A Communication Robot in a Shopping Mall

Takayuki Kanda, *Member, IEEE*, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, *Member, IEEE*, and Norihiro Hagita, *Senior Member, IEEE*

***Abstract*—This paper reports our development of a communication robot for use in a shopping mall to provide shopping information, offer route guidance, and build rapport. In the development, the major difficulties included sensing human behaviors, conversation in a noisy daily environment, and the needs of unexpected miscellaneous knowledge in the conversation. We chose a network-robot system approach, where a single robot's poor sensing capability and knowledge are supplemented by ubiquitous sensors and a human operator. The developed robot system detects a person with floor sensors to initiate interaction, identifies individuals with radio-frequency identification (RFID) tags, gives shopping information while chatting, and provides route guidance with deictic gestures. The robot was partially teleoperated to avoid the difficulty of speech recognition as well as to furnish a new kind of knowledge that only humans can flexibly provide. The information supplied by a human operator was later used to increase the robot's autonomy. For 25 days in a shopping mall, we conducted a field trial and gathered 2642 interactions. A total of 235 participants signed up to use RFID tags and, later, provided questionnaire responses. The questionnaire results are promising in terms of the visitors' perceived acceptability as well as the encouragement of their shopping activities. The results of the teleoperation analysis revealed that the amount of teleoperation gradually decreased, which is also promising.**

***Index Terms*—Information-providing, network robot system, robots for shopping mall, route guidance, social human–robot interaction.**

I. Introduction

As ROBOTS move from laboratories and into our daily lives, they are expected to interact with people and support daily activities. In particular, humanoid robots are already being used to provide help with physical activities [1]. Moreover, researchers have started to consider how humanoid robots might be suitable for communication with humans [2]–[5]. Their human-like bodies enable them to perform natural-gaze motion [6] and deictic gestures [7]. These features of humanoid robots will allow them to perform such communicative tasks in human society as route guidance and explanations of exhibits.

The applications of robotic systems continue to expand. Whenever a new application is to be covered by a robotic system, the challenge is to identify a possible combination of system components, e.g., sensors and actuators, and a way to integrate them. In robotics, the success of such development has been accumulated as useful scientific knowledge. For instance, previous studies have revealed that robots can be used as museum guides [8], [9] in city exploration [10], as receptionists who assist visitors [11], as peer tutors in schools [12], in mental healthcare for elderly people [13], in autism therapy [14], [15], and in childcare [16].

This paper[1] reports our challenges in applying a robotic system to an information-providing task in the daily environment of a shopping mall. The difficulties included sensing the human behaviors and conversation in a noisy daily environment, and the unexpected needs of various information during conversation. Our approach used a network-robot system to supplement the robot's poor sensing capability and a human operator to support the robot's conversational and informational capability.

This paper consists of the following sections. In Section II, we report how we designed the role of the robot. In Section III, we report its implementation. In Sections IV and V, we report its influence on people.

II. Role of the Robot

What kind of tasks do people want robots to perform in their daily lives? According to a Japanese government report [18], a majority of respondents want robots that provide information at such public spaces as train stations and shopping malls,[2] while people also want robots to do such physical tasks as toting luggage. Therefore, we decided to explore an information-providing task for a robot in a public space, specifically, a guide robot in a large mall (see Fig. 1).

The next question addresses the roles of a guide robot in a mall. Many other facilities, such as maps and large screens, provide information (see Fig. 2). In contrast, a robot offers unique features based on its physical presence, its interactivity, and its capability for personal communication. We defined three roles based on these features.

Manuscript received November 12, 2009; revised May 10, 2010; accepted July 22, 2010. Date of publication August 26, 2010; date of current version September 27, 2010. This paper was recommended for publication by Associate Editor K. Yamane and Editor L. Parker upon evaluation of the reviewers' comments. This work was supported by the Ministry of Internal Affairs and Communications of Japan. This paper was presented in part at the 2009 ACM/IEEE International Conference on Human–Robot Interaction.

The authors are with the Advanced Telecommunications Research Institute International Intelligent Robotics and Communication Laboratories, Kyoto 619-0288, Japan (e-mail: kanda@atr.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2010.2062550

[1]This paper is an extended version of a conference paper [17] with additional description of technical implementation and discussions.

[2]This might be relatively high in Japan compared with other countries: 76.2% of the respondents think it is good to have robots at transportation facilities, such as train stations, and 87.5% think that at these places, guidance is a good task for robots; 64.2% think it is good to have robots at commercial places, such as a shopping mall, and 87.9% think that at these places, guidance is a good task for robots.

Fig. 1. Shopping mall.

Fig. 2. Large information screen in shopping mall.

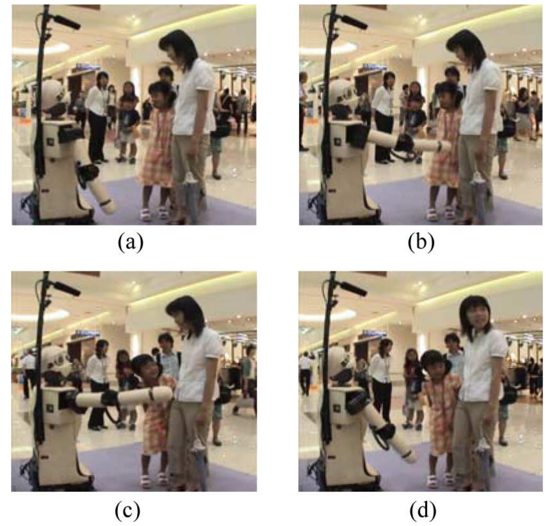

Fig. 3. Robot guiding a customer with deictic representation. (a) Person asks for route. (b) It points and looks that way. (c) "Please go that way." (d) "After that...".

Role 1: Guiding

The size of shopping malls continues to increase. Sometimes, people get lost and ask for directions. Even though a mall has maps, many still prefer to ask for help, especially because some information is not shown on a map; people ask, "Where can I buy an umbrella?" Here, a route-guidance service is needed. People even ask strange questions, e.g., "Where can I print a digital camera?" One author was actually asked this in a mall, which suggests the need of human support for the robot. Even if a robot has great language-interpretation capability, people might not use correct language.

In contrast with a map or other facilities, a robot has the following unique features. It has a physical existence, it is colocated with people, and it is equipped with human-like body characteristics. Thus, as shown in Fig. 3, a robot can naturally explain a route by pointing like a human, looking in the same direction as the person is looking, and using such reference terms as "this way." Since the safe locomotion speeds of robots remain very slow, we arranged the route guidance as an information-providing service at a certain location instead of a service that physically accompanies people to destinations.

Role 2: Building Rapport

From the customer's viewpoint, since the robot is a representative of the mall, it needs to be friendly to customers. In addition, since people repeatedly visit malls, a robot needs to naturally repeat interaction with the same person; thus, a function that builds a rapport with each customer is useful. The importance of building rapport has been studied in human–computer interaction (HCI) in the context of affective computing [19].

Moreover, one future scenario in this direction is a function of customer-relationship management. Previously, this was done by humans; for example, in a small shop, the shopkeeper remembers the "regulars" and molds communication to each individual. Typically, shopkeepers are particularly cordial to good customers. Recently, since managing numerous customers has become too unwieldy, information systems have assumed this role, in part, such as the mileage services of airline companies and the point systems of credit cards and online-shopping services, such as Amazon. However, these information systems do not provide *natural* personalized communication like humans. In contrast, we believe that in the future a robot might provide natural communication and personalized service for individual customers and, thus, develop a relationship or a rapport with them.

Role 3: Advertisements

From the mall's point of view, advertising is an important function. For instance, posters and signs are placed everywhere in malls. Recently, information technologies are also being used for such purposes. Fig. 2 shows a large 2.5 m $\times$ 5 m screen that provides shopping information to customers in the shopping mall, where we conducted our field trial. The screen shows such shop information as locations in the mall and features of shops' products.

We believe that a robot can also be a powerful tool for this purpose. Since a robot's presence is novel, it can attract people's attention and redirect their interest to the provided information [20]. In addition, it can furnish information in the same way that people talk to each other; it can describe shops and products from its first-person viewpoint (see Section III-D4).

III. Hardware and System Design

Fig. 4 shows an overview of the system architecture. The robot identifies a person with an radio-frequency identification (RFID) tag reader, continues to track his/her position with floor sensors, and tracks the person's face with its camera. As in a Wizard of Oz (WOZ) method, speech recognition is conducted by a

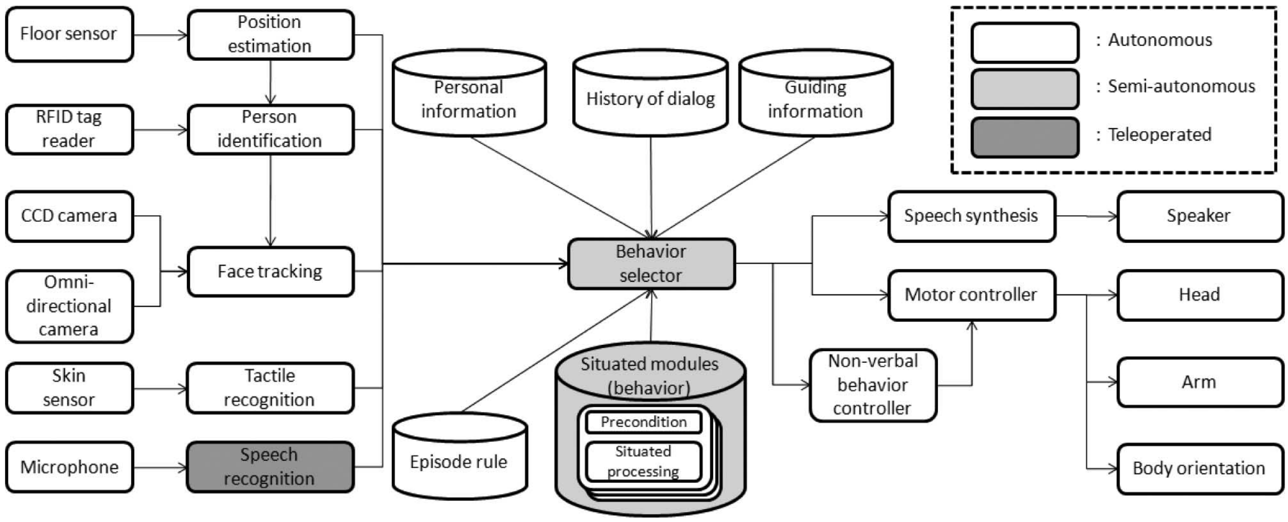

Fig. 4. System overview.

human operator. This information is sent to a behavior selector that chooses an interactive behavior based on preimplemented rules called *episode rules*, which are the history of previous dialogues with this person and her/his personal information.

