic Systems (DARS) 2008.

35. S. Ikemoto, H. B. Amor, T. Minato, H. Ishiguro, and B. Jung. Physical Interaction Learning: Behavior Adaptation in Cooperative Human-Robot Tasks Involving Physical Contact. 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009.

36. Raj M. Ratwan, J. Malcolm McCurry, and J. Gregory Trafton, "Single Operator, Multiple Robots: An Eye Movement Based Theoretic Model of Operator Situation Awareness," in 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI) 2010, pp. 243-250, March 2010.

Masahiro Shiomi received M. Eng and Ph.D. degrees in engineering from Osaka University, Osaka, Japan in 2004 and 2007, respectively. From 2004 to 2007, he was an Intern Researcher at the Intelligent Robotics and Communication Laboratories (IRC) and is currently a Researcher at IRC at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan. His research interests include human-robot interaction, interactive humanoid robots, networked robots, and field trials.

Daisuke Sakamoto received his B.A.Media Architecture, M.Systems Information Science, and Ph.D. in Systems Information Science from Future University-Hakodate in 2004, 2006, and 2008, respectively. From 2006 to 2008, he worked for ATR Intelligent Robotics and Communication Labs. (IRC) as Intern Researcher. From 2008 to 2010, he worked at The University of Tokyo as a Research Fellow of the Japan Society for the Promotion of Science, and continues to be involved in the JST ERATO Igarashi design interface project as a Collaborator, and ATR IRC as a Cooperate (Visiting) Researcher. He is now a Researcher of JST ERATO Igarashi Design Interface Project. His research interests include Human-Robot Interactions, Human-Computer Interaction, and Interaction Design of Interactive Robots.

Takayuki Kanda received his B. Eng, M. Eng, and Ph. D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively. From 2000 to 2003, he was an Intern Researcher at the ATR Media Information Science Laboratories and is currently a Senior Researcher at the ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. His current research interests include intelligent robotics, human-robot interaction, and vision-based mobile robots. Dr. Kanda is a member of ACM, the Robotics Society of Japan, and the Information Processing Society of Japan.

Carlos Toshinori Ishi received the B.E. and M.E. degrees in electronic engineering from the Instituto Tecnologico de Aeronautica (Brazil) in 1996 and 1998, respectively. He received the PhD degree in engineering from The University of Tokyo (Japan) in 2001. He worked at the JST/CREST Expressive Speech Processing Project from 2002 to 2004 at ATR Human Information Science Laboratories. He is currently with ATR Intelligent Robotics and Communication Laboratories, since Jan. 2005. His major interests are on prosodic and voice quality information extraction from speech signals, linguistic and paralinguistic information processing, auditory processing, sound source localization and noise suppression processing, and application for communication robots. Dr. Ishi is a member of the Acoustic Society of Japan, and the Robotics Society of Japan.

Hiroshi Ishiguro received his D. Eng. degree from Osaka University, Japan in 1991. In 1991, he started working as a Research Assistant in the Department of Electrical Engineering and Computer Science, Yamanashi University, Japan. He moved to the Department of Systems Engineering, Osaka University, Japan, as a Research Assistant in 1992. In 1994, he became an Associate Professor of the Department of Information Science, Kyoto University, Japan, and started research on distributed vision using omnidirectional cameras. From 1998 to 1999, he worked in the Department of Electrical and Computer Engineering, University of California, San Diego, USA, as a visiting scholar. He has been a Visiting Researcher at the ATR Media Information Science Laboratories since 1999 and developed Robovie, an interactive humanoid robot. In 2000, he moved to the Department of Computer and Communication Sciences, Wakayama University, Japan, as an Associate Professor and became a Professor in 2001. He is now a Professor of the Department of Adaptive Machine Systems, Osaka University, Japan, and a group leader at ATR Intelligent Robotics and Communication Laboratories.

Norihiro Hagita received B.S., M.S., and Ph.D. degrees in electrical engineering from Keio University, Tsuruoka City, Japan in 1976, 1978, and 1986, respectively. From 1978 to 2001, he was with the Nippon Telegraph and Telephone Corporation (NTT). He joined the Advanced Telecommunications Research Institute International (ATR) to establish the ATR Media Information Science Laboratories and the ATR Intelligent Robotics and Communication Laboratories in 2001 and 2002, respectively. He is now a director at ATR ICT Environment Research Laboratories Group and ATR Intelligent Robotics and Communication Laboratories. His current research interests include communication robots, network robot systems, interaction media, and pattern recognition.He is a Fellow of the Institute of Electronics, Information, and Communication Engineers, Japan and the ATR. He is also a member of the Robotics Society of Japan, the Information Processing Society of Japan, and The Japanese Society for Artificial Intelligence. He is a Co-Chair for the IEEE Technical Committee on Networked Robots.