Our architecture is behavior-based, where low-level submodules can be restricted by modules in higher layers. Perhaps uniquely, there is no central module that coordinates all other modules, unlike sense-plan-act-type architectures, where all information is once corrected in such a coordination module. The top of the architecture, in terms of coordination, is the situated module. Depending on the situation, a situated module can override the default structure of the coordination. In terms of information sharing among modules, our model is a very simple blackboard model, where all modules update shared variables, which are placed in the shared memory.

This section reports our design considerations behind this architecture and the details of each mechanism in the architecture.

A. Design Considerations

Based on the three roles to be provided, we explored a combination of hardware and infrastructure. Some researchers are studying a stand-alone robot that has all sensing, decision making, knowledge processing, and acting capabilities. In contrast, others are focusing on a combination of robots, ubiquitous sensors, and humans. Considering the complexity of the given task, we chose the latter strategy, which is known as a network-robot system [21], in which a robot's sensing capability and knowledge are supplemented by ubiquitous sensors and a human operator.

Moreover, we made two decisions about the hardware/software design. First, we decided to use human operators for speech recognition and to supplement knowledge gaps. Our robot system is designed to operate without an operator; however, when providing information, instability and awkwardness cause disappointment, and the quality of current speech-recognition technology remains far from adequate for our needs. For instance, a speech-recognition system prepared for a noisy environment, which performs at 92.5% word accuracy in 75 dBA noise [22], resulted in only 21.3% accuracy in a real environment [23]. This reflects the natural manner of daily utterances, changes in voice volume among people and even within the same person, and the unpredictability of noise in a real environment. Therefore, since a speech-recognition program causes so many recognition errors, the robots would have to ask for clarification too often.

Using human operators is quite common in HCI and human–robot interaction for prototyping and is known as WOZ approach [24]. In addition, since our vision is to use a human operator for more than making a prototype, we believe that a robot can start working in daily environments with human operators with a technique that minimizes the task load of operators, such as one that allows a single operator to control four robots [25].

Second, we restricted the robot from moving around its environment. Since the robot is mobile, this decision appears disappointing; however, in negotiations with the mall management, considering the current robot's capability in terms of its perception of people movements, we made this decision to reduce such safety risks as contact with people or such business risks as crowds in the middle of the corridor or in front of a shop.

B. Hardware and Sensing System

Robovie, i.e., a humanoid robot, was used for this study. Among its sensing capabilities, we focused on three capabilities: position estimation of people, person identification, and face tracking. Since the first two are difficult for a robot in a noisy real world, we decided to use ubiquitous sensors. Face tracking was primarily done with its own sensor by incorporating information from ubiquitous sensors. The following section describes the details of the hardware and our decisions.

1) Robovie's Hardware: Robovie is an interactive humanoid robot characterized by its human-like physical expressions and its various sensors [26] (see Fig. 5). Robovie has a head, two arms, a body, and a wheeled-type mobile base. Its height and weight are 120 cm and 40 kg. It has the following degrees of

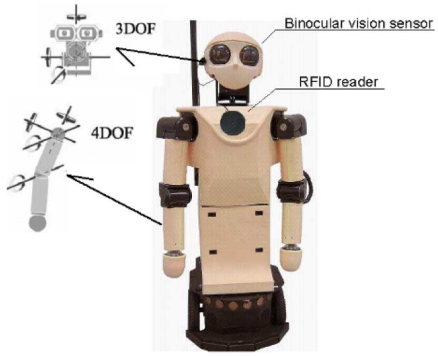

Fig. 5. Robovie-IIF.

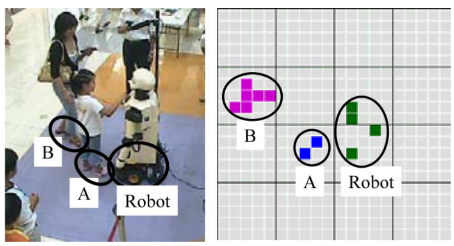

Fig. 6. Floor sensors.

freedom (DOFs): two for its wheels, three for its neck, and four for each arm. On its head, it has two charge-coupled device (CCD) cameras and a speaker. An omnidirectional camera and a microphone are attached on an extended pole connected to its shoulder. We used a corpus-based speech synthesis [27] to generate speech.

The version used for this study is Robovie-IIF, which has tactile sensor elements embedded in the soft skin that covers its whole body. With the tactile sensor, the robot recognizes such types of touching as petting and stroking as well as the place being touched [28].

2) Position Estimation: We used external sensors for detecting and tracking people's positions around the robot, since we need such highly accurate detection to robustly operate the robot in crowded environments. We chose floor sensors, because they are the most robust device in terms of stability to detect a person's presence, and their area can also be used to indicate the conversational distance of the robot, where people can interact with it.

We installed 16 floor sensors units, VS-SS-F (Vstone Corporation, Osaka, Japan), around the robot that covered a 2 m $\times$ 2 m area (see Fig. 6). Each sensor unit is 50 cm $\times$ 50 cm with 25 ON–OFF pressure switches. The sensor resolution is 10 $\times$ 10 cm^2. Each sensor provides binary output, regardless of whether there is pressure on it. These outputs are read through an RS-232 C interface with a sampling frequency of 5 Hz.

We used a Markov Chain Monte Carlo method and a bipedal model to estimate people's positions with a model of a person's gait pattern [29]. This method provided robust position estimation up to a few persons with 21 cm average position error.

3) Person Identification With RFID Tag: Various techniques for *person identification* exist, including computer vision to recognize faces, active-type RFIDs, and passive-type RFIDs. For person identification, we employed a passive-type RFID tag because of their 100% accuracy in person identification. Such accuracy is crucial, since the misidentification of a person causes embarrassing interaction in human communication. One downside is that a system based on passive-type RFID tags requires intentional user contact with an RFID reader; since passive-type RFIDs are already widely adopted for train tickets in Japan, people are accustomed to using them. We do not consider this problematic.

Fig. 7. RFID tag and reader.

The left side of Fig. 7 shows a passive-type RFID tag (Texas Instruments Incorporated, RI-TRP-WRHP) embedded in a cellular phone strap that uses a frequency of 134.2 kHz. The accessory is 4 cm high. The RFID tag's reader is attached to the robot's chest. Since a passive-type RFID system requires a contact distance for reading, users were instructed to place the tag on the tag reader for identification and to interact with the robot (see Fig. 7, right). We provided this RFID tag to mall customers, who registered for the field trial.

Once an RFID tag is read by the reader, the system associates the person's ID with the person detected by the floor sensor. When multiple persons are tracked with the floor sensors, it associates the ID to the nearest person. Once the ID is associated, it keeps tracking the ID until the person with the ID leaves the area on the floor sensors. The location of the person with the ID is repeatedly used during interaction. The robot orients the body direction and maintains eye contact; the interaction is concluded when the person leaves.

4) Face-Tracking System: We developed a face-tracking algorithm for a communication robot that integrates information from both foveal and omnidirectional visions and actively controls the robot's head orientation [30]. However, in a real environment, false-positive faces were frequently detected, which largely hindered the performance.

Thus, in addition to this basic mechanism, we used information from person tracking and identification. The search area of a person's face is limited to the area, where people are detected by floor sensors. When the person is identified by the RFID tags, the system retrieves the person's height information, which is prestored to vertically limit the search range. With these combinations, the robot was usually able to orient its gazing direction to a user's face.

C. Mechanism for Generating Interactive Behaviors

The mechanism for generating interactive behaviors follows the architecture reported in [26]. For this study, we added databases for repeated interaction and route guidance and

TABLE I
DATA SCHEMA OF PERSONAL INFORMATION TABLE

ID	Name	Nickname	Height	Age	useDays
1	Yamada	Daddy	180	40	2
2	Inoue	Ino	160	24	1
3	Sato	Sattyan	120	12	1

TABLE II
DATA SCHEMA OF INTERACTION HISTORY TABLE

ID	Time	Module ID	Result value
1	136	GUIDE_CAFE	1 (success)
1	149	ASK_FAVORITE	2 (ice cream)
1	160	BYE	1 (success)
2	200	HELLO	1 (success)
2	220	OFFER_ROUTE	120 (asked restroom)

TABLE III
DATA SCHEMA OF GUIDING INFORMATION TABLE

Shop Name	ID	X	Y	Floor	Relay point
CAFE	1	300	1000	3	Escalator_1
DRUG_STORE	2	1000	2000	4	Elevator_3
RESTROOM	3	5000	300	2	null

TABLE IV
PSEUDOCODE OF "SHAKE_HANDS" BEHAVIOR

```
Precondition: IsHumanExist() == true
Situated processing:
Initially, returnValue = 0;
Say ("Let's shake hands");
PlayMotion ("shake_hands");
startTime=GetNowTime();
while ( GetNowTime() – startTime < 3 seconds )|
    if (IsReactTactileSensor( RIGHT_HAND ) == true )
        returnValue = 1;
        break;
    end if
end while
if (returnValue ==1)
        Say ("Thank you");
end if
return returnValue;
```

extended the behavior selection mechanism to be applicable to different individuals. In this section, we briefly explain the basic mechanism (see [26] for more details) and describe the new features of knowledge representation and behavior selection.

1) Knowledge Representation: The robot system has three databases: personal information, dialog history, and guidance information (see Fig. 4). Table I shows the data structure for personal information. Each person has a unique ID that is associated with RFID tags. Participants also furnished their names (only for administrative purposes), nicknames, height, and age. The robot used nicknames in the interaction. This design addressed privacy concerns, since we found in our preliminary study that some people prefer that their real names not be used in a public environment [31]. The table shows some *variables*. For example, the "useDays" *variable* is incremented when the person visited the robot on a different day from last visit.

Table II shows the data structure for dialog history, which is simply stored as a sequence of the history of the executed *situated modules*, and includes information about how participants answered questions from the robot as the result values of the *situated module*. The history is stored in the database when the person finished interaction with the robot and restored to the memory from the database when the person visits the robot again. In combination with the *episode rules*, this dialog history enables the robot to adapt its interaction to each person over time. Table II indicates a brief example, where Mr. Yamada answered "I like ice cream" for ASK_FAVORITE behavior, and Mr. Sato asked about a restroom location.

Table III shows the data structure for the guiding information that is used by the route-guidance behaviors. There are unique IDs for each shop, name, *x*–*y* coordinates, a floor, and a *relay point*, which is used when the destination is outside the visible distance. For example, when it provides directions to the CAFE, it says, "Please go this way (*pointing*), and take the first escalator to go to the third floor. The coffee shop is around there." When the destination is within a visible distance, the *relay point* is not used. For example, the robot gives directions to RESTROOM: "Please go this way (*pointing*); you will see the restroom."

2) Dialog Control:

a) Behavior (situated module): The central idea in our architecture is the concept of *situated modules* (also referred to as "behavior" in this paper) and *episode rules* [26]. The robot only executes one *situated module* at each moment. Each *situated module* controls the robot's utterances, its gestures, and its nonverbal behaviors in reaction to a person's action.

In concrete, each *situated module* consists of precondition and situated processing parts. Table IV shows an example of the pseudocode of one *situated module*. In this example, the precondition part is used to determine whether this *situated module* is executable at a current moment. It verifies the presence of a person in front of the robot, as detected by the floor sensor. The behavior selector only chooses a *situated module* that is executable at the current moment. Once this module is selected for execution, the system starts to execute the situated processing part of the module. The robot offers to shake hands by saying "Let's shake hands", and waits for input from a tactile sensor to react to the handshake. The result of the situated recognition result is returned to the behavior selector so that the system can use it for choosing the next behavior. In this example, it returns 1, which represents a success when a reaction is observed at the tactile sensor, which is considered a successful reaction in the handshaking behavior. It returns 0, which represents a failure when there is no reaction until the timeout time.

b) Episode rules: The transitions among behaviors are precoded but not in a strictly sequential way. Our architecture allows us to describe much flexible relationships. We set up

TABLE V
GRAMMAR OF EPISODE RULES

1. <ModuleID=result_value>…<…>NextModule
2. (<ModuleID1=result_value1>|<ModuleID2=result_value2>)...
3. (...){n,m}...
4. !<…>NextModule
5. ^<ModuleID=^result_value>NextModule
(1: basic structure of describing executed sequence, 2: "OR", 3: repetitions, 4: negation of *episode rule*, 5: negation of Module ID and result value)

a large pool of rules named *episode rules*, which describe the transition rules between the *situated modules*. These transitions rules are defined in the same way as regular expressions. With this mechanism, we can specify complex conditions (i.e., high-description capabilities) in a reasonable development time. In addition, the pool mechanism allows us to flexibly add new behavior and episodes rules.

In concrete, each moment when a *situated module* finishes its execution, the behavior selector chooses the next *situated module* by referring to all preimplemented *episode rules* and the execution result of the current *situated module*.

Table V shows the basic grammar and examples of the *episode rules*. Each situated module has a unique identifier called a Module ID. "<ModuleID = result_value>'' refers to the execution history and the result value of the situated modules. "<ModuleID1 = result_value1> <ModuleID2 = result_value2>…" means the referring rule of the previously executed sequence of situated modules (see Table V, item 1). "<…><…>'' denotes a selective group (OR) of the executed *situated modules*, and "(…)" means a block that consists of a *situated module*, a sequence of *situated modules*, or a selective group of *situated modules* (see Table V, item 2). Similar to regular expressions, we can describe the repetition of the block as "(… $\{n, m\}$," where n gives the minimum number of times that match the block, and m gives the maximum (see Table V, item 3). We can specify the negation of the whole *episode rule* with an exclamation mark "!". For example, "! <…>…<…>NextModuleID" (see Table V, item 4) means the module of NextModuleID will not be executed when the *episode rule* matches the current situation specified by "<…>…<…>.'' The negation of a Module ID or a result value is written with a caret: "^" (see Table V, item 5).

c) Using dialog history: The robot has a mechanism for adjusting its interactive behaviors to each individual based on its dialog history. For example, the robot asks whether the person likes icecream on day one; when the person revisits the robot on day two, the robot starts a conversation: "Last time, you said that you liked ice cream. Well, I asked about ice cream in this shopping mall, and I found…"

We extended our architecture to realize such an interaction. Two types of information are stored: abstracted information in personal information tables (see Table I) and raw execution logs of *situated modules* in interaction-history tables (see Table II). When the interacting person is identified by RFID tag, the robot updates the personal information stored in the personal-information table (see Table I), i.e., incremented *useDay variable* if this person did not visit the robot on this day. The system starts to store the dialog history in the database after the end of each execution of *situated modules.*

We extended the *episode rule* to include the notion of *condition*, which is defined by an equation with a *variable* (e.g., *useDays* $==$ 0 or *useDays* $>$ 0), where *variables* are stored in the personal-information table (see Table I). If a *condition* is used, the rule will only be applicable if the condition is satisfied. The following is the basic grammar of an extended *episode rule*:

$$Extended\ episode\,rule = episode\,rule,\ condition.$$

Fig. 8 shows an example of a *situated module* transition with *episode rules. Episode rules* 1 and 3, which are typical examples of an *episode rule* with a condition, are applicable for the situation after the HELLO *situated module* returns value 1. They use the *useDays* variable to define their condition; on the first day, the behavior selector chooses the SHAKE_HANDS *situated module* after HELLO. Since the *useDays* variable is at 0 at the interaction on the first day, *episode rule* 1 is applicable, but not *episode rule* 3. On the other hand, on the second day, the behavior selector chooses THANKS_REVISIT after HELLO because the *useDays* variable has been incremented to 1 now; the behavior selector referred *episode rule* 3, since it is applicable, but not *episode rule* 1.

Episode rule 4 is an alternative way of using the dialog history. In this case, the *episode rule* has no condition. On day two, the behavior selector chooses YOU_LIKE_ICECREAM by referring to *episode rule* 4 after END_DAILY_INTERACTION. *Episode rule* 4 matches because ASK_FAVORITE was executed on day one, which resulted in value 2 (ice cream), after END_DAILY_INTERACTION was conducted. With these mechanisms, in the YOU_LIKE_ICECREAM behavior, the robot says, "Last time, you said that you like ice cream. I asked about ice cream in this shopping mall and have a recommendation for you. On the fourth floor, …"

3) Nonverbal Behaviors: While situated descriptions are written in the *situated modules,* a separate controller exists for the generally-applicable low-level control of nonverbal behaviors that controls the robot's body orientation [see Fig. 9 (b)] and its gaze direction. For these functionalities, information from the position and face-tracking modules is used. The *situated modules* can restrict the execution of these nonverbal behaviors if needed; in other words, unless restricted, the robot generally performs these nonverbal behaviors.

The *situated modules* sometimes require the nonverbal behavior controller to perform gestures, such as pointing and gazing at, in a guiding behavior. This explicit control of gesture overrides the control of actuators usually used for the earlier nonverbal behaviors.

Reactive behaviors are implemented as well. When a person touches the tactile sensors of the robot, the robot directs its gaze to the part being touched and overrides the gaze control usually used to maintain eye contact. This functionality expresses its lifelikeness.

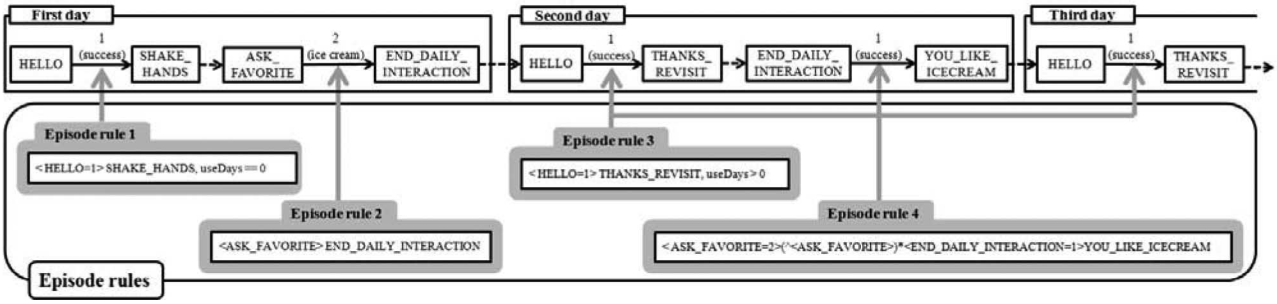

Fig. 8. Examples of transitions of behaviors.

Fig. 9. Typical sequence of robot's behavior. (a) Waiting for a human. (b) Finding a human. (c) Having a dialogue. (d) Ending the dialogue.

D. Design of Interactive Behaviors

1) General Design: We set two basic policies for designing the robot's interaction. First, it assumes the communication initiative, introduces itself as a guide robot, asks about places, and then provides information in response to user requests. This way, customers clearly understand that the robot is engaged in route guidance.

Second, the manner of its utterances and other behaviors are prepared in an affective manner [19], not in a reactive manner. The robot engages in human-like greetings, reports its "experience" with products in shops, and tries to establish a relationship (rapport) [32] with individuals. This is very different from master–slave-type communication, where a robot prompts a user to provide a command.

2) Guiding Behavior: Two types of behaviors were prepared for guiding: *route guidance* and *recommendation*. The former is a behavior in which the robot explains a route to a destination with utterances and gestures (see Fig. 3). The robot points in the first direction and says, "Please go that way" with an appropriate reference term chosen by an attention-drawing model [33]. It continues the explanation: "After that, you will see the shop on your right." Since the robot knows all of the mall's shops and facilities (restrooms, exits, parking lots, etc.), it can provide directions for 134 destinations.

In addition, for situations, where a user has not decided where to go, we designed *recommendation* behaviors in which the robot suggests restaurants and shops. For example, when a user asks, "Where is a good restaurant?" the robot starts a dialogue by asking a few questions, such as "What kind of food would you like?", and accordingly chooses a restaurant to recommend.

3) Rapport-Building Behavior: For persons wearing RFID tags, the robot starts to build rapport through a dialog that consists of the following three policies.

a) Self-disclosure: The importance of self-disclosure for humans to express friendship has long been studied. Bickmore and Picard used this strategy in relational agents for building relationships with users [32]. Gockley *et al.* made a receptionist robot that tells new stories and successfully attracted people to interact with it [11]. In our previous study, which was successful, our robot disclosed a secret [34]. In this study, we follow the same strategy: letting the robot perform self-disclosure. For example, the robot mentions its favorite food, "I like *takoyaki*," and its experiences, such as, "This is my second day working in this mall."

b) Explicit indication of person being identified: Since in our previous studies, we found that people appreciated having their names used by robots [12], we retained this strategy. The robot greets a person by the name under which he/she registered, such as "Hello, Mr. Yamada." In addition, it uses the history of previous dialogues to show that the robot remembers the person. For example, on day one, if the robot asked, "Do you like ice cream?" and the person answered "Yes," the robot says "OK, I'll remember that"; on day two, the robot says, "I remember that you said you liked ice cream, so today, I'm going to tell you my favorite flavor of ice cream."

c) Change in friendliness of behaviors: For a person who repeatedly visits the mall and the robot, the robot gradually changes its behavior to show a more friendly attitude. For example, on day one, it says, "I'm a little nervous talking with you for the first time," but on day three, it says, "I think we are friends" to show its warm attitude toward the person.

4) Behavior for Advertisement: The robot is also designed to provide advertisements about shops and products in a manner that resembles *word of mouth*. When the robot begins a conversation with a customer, it starts with a greeting and then engages in *word of mouth* behavior as a form of casual chat. It affectively

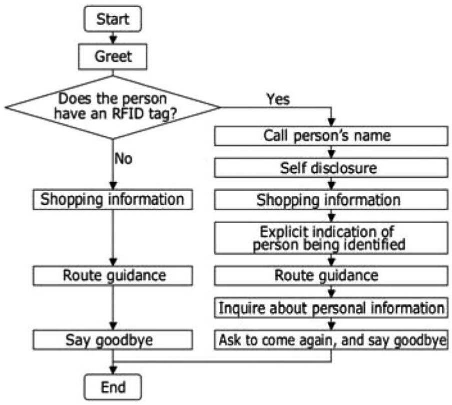

Fig. 10. Flow of robot's dialogue.

reports its pretended experiences about products in shops. For example, the robot might say, "Yesterday, I ate a crêpe in the food court. It was nice and very juicy. I was surprised!" or "The beef stew at Bombardier Jr. was good and spicy. The egg was really soft, too, which was also very nice." We implemented five topics per day and changed them every day so that daily shoppers did not get bored with this behavior.

5) Implementation of Behaviors and Episode Rules: The aforementioned behaviors are all implemented into *situated modules* and connected with the *episode rules*. They form an interaction flow as follows. As shown in Fig. 9, when no person is present, the robot remains in a wait mode. When a person steps on the floor sensor, the system detects this person-arriving event and transits the behavior from waiting behavior to greeting behavior. This starts a dialog. When the person leaves, it also detects this event and transits from interacting to waiting behavior.

Fig. 10 shows a simplified description of the overall structure of the interaction flow. Note that Fig. 10 is the summary, but not the details of the implementation, such as how a single sequence is formed. The actual transitions among behaviors are more complex and branched for variety, as well as for different conditions, as described by the episode rules. When the robot greets that person, if the person touches his/her RFID tag to the reader, the robot starts the flow in the first branch of Fig. 10. It calls the person's name, provides daily shopping information, chats about the person's preferences, and offers route guidance. An example of such a dialog is shown in Table VI. If the person does not have an RFID tag, it engages in the simpler interaction of providing shopping information and route guidance. Each behavior usually lasts about 5–15 s. A total of 1759 behaviors and 1015 *episode rules* were finally implemented.

TABLE VI
EXAMPLE OF A DIALOG

(Day one) R denotes robot, and H denotes a human
R: Hello, Ms. Suzuki.
R: Nice to meet you.
H: Nice to meet you.
R: Let's shake hands. (*They shake hands*)
R: Thank you.
R: Today, I ate a rice ball in the food court, which was hot and good!
R: By the way, my job is to provide route guidance.
R: I can guide you anywhere in AEON. (*Shopping mall*)
R: Where would you like to go?
H: I'm looking for an ATM.
R: Well, let's see. . .
R: Please go this (*pointing*) way, and you will see two ATMs on your right.
R: Do you understand?
H: Yes. Thank you.
R: By the way, there is a movie theater in the shopping mall. Have you been there before?
H: Yes.
R: Ok, I'll remember that.
R: Please come see me again. Goodbye!
H: Goodbye!

(Day two)
R: Hello, Ms. Suzuki.
R: Thank you for coming to see me again!
R: You came here on the way back from work the other day, right?
H: Yes.
R: Do you like movies?
H: Yes.
R: OK I'll remember that. Next time, I'd like to recommend some movies.
…

E. Operator Control

1) General Principle: Three types of control were enabled for the human operators: *substitute for speech recognition, supervisor of behavior selector,* and *knowledge provider*, all of which were controlled by the user interface shown in Fig. 11. In addition, we established an important principle for the operator: Minimize the number of operations. This principle was set to study the potential of robot autonomy. Except for substituting for speech recognition, the operator only helped the robot when intervention was truly needed. For example, even if a user interrupted the flow and asked, "How old are you?" (a frequent question), the operator did not operate the robot. If the user continued to repeat the question without showing any signs of stopping, the operator selected the robot's behavior or even typed an utterance to answer. When the operator needed a few seconds to make a decision and type an utterance, he/she executes a *conversational filler* behavior [35] to notify listeners that the robot will soon respond.

2) Substitute for Speech Recognition: When a robot performs a behavior in which it asks a question, the teleoperation system prompts the operator to choose words from the list expected for this situation. As shown in Fig. 11, "*substitute for speech recognition,*" the word list with buttons will be displayed on the right side of the interface. The displayed word list changes depending on the robot's behavior. By clicking the buttons, string information will be sent to the robot as a speech-recognition result. For example, when the robot asks, "I can give you the route. Where would you like to go?" the teleoperation system shows a list of places. When the robot asks, "Do you like ice cream?," it shows a simple choice of "yes," "no," and

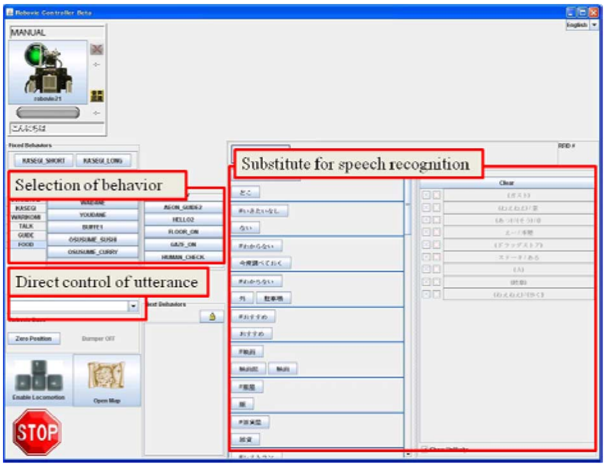

Fig. 11. User interface for robot control.

"no response" to the operator. Here, the operator behaves as a speech-recognition software. After the operator chooses the words, the robot autonomously continues the dialogue.

3) Supervisor of Behavior Selector: Significant degrees of uncertainty exist about user behavior toward the robot. Sometimes, people asked unexpected things, even though the robot has a behavior to answer the question; here, the problem is the lack of *episode rules*. The interface displays a list of behaviors in its upper left side (see Fig. 11), which is denoted with "selection of behavior." The operator can choose a behavior list by selecting each tab, which includes the behaviors of identical categories, such as guiding, greeting, etc. By clicking the buttons, information about the next behavior is sent to the robot; in other words, the operator behaves in the same way as a behavior selector function. For example, although the robot has behaviors to guide and explain all of the shoe stores, it was confused when a user asked about a "shop for children's shoes" because this phrase was not in the speech-recognition dictionary. In such situations, the operator selects the next behavior for the robot.

In case a person came up with a request out of context (e.g., requested to do handshaking while the robot gave information about the mall), first the operator would ignore and see whether the person followed the lead from the robot, and if this did not happen, the operator would then control the behavior selection to match the robot's behavior to the person's request. Note that during the field trial reported in this paper, we have observed that most of the time, people were willing to accept such situations, and were following the robot's conversation lead.

After this type of operation, developers manually updated the word dictionaries for speech recognition and the *episode rules* based on the operation histories so that the robot can autonomously select its next behavior for such a future request.

4) Knowledge Provider: With current technology, only humans can provide knowledge to the robot. Developers input knowledge in advance as a form of behavior, but this in-advance effort is limited to situations that the developers can imagine; in reality, much knowledge for unexpected activities is needed. To solve such problems, the interface provides a text-to-speech function for the operator, who can type the sentence in the interface's middle left (see Fig. 11) in a text box called the "direct control of utterances."

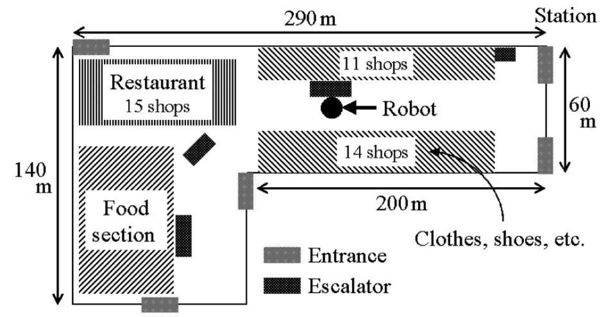

Fig. 12. Map of mall's main floor.

For example, although the robot has behaviors for all restaurants, when asked about a Japanese restaurant, the robot could not say something like, "There are two Japanese restaurants: a *sushi* restaurant and a *soba* restaurant. Which do you prefer?" For such a case, the operator directly typed the sentence so that the robot could respond. Later, developers manually added the appropriate behaviors for the situation.

5) Transition Between Autonomous and Teleoperated Mode: For the role of substitute of speech recognition, it is the system that transits the mode from autonomous to teleoperated, in a similar way as when the robot system itself sends a request to the autonomous speech recognizer. For the other two roles, supervisor of behavior selector, and knowledge provider, it is the operator, who transits the mode.

When these types of operation occur, the currently-running behavior is interrupted. In case the operator provides new knowledge, the system stays in the teleoperated mode, until the operator completes the direct control and chooses the next behavior to execute. After the operator chooses the next behavior, the system goes back to autonomous mode and keeps choosing the next behavior autonomously.

Note that such interruption of the currently-running behavior seemed to not have any negative effect; in this situation, the operator usually controlled the robot because the interacting person required something different from the currently-running behavior, thus people simply responded to the newly executed behaviors, since they were not interested in the currently running behavior.

IV. Field Trial

A field trial was conducted at a large, relatively new shopping mall consisting of three floors for shopping, one for parking, 150 stores, and a large supermarket. The robot was placed in a main corridor of the mall (see Figs. 12 and 13) weekday afternoons from 1 to 5 P.M. for five weeks (from July 23 to August 31, 2007, except for a busy holiday week in the middle of August). This schedule was set based on an agreement with the mall management to avoid busy times and, thus, prevent situations where too many people might crowd around the robot.

The robot was accessible to all visitors. Those who signed up for the field trial (participants) received a passive-type RFID

Fig. 13. Field trial environment.

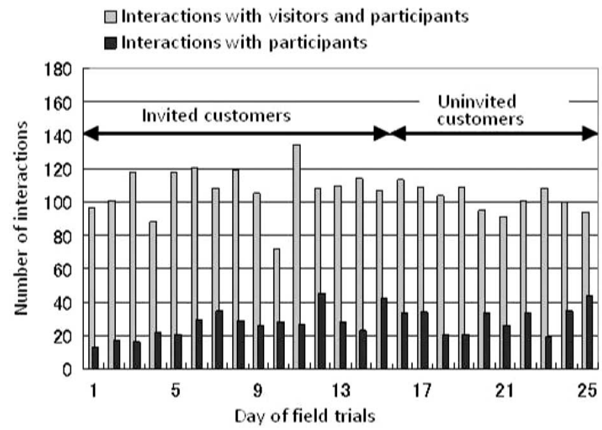

Fig. 14. Number of visitors and participants.

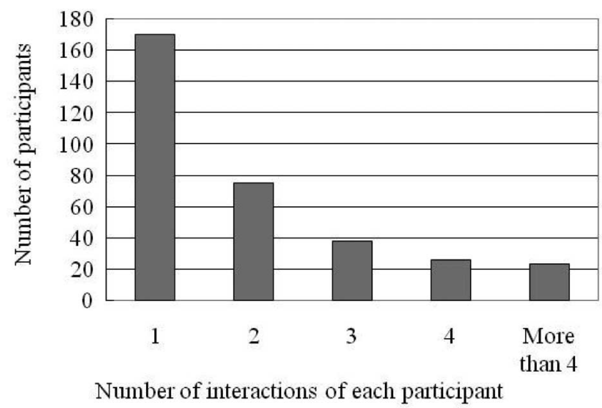

Fig. 15. Visits of each participant.

embedded in a cellphone strap (see Fig. 7). We recruited these participants by two methods: 1) a flyer distributed to residents around the mall and 2) on-site sign-up during the first three weeks while our staff approached visitors who seemed interested in the robot. The participants signed consent forms when they enrolled and filled out questionnaires after the field trial. They were not paid, but they were allowed to keep their RFID tags.[3]

The purpose of the field trial was to test the hardware and system design for the designated three roles in a real context. A questionnaire, which was designed to measure the usefulness of the robot for these roles, asked for impressions of the robot, the usefulness of the route guidance, and the effect of its advertisements.

V. Results

A. Observations

1) Transition of Interactions: Fig. 14 shows the number of interactions in which the robot engaged. One of them represents an interaction that continued with the visitor until the robot said goodbye. During the first three weeks, our staff invited visitors for registration and interaction with the robot. From the fourth week, our staff stood near the robot for safety. There were averages of 105.7 interactions each day. As the graph shows, the number of interacting persons did not differ over the five-week period. Multiple persons interacted with the robot (an average of 1.9 persons per interaction).

In all, 332 participants signed up for the field trial and received RFID tags; 37 of these participants did not interact with the robot at all. As shown in Fig. 15, 170 participants visited once, 75 visited twice, 38 visited three times, and 26 visited four times; the remaining 23 participants visited from five to 18 times. On average, each participant interacted 2.1 times with the robot, indicating that they did not repeat their interaction very much. One obvious shortcoming was the trial duration; since every day many nonparticipant visitors waited in line to interact with the robot, some reported that they hesitated to interact with it because of the crowds. Fig. 14 shows the number of participants who interacted each day: an average of 28.0 persons.

[3]The experimental protocol was reviewed and approved by our institutional review board.

2) Scenes of Interaction: Visitors who typically interacted with the robot only once or twice often passively interacted with it. They observed how the robot behaved and provided minimum response. For example, when the robot offered route guidance, they asked a simple destination, such as the restrooms. They answered yes or no to the robot's questions but did not give more than such simple answers. They seemed satisfied after they learned the capability of the robot, e.g., that it could point in the correct direction to the destination they asked about.

In contrast, some visitors interacted with the robot more than a few times, showing that they were absorbed in interacting with the robot. Such groups were mainly composed of children and mothers. As a case study, we describe some scenes of their interactions. We emphasize these cases because phenomena of interest often appear among people, who show such a large involvement. Various reasons might explain why people did not repeatedly visit the robot, including a lack of opportunity, crowds around the robot, and the lack of the robot's capabilities. In the future, if such a robot were placed in a daily environment for a long time, we believe that such cases, where people show a lot of involvement, will be more frequent, because we expect greater opportunities for interaction, less hesitation based on crowds, and better robot capabilities reflecting technological advancements.

Fig. 16. Child who behaved socially.

Fig. 17. Engagement with familiarization behavior.

Fig. 18. Interacting with family members.

Fig. 19. Revisiting with others.

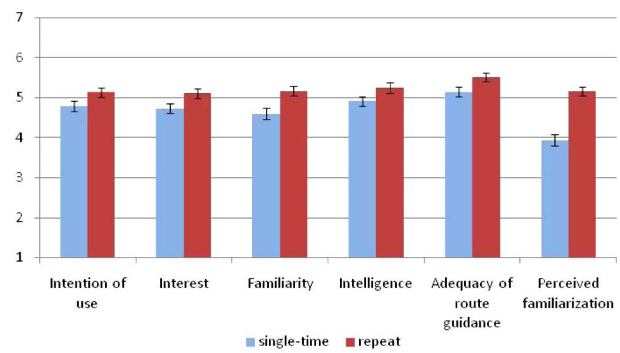

Fig. 20. Comparison of repeat and single-time visitors.

a) Case 1 (Engagement and social interaction): One notable case involved a boy, who behaved socially to the robot. Fig. 16 shows a scene when the boy first met the robot. When the robot greeted him with "Nice to meet you," the boy bowed. During the conversation, he provided many responses after the robot explained something. For instance, when the robot recommended visiting a *takoyaki* shop, the boy answered "OK!"

His engagement did not diminish after repetitions. Fig. 17 shows a scene when he visited the robot for the fourth time. When the robot said, "Let me tell you my secret," he stepped forward to the robot, showing his strong interest in the robot's secret. The boy visited six times during this field trial, and until the end, he showed such engagement in the interaction.

b) Case 2 (Interacting together): Few people interacted with the robot alone; instead, they came with family members. The second case involves a mother and a young girl, who visited six times. There were a number of scenes where we observed their interaction together. For example, when the robot said "Hello," the mother and the girl both simultaneously responded with "Hello." Fig. 18 shows a scene during the second visit when the robot said, "Please touch me" (left). Both stretched their hands out to the robot's head and patted it (right).

c) Case 3 (Revisiting with others): Similar to Case 2, another interesting case is a lady, who visited five times. On the first time, she visited the robot alone (see Fig. 19, left). However on the second time, she brought two young boys, who appeared to be her grandsons and showed them how to interact with the robot (see Fig. 19, right). We also observed another similar case, where a lady brought her friends the next time.

B. Visitors' Perception

1) Perception of Participants: When the field trial was finished, we mailed questionnaires to the 332 participants and received 235 answers. All items were on a 1-to-7 point scale, where 7 is the most positive, 4 is neutral, and 1 is the most negative.

a) Impressions of robot: The questionnaire included items about "intention of use (studied in [36])," "(degree of) interest," "familiarity," and "intelligence," all of which resulted in scores around 5.0 (see Fig. 20). Many positive, free-answer comments described the robot as cute and friendly.

b) Route guidance: The questionnaire answers about the adequacy of route guidance averaged 5.3 points. The following free-answer comments were made.

1) The robot correctly answered when I asked about a particular shop.
2) I am surprised that its route guidance was so detailed.
3) Its route guidance was appropriate and very easy to understand.
4) The robot was useful for questions that I hesitated to ask a person because they seemed too simple.

c) Providing information: The questionnaire answers about the usefulness and interest in the information averaged 4.6 and 4.7 points, respectively. Ninety-nine of 235 participants reported that they visited a shop mentioned by the robot, and 63 participants bought something based on the information provided by the robot. We particularly asked about reasons in a free-description question and received the following comments.

1) The robot recommended a kind of icecream that I had not eaten before, so I tried it.
2) The movie mentioned by the robot sounded interesting.
3) Since Robovie repeatedly mentioned crêpes, my child wanted to eat one.

These results suggest that the robot's information-providing function affected them, increased their interest in particular shops and products, and even encouraged them to actually buy products.

d) Building rapport: The questionnaire answers about the degree of perceived familiarization averaged 4.6 points. In the free-description form, comments included as following.

1) Since it said my name, I felt the robot was very friendly.
2) The robot was good, since it gradually became familiar with me.
3) I am surprised that the robot had such a good memory.
4) My child often said, "Let's go to the robot's place," and this made visiting the mall more fun.
5) The robot was very friendly. I went with my five-year-old daughter to interact with the robot; on the last day, she almost cried because it was so sad to say goodbye. She will remember it as an enjoyable event: at home, she imitates the robot's behavior and draws pictures of it.

2) Comparison of Repeat and Single-Time Visitors: To study how effectively the robot functioned in the building rapport role, we compared repeat and single-time visitors. We classified the 235 participants, who returned questionnaires based on whether they visited the robot more than once. An analysis of variance (ANOVA) was conducted for the ratings of their impressions (see Fig. 20). There were significant or almost significant differences [$F(1233) = 4.32$ ($p < 0.05$), 4.62 ($p < 0.05$), 9.98 ($p < 0.01$), 3.56 ($p < 0.1$), 5.44 ($p < 0.05$), and 49.39 ($p < 0.01$)]. Repeat visitors had better impressions; a particularly notable result is the difference about the perceived familiarization, which is largely high for the repeat visitors.

3) Comparison With an Information Display: We asked participants how often they were influenced by information displays in the same mall (see Fig. 2). In the questionnaires, participants answered the following: "Usefulness of information provided by display/robot," "Interest in shops mentioned by display/robot," "Visiting frequency triggered by display/robot," and "Shopping frequency triggered by display/robot." The order of the questions about the display and robot was counterbalanced.

Fig. 21 shows the comparison results. There were significant differences ($F(1229) = 40.96$, 69.52, 36, 19, and 7.66, $p < 0.01$ for all four items). Thus, for the participants, the robot provided more useful information and elicited more shopping.

4) Integrated Analysis:

a) Structural-equation modeling: Structural-equation modeling (SEM) is a relatively new statistical-analysis method for revealing the relationships behind observed data. We analyzed the relationships among impression, the perceived usefulness, and the effect on shopping behavior using SEM because it enables us to analyze relationships among multiple variables and analyze cause-and-effect relationships as well.

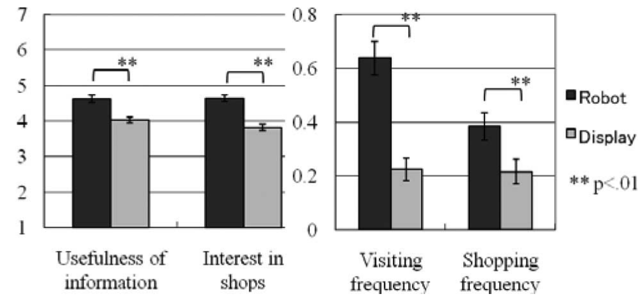

Fig. 21. Comparison of robot and display.

Unlike controlled experiments in a laboratory, where we simplify related variables, in a field trial, we need to retrieve knowledge from complex variables.

Its process resembles factor analysis to reveal latent variables and regression analysis to associate variables to produce a graphical model of causal-result relations. The analysis consists of the following steps.

1) *Modeling:* Run the following step to find a model that reasonably explains the observed variables.
   a) *Make a hypothesized model with latent variables:* Based on the theoretical or hypothetical relationships among observed variables, make a model, where latent variables are placed among the variables.
   b) *Calculate path coefficients:* Based on the given model, path coefficients are calculated in a similar manner as factor and regression analysis.[4]
   c) *Confirm the fitness of paths:* For all paths in the model, test whether they are significant.
   d) *Confirm the model's fitness:* Indicators of the model's fitness include the goodness-of-fit index (GFI), the adjusted GFI (AGFI), the comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA). According to a textbook [38], the desired range of the indicators should be as follows: GFI, AGFI ≥ 0.90, CFI ≥ 0.95, and RMSEA ≤ 0.05.
2) *Analyze the best-fit model:* Perhaps, more than one model can explain the observed variables. There is an indicator Akaike information criterion (AIC) for the best fitness of this model. The model with the minimum AIC value is considered the best among the models with identical variables.

More detailed explanation can be found in many textbooks [37], [38]. The following paragraphs report how we applied this technique to our data.

b) Analysis results: For the modeling, our hypothesis states that participant interaction experiences with the robot (observed as impression and day of visit) affected their shopping behavior as an advertisement effect. Our model included the latent variables of advertisement and interest effects as possible consequences. We added the latent variables of the impression

[4]A statistic software package usually runs this process. We used Amos 6.

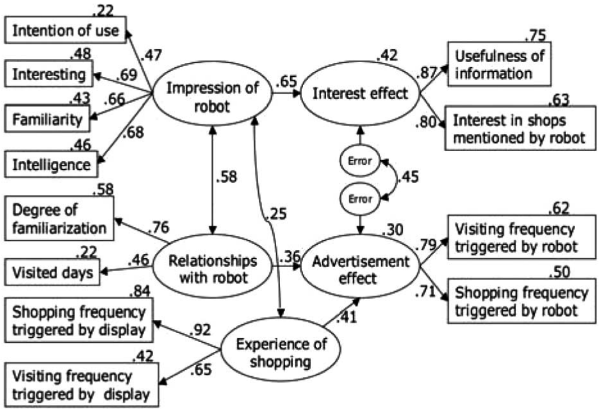

Fig. 22. Results of integrated analysis: How participant impressions related with shopping behaviors.

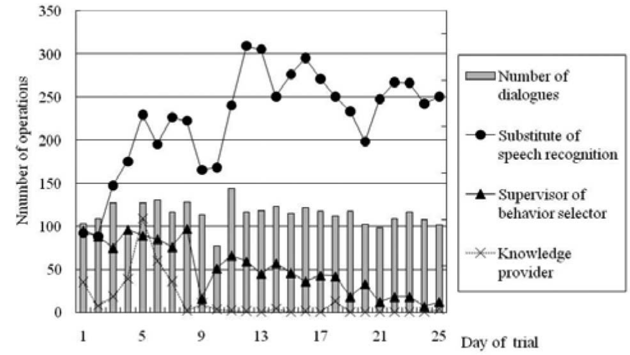

Fig. 23. Operations by operator.

of the robot, the rapport established with it, and the experience of shopping as possible causal factors.

Fig. 22 shows the best-fit model produced by SEM. In Fig. 22, for readability, we did not draw the error variables that are associated with only one variable. The variables in the squares are the observed variables (such as the questionnaire items), and those in the circles are the latent variables retrieved by the analysis (named by us). The numbers around the arrows (paths) are the values of the path coefficients, similar to coefficients in regression analysis. The numbers on the variables show the coefficient of determination R^2. Thus, 30% of the "advertisement effect" is explained by the factors of "relationships with robot" and "experience of shopping," and 42% of the "interest effect" is explained by the "impression of robot" factor.

Regarding the model's validity, the analysis results show good fitness in the appropriateness indicators of GFI = 0.957, AGFI = 0.931, CFI = 0.987, and RMSEA = 0.028. Each path coefficient is significant at a significance level of 1%. It has the minimum AIC value of 115.9 among the other possible models.

Other models do not show appropriate fitness and have larger AIC values. For example, a model with one extra path from "impression of robot" to "advertisement effect" results in an AIC value of 116.9, and this path itself is not significant (coefficient = −0.10, p = 0.36). This suggests that "advertisement effect" is not directly affected by "impression of robot." Models with different latent variables show less desirable fitness. For example, if we only put two latent variables on the left side, merging "impression of robot" and "relationships with robot," the analysis produces less appropriate fitness indicators GFI = 0.937, AGFI = 0.900, CFI = 0.941, and RMSEA = 0.060, and a larger AIC value of 143.5.

c) Interpretation: The obtained model leads to an interesting interpretation and suggests that the participants who positively evaluated the impression of the robot tended to be positive about the interest effect (coefficient = 0.65); however, the advertisement effect is not associated with impression of the robot, but with the relationships formed with it (coefficient = 0.36). Therefore, the factor of the participant's relationship with the robot explains 13% of the deviation of the advertisement effect. Although this ratio might not be so high, we believe that it is interestingly high for such shopping behavior, since shopping behavior largely depends on people's situations, including disposable income, interests, time, and the occasion. This implies that the development of relationships with the robot would increase the advertisement effect. Although for improving relationships, impressions could be important.

C. Operator Involvement

1) Number of Operations: Since this study was conducted with a human operator, it is useful to show how often the robot was under his/her control. Fig. 23 shows the number of operations. As described in Section III-B, one operator role was to "substitute for speech recognition," which we expect to be automated in the future. The operator did this two or three times per dialog.

In contrast, the results show that the operator's loads for the remaining two roles, i.e., "supervisor of behavior selector" and "knowledge provider," were relatively small, particularly after day 10. Up to day 10 (see Fig. 23), the number of operations was sometimes large due to unimplemented features (explained in the following). After day 10, 254.2 "substitute for speech recognition," 1.7 "knowledge provider," and 13.4 "supervisor of behavior selector" operations were conducted per day. The amount of operation in terms of provided information from the operator [16] averaged 1 byte per 24.8 s.

We believe that these results are promising. We expect that the role of "substitute for speech recognition" will disappear in the near future with advances in speech-recognition technology. In contrast, the remaining two roles will be difficult to perform autonomously, but the number of operations related to these two roles was relatively small.

2) Transition of Operations With Additionally Implemented Behaviors/Rules: During the field trial, we continued to implement new interactive behaviors to supplement the missing knowledge that the operator needed. On average, we added 0.2 interactive behaviors per day to reduce the "knowledge provider" task and 3.4 rules per day for transition among behaviors to reduce the "supervisor of behavior selector" task.

This continued implementation of new interactive behaviors gradually decreased the number of operations related to the "supervisor of behavior selector" and "knowledge provider" roles. Such improvement shows one promising case of robot development that occurred in a real field under the supervision of a human operator.

VI. Discussion

A. Implications on Hardware and System Design

To cope with the real-world difficulties, we adopted a network-robot-system approach, where the lack of a single robot's capabilities was supplemented by ubiquitous sensors and a human operator.

For *person identification*, we used passive-type RFID tags. Based on the person-identification function, as intended, the robot successfully established a rapport. Many people commented that they appreciated its function of saying their names, and the questionnaire results showed that it provided familiarized impressions for repeat visitors. Concerns might arise that using RFIDs distracts visitors from human-like interactions with the robot; however, the participant questionnaire comments revealed no such reactions or complaints about touching the RFID or deviation from a human-like manner. For further development implications, although we believe that this was a reasonable choice for now, in the near future, other methods will probably be available that work in a human-like manner; when available, such a method that replaces passive-type RFID tags will increase the robot's perceived intelligence.

We used floor sensors for *positioning*. This was also effective. They enabled the robot to correctly identify the existence of the interacting person and the robot's position so that it could provide accurate route guidance. Another interesting characteristic also surfaced. The floor sensors created a visually separated region around the robot that resembled the conversation distance of people; therefore, it was clear for the visitors, where they could interact with the robot. This feature benefited our case, since we designed the robot to stay in the same place to avoid large crowds around it. However, for other settings, this might be a negative feature. For example, since it would be natural for the robot to move around an environment; floor sensors would simply limit its locomotion area. It is not reasonable to cover an entire environment with floor sensors. Instead, we need to use an alternative positioning device, such as one with laser range finders [39].

B. Implications on Semiautonomous System Design

Teleoperation is often used when technology remains immature. Robots designed for space exploration or search and rescue also use teleoperation techniques. We believe that our study reveals the possibility of using teleoperation to realize useful communication robots for daily life. However, this does not diminish the importance of autonomy. Teleoperated robots need a highly autonomous system; otherwise, teleoperation will become too complex for real-time operation. In fact, we prepared autonomous systems for finding people, identifying them, and controlling the dialog flow, the verbal and nonverbal behaviors, and the actuations. The experimental results indicate that the operator only provided 1 byte of command per 24.8 s.

One of the biggest task loads for the operator was the substitution for speech recognition, which will be replaced with a fully autonomous system in the near future. When robots can be equipped with such a robust speech-recognition system, they will probably operate autonomously in tasks similar to that reported in this paper almost all the time. Consequently, the achievement of this level of speech recognition would make an autonomous robot nearly practical for our purpose.

Meanwhile, we wonder whether we can create truly autonomous robots for such services as information-providing and route guidance. In our study, an operator helped us to prepare additional information for the robot's use. In particular, in the beginning, we experienced difficulty completely predicting communication patterns that visitors would prefer and information visitors would need. The required knowledge will always change over time, e.g., new shops will open and new trends will emerge that will make visitors look for new products. Consequently, we believe that value remains in having a monitoring person behind the robot system, who can control the flow of conversation and supplement information when the robot lacks up-to-date information.

Concern might also surface whether an operator needs to constantly monitor the robot to check that its knowledge is sufficient. As in other shared autonomy studies in robotics, we need to shift the mode of teleoperation from full teleoperation to adjustable autonomy [40] or sliding autonomy [41], where the system takes a greater initiative in using human support. We have started to explore such semiautonomous control by modeling a critical part of the interaction to be supervised [25] and estimating error situations [23]. We probably need to develop more sophisticated techniques to actively detect error states in conversation, where the robot needs extra knowledge, and summon human operator support.

C. Implications of the Role of the Robot

One of this study's challenges was to assign three different roles to the robot: guiding, building rapport, and advertising. While guiding is beneficial for visitors, advertising is more beneficial for the mall. Do these roles conflict? Did visitors accept such a robot?

The participant comments we received included no complaints about this situation. In the design, we clarified that the robot is affiliated with the shopping mall. It physically stays in the mall, has professional knowledge about the mall, and welcomes participants as a representative of the mall. Its role resembles that of the shopping mall's personnel.

In the advertisement behavior, the robot gave word-of-mouth recommendations and talked about food and restaurants, even though the robot actually did not have such experiences. Did people believe that the robot ate something? Of course not. However, regarding such information purportedly based on its "personal experience," people were not distracted by a robot that talked about eating and restaurants. No comments were made about any problems about the food and restaurant information. People just seemed to accept what the robot said. They knew it was a robot. Their reactions to animation characters who talk about food and restaurants would probably be similar. At the same time, some people reported that they were influenced by the information provided by the robot to visit shops.

D. Implications of Ethical Issue

First, this study was conducted as an academic trial to observe what happens when robots behave under such conditions. People, often with children, engaged with a friendly robot in order to entertain them and to get information, which, in turn, elicited people to do more shopping. These facts must be considered carefully when people design such robots for commercial purposes. Since ethical judgments reflect societal standards, guidelines must be created before robots start to perform such roles. This study itself could be used as evidence when people discuss ethical issues. Does this resemble a "subliminal" effect? We do not think so since everything that happens in the interactions with the robot is perceivable. Of course, current guidelines for similar issues, e.g., TV commercials, can probably be applied to robots, e.g., they should not lie.

We also note that the robot is clearly not owned by the users but is affiliated with the mall. Usually, in HCI studies, ethical issues become a problem when people's devices (owned by them) start to behave against them, e.g., send their personal information elsewhere unintentionally. In case with a robot in a mall, this is not a case.

E. Novelty Effect: Is It Worrisome?

Since robots remain very novel to the general public, perhaps the robot's effects on shopping activities only arose due to the novelty effect. We believe that such a novelty effect is reflected in the experimental results; however, it is impossible to eliminate such effects in today's world: There is no place where robots are just ordinary. Until robots become widely distributed, it is difficult to see how people perceive them without having a novelty effect. Even if one person is well habituated to a robot, many other inexperienced people will also see the robot. Instead, we believe that the novelty effect can be calculated along with the effect of using robots in advertisements. People like something new.

Moreover, we believe that we should study how robots affect people despite the existence of the novelty effect. Since robots are novel to people, they do not know whether robots can assist them. We need to study how robots are perceived as such a new entity. One participant made an interesting comment: He prefers using the robot to retrieve route guidance, since he hesitates to ask such a simple question to a human shopkeeper. Such comments might indicate how people perceive robots beyond the simple novelty effect.

F. Perspective for Better Communication Robots in Public Spaces

Although our developed robot successfully operated the route-guidance service, two major remaining problems must be solved in the future. First, the robot did not fully support people's motivation to repeatedly visit it. The study failed to show whether the robot could elicit spontaneous repeated interactions; a limited number of participants visited repeatedly. This might be due to the limited duration of the field trial or to the robot's limited capability. This aspect should be further explored in future studies.

Second, the robot's task was limited to a simple one. The complexity of the required knowledge will increase if the robot's task includes more than just providing route guidance and recommendations. For example, a shopkeeper robot, in a particular store, would need much greater information that incorporates recent trends in people's daily lives.

G. Limitations

The ability to generalize the field-trial findings is limited for several reasons. First, as discussed earlier, the novelty effect encourages people to visit the robot. This factor is unavoidable and might limit the validity of the findings in the future. In addition, since the findings were obtained with Japanese participants, they might not be directly applicable to other countries.

Second, since the participants in the field trial were self-selected and the questionnaire results were mailed, a mental bias might exist in the evaluation of the robot performance. Nevertheless, we believe that reporting the findings is useful, since what we can investigate in field trials is restricted. Note that we informed all participants that they were required to respond to this questionnaire after the field trial as a condition for participation; thus, the responses were more than simply positive reactions to the robot. In the future, we should verify the effects of each design in a controlled laboratory study.

The comparison of the display and this robot was conducted in a quasi-experimental approach, since this study was conducted in a field, where we could not fully control all relevant factors. Therefore, the following factors varied between conditions.

1) *Duration of comparison:* We asked about participant experiences during the field trial for the robot; for display, we asked experiences during the four months (from the mall's opening until the field trial's end). This arrangement was to include display's possible novelty effect (it could be novel to participants, since it is very large).
2) *Way of providing information:* The display shows information about a shop by highlighting its particular information. The target shop is switched about once a minute.
3) *Participant interest:* The participants might be more interested in the robot than other mall visitors, since participation in the field trial most likely reflected interest in the robot. However, this is a limitation of our study as a field trial, which needed spontaneous participation; for example, participants were required to register for the RFID tags. Nevertheless, we believe the comparisons are still useful to understand the phenomena elicited by the robot.

VII. Conclusion

This paper reports on a study of a communication robot for a shopping mall. We developed a robot that provides route guidance and other shopping information for visitors of a shopping mall. A five-week field trial was conducted, where the robot engaged in total 2642 interactions, with 235 participants signed up to use RFID tags for repeated interaction.

The study provided findings covering a series of topics from design consideration to user feedback.

1) It gave an example of societal roles of a communication robot in shopping mall, i.e., guiding and advertisement.
2) It demonstrated proof-of-concept of a network robot system, i.e., a robot system connected with ubiquitous sensors and a human operator, in a real everyday context.
3) It showed successful use of behavior-based software architecture for a communication robot.
4) It reported on successful behavior design for guiding with deictic gestures, rapport-building through repeated interaction, and word-of-mouth-type advertizing.
5) It demonstrated one feasible path of deploying a communication robot, where the results of teleoperation inform iterative developments, resulting in the gradual decrease of the amount of teleoperation.
6) It provided field data about how people interact with such a robot.
7) It empirically showed that people were influenced by the robot for shopping behavior.

Acknowledgment

The authors would like to thank the administrative staff at the Takanohara AEON store for their cooperation. They would also like to thank Dr. Akimoto, Dr. Miyashita, Dr. Sakamoto, Mr. Glas, Mr. Tajika, Mr. Nohara, Mr. Izawa, and Mr. Yoshii for their help.

References

[1] T. Mukai, M. Onishi, T. Odashima, S. Hirano, and Z. Luo, “Development of the tactile sensor system of a human-interactive robot ‘RI-MAN,” *IEEE Trans. Robot.*, vol. 24, no. 2, pp. 505–512, Apr. 2008.

[2] T. Takeda, Y. Hirata, and K. Kosuge, “Dance partner robot cooperative motion generation with adjustable length of dance step stride based on physical interaction,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 3258–3263.

[3] T. Spexard, M. Hanheide, and G. Sagerer, “Human-oriented interaction with an anthropomorphic robot,” *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 852–862, Oct. 2007.

[4] J. G. Trafton, N. L. Cassimatis, M. D. Bugajska, D. P. Brock, F. E. Mintz, and A. C. Schulz, “Enabling effective human-robot interaction using perspective-taking in robots,” *IEEE Trans. Syst., Man, Cybern., Part A: Syst. Humans*, vol. 35, no. 4, pp. 460–470, Jul. 2005.

[5] R. Stiefelhagen, H. K. Ekenel, C. Fügen, P. Gieselmann, H. Holzapfel, F. Kraft, K. Nickel, M. Voit, and A. Waibel, “Enabling multimodal human-robot interaction for the karlsruhe humanoid robot,” *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 840–851, Oct. 2007.

[6] B. Mutlu, J. Forlizzi, and J. Hodgins, “A storytelling robot: Modeling and evaluation of human-like gaze behavior,” in *Proc. IEEE Int. Conf. Humanoid Robots*, 2006, pp. 518–523.

[7] B. Scassellati, “Investigating models of social development using a humanoid robot,” in *Biorobotics*, B. Webb and T. Consi, Eds. Cambridge, MA: MIT Press, 2000.

[8] W. Burgard, A. B. Cremers, D. Fox, D. Hänel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, “The interactive museum tour-guide robot,” in *Proc. Nat. Conf. Artificial Intell.*, 1998, pp. 11–18.

[9] R. Siegwart, K. O. Arras, S. Bouabdallah, D. Burnier, G. Froidevaux, X. Greppin, B. Jensen, A. Lorotte, L. Mayor, M. Meisser, R. Philippsen, R. Piguet, G. Ramel, G. Terrien, and N. Tomatis, “Robox at Expo.02: A large scale installation of personal robots,” *Robot. Autonomous Syst.*, vol. 42, no. 3, pp. 203–222.

[10] A. Bauer, K. Klasing, G. Lidoris, Q. Mühlbauer, F. Rohrmüller, S. Sosnowski, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss, “The autonomous city explorer: Towards natural human-robot interaction in urban environments,” *Int. J. Soc. Robot.*, vol. 1, no. 2, pp. 127–140, 2009.

[11] R. Gockley, J. Forlizzi, and R. Simmons, “Interactions with a moody robot,” in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, 2006, pp. 186–193.

[12] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, “Interactive robots as social partners and peer tutors for children: A field trial,” *Human Comput. Interaction*, vol. 19, no. 1–2, pp. 61–84, 2004.

[13] K. Wada and T. Shibata, “Living with seal robots—Its sociopsychological and physiological influences on the elderly at a care house,” *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 972–980, Oct. 2007.

[14] K. Dautenhahn and I. Werry, “A quantitative technique for analysing robot-human interactions,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2002, pp. 1132–1138.

[15] H. Kozima, C. Nakagawa, and Y. Yasuda, “Interactive robots for communication-care: A case-study in autism therapy,” in *Proc. IEEE Int. Symp. Robots Human Interactive Commun.*, 2005, pp. 341–346.

[16] F. Tanaka, A. Cicourel, and J. R. Movellan, “Socialization between toddlers and robots at an early childhood education center,” *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 46, pp. 17954–17958, 2007.

[17] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, “An affective guide robot in a shopping mall,” in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, 2009, pp. 173–180.

[18] Research study for the scope of the strategy map of robotics technology, 2005. (Available at http://www.nedo. go.jp/ database/index.html, with index code 100007875) (in Japanese).

[19] R. W. Picard, *Affective Computing*. Cambridge, MA: MIT Press, 1997.

[20] T. Kanda, D. F. Glas, M. Shiomi, H. Ishiguro, and N. Hagita, “Who will be the customer?: A social robot that anticipates people’s behavior from their trajectories,” in *Proc. 10th Int. Conf. Ubiquitous Comput.*, 2008, pp. 380–389.

[21] A. Sanfeliu, N. Hagita, and A. Saffiotti, “Special issue: Network robot systems,” *Robot. Autonomous Syst.*, vol. 56, pp. 793–797, 2008.

[22] C. T. Ishi, S. Matsuda, T. Kanda, T. Jitsuhiro, H. Ishiguro, S. Nakamura, and N. Hagita, “A robust speech recognition system for communication robots in noisy environments,” *IEEE Trans. Robot.*, vol. 24, no. 3, pp. 759–763, Jun. 2008.

[23] M. Shiomi, D. Sakamoto, T. Kanda, C. T. Ishi, H. Ishiguro, and N. Hagita, “A semi-autonomous communication robot—A field trial at a train station,” in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, 2008, pp. 303–310.

[24] A. Green, H. Huttenrauch, and K. S. Eklundh, “Applying the wizard-of-oz framework to cooperative service discovery and configuration,” in *Proc. IEEE Int. Workshop Robots Human Interactive Commun.*, 2004, pp. 575–580.

[25] D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, “Simultaneous teleoperation of multiple social robots,” in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, 2008, pp. 311–318.

[26] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, “Development and evaluation of interactive humanoid robots,” in *Proc. IEEE*, 2004, vol. 92, pp. 1839–1850.

[27] H. Kawai, T. Toda, J. Ni, M. Tsuzaki, and K. Tokuda, “XIMERA: A new TTS from ATR based on corpus-based technologies,” in *Proc. 5th ISCA Workshop Speech Synthesis*, 2004, pp. 179–184.

[28] T. Tajika, T. Miyashita, H. Ishiguro, and N. Hagita, “Automatic categorization of haptic interactions—What are the typical haptic interactions between a human and a robot?,” in *Proc. IEEE Int. Conf. Humanoid Robots*, 2006, pp. 490–496.

[29] T. Murakita, T. Ikeda, and H. Ishiguro, “Human tracking using floor sensors based on the markov chain monte carlo method,” in *Proc. Int. Conf. Pattern Recognition*, 2004, pp. 917–920.

[30] M. Shiomi, T. Kanda, N. Miralles, T. Miyashita, I. Fasel, J. Movellan, and H. Ishiguro, “Face-to-face interactive humanoid robot,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep., 2004, pp. 1340–1346.

[31] S. Koizumi, T. Kanda, M. Shiomi, H. Ishiguro, and N. Hagita, “Preliminary field trial for teleoperated communication robots,” in *Proc. IEEE Int. Symp. Robots Human Interactive Commun.*, 2006, pp. 145–150.

[32] T. W. Bickmore and R. W. Picard, “Establishing and maintaining long-term human-computer relationships,” *ACM Trans. Computer-Human Interaction*, vol. 12, no. 2, pp. 293–327, 2005.

[33] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, “Human-like conversation with gestures and verbal cues based on a three-layer attention-drawing model,” *Connection Sci.*, vol. 18, no. 4, pp. 379–402, 2006.

[34] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "A two-month field trial in an elementary school for long-term human-robot interaction," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 962–971, Oct. 2007.
[35] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, "How quickly should communication robots respond?," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, 2008, pp. 153–160.
[36] M. Heerink, B. Kröse, B. Wielinga, and V. Evers, "Enjoyment, intention to use, and actual use of a conversational robot by elderly people," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, 2008, pp. 113–119.
[37] H. Toyoda, *Structural Equation Modeling*. Tokyo, Japan: Tosho, 2007 (in Japanese).
[38] D. W. Kaplan, *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: Sage, 2000.
[39] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Laser-based tracking of human position and orientation using parametric shape modeling," *Adv. Robot.*, vol. 23, pp. 405–428, 2009.
[40] M. A. Goodrich, T. W. McLain, J. D. Anderson, J. Sun, and J. W. Crandall, "Managing autonomy in robot teams: Observations from four experiments," in *Proc. ACM/IEEE 2nd Annu. Conf. Human-Robot Interaction*, 2007, pp. 25–32.
[41] B. P. Sellner, L. M. Hiatt, R. Simmons, and S. Singh, "Attaining situational awareness for sliding autonomy," in *Proc. ACM/IEEE 1st Annu. Conf. Human-Robot Interaction*, 2006, pp. 80–87.

Takayuki Kanda (M'04) received the B.Eng, M.Eng, and Ph. D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively.

From 2000 to 2003, he was an Intern Researcher with the Advanced Telecommunications Research Institute International (ATR) Media Information Science Laboratories. He is currently a Senior Researcher with the ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. His research interests include intelligent robotics, human–robot interaction, and vision-based mobile robots.

Dr. Kanda is a member of the Association for Computing Machinery, the Robotics Society of Japan, and the Information Processing Society of Japan.

Masahiro Shiomi received the 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 with the Intelligent Robotics and Communication Laboratories, Kyoto, Japan, where he is currently a Researcher with the Advanced Telecommunications Research Institute International. His research interests include human–robot interaction, interactive humanoid robots, networked robots, and field trials.

Zenta Miyashita received the M.Eng. degree in engineering from Osaka University, Osaka, Japan, in 2008.

From 2005 to 2008, he was an Intern Researcher with the Intelligent Robotics and Communication Laboratories, Kyoto, Japan, where he is currently with the Advanced Telecommunications Research Institute International.

Hiroshi Ishiguro (M'01) received the D.Eng. degree from Osaka University, Osaka, Japan, in 1991.

In 1991, he was a Research Assistant with the Department of Electrical Engineering and Computer Science, Yamanashi University, Yamanashi, Japan. In 1992, he joined as a Research Assistant with the Department of Systems Engineering, Osaka University. In 1994, he became an Associate Professor with the Department of Information Science, Kyoto University, Kyoto, Japan, where he was engaged in research on distributed vision using omnidirectional cameras. From 1998 to 1999, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of California, San Diego. In 1999, he was a Visiting Researcher with the Advanced Telecommunications Research Institute International (ATR) Media Information Science Laboratories, where he developed Robovie, which is an interactive humanoid robot. In 2000, he was an Associate Professor with the Department of Computer and Communication Sciences, Wakayama University, Wakayama, Japan, where he became a Professor in 2001. He is currently a Professor with the Department of Adaptive Machine Systems, Osaka University, and a Group Leader with the ATR Intelligent Robotics and Communication Laboratories, Kyoto.

Norihiro Hagita (M'85–SM'99) received the 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, Japan. He joined the Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan, to establish the ATR Media Information Science Laboratories and the ATR Intelligent Robotics and Communication Laboratories in 2001 and 2002, respectively. His current research interests include communication robots, network robot systems, interaction media, and pattern recognition.

Dr. Hagita is a Fellow of the Institute of Electronics, Information, and Communication Engineers, Japan. 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.