International Journal of Social Robotics (2020) 12:549–566
https://doi.org/10.1007/s12369-019-00587-y

ORIGINAL RESEARCH

Human-Like Guide Robot that Proactively Explains Exhibits

Takamasa Iio[1] · Satoru Satake[2] · Takayuki Kanda[3] · Kotaro Hayashi[4] · Florent Ferreri[2] · Norihiro Hagita[2]

Accepted: 31 August 2019 / Published online: 13 September 2019

Abstract
We developed an autonomous human-like guide robot for a science museum. Its identifies individuals, estimates the exhibits at which visitors are looking, and proactively approaches them to provide explanations with gaze autonomously, using our new approach called speak-and-retreat interaction. The robot also performs such relation-building behaviors as greeting visitors by their names and expressing a friendlier attitude to repeat visitors. We conducted a field study in a science museum at which our system basically operated autonomously and the visitors responded quite positively. First-time visitors on average interacted with the robot for about 9 min, and 94.74% expressed a desire to interact with it again in the future. Repeat visitors noticed its relation-building capability and perceived a closer relationship with it.

Keywords Human–robot interaction · Service robots

1 Introduction

Guidance and information-providing are promising services for social robots. Many robots have already been developed for daily contexts, such as museums [1, 2], exposition [3], reception [4], shops [5], stations [6], and shopping malls [7]. Since the effect of such human-like behavior as gaze and pointing behavior in information-providing is well known in HRI studies, recent guide robots often interact with users in human-like ways.

In contrast, it remains an open question how to successfully create human-like, autonomous guide robots. Since many robots are operated in actual environments in the daily contexts of humans [1–7], they have to face actual environment challenges. Automatic speech recognition (ASR) is still a complicated problem in such noisy environments. In fact, although previous studies used GUI [2, 5], keyboard input [4], and a human wizard (operator) [6, 7], ASR still generally fails. For instance, ASR that is prepared for a noisy environment only worked with 20% success in a real environment [6]. The locations of visitors were successfully used [2–8], and people were sometimes identified by RFID tags [4, 7]. However, if we use non-spoken input, like keyboards or GUIs, interaction is limited (e.g., areas where users can stand are restricted) and diverges from interaction that resembles that of humans. Must we wait for complete ASR capability for human-like robots?

In order to expand HRI studies, we should challenge to develop a human-like guide robot.

As Pateraki and Trahanias point out [9], such a robot need to have not only navigation functions such as localization, collision avoidance and path planning, but also functions to interact naturally with humans. The functions are, for example, detecting and tracking humans, estimating human attention, approaching to a human in humanlike manner, and providing appropriate information, etc. These functions have been studied so far in the field of HRI. However, as far as we know, there are few human-like guide robot systems into which those functions are integrated in a real environment. Therefore, in this study, we develop such a human-like guide robot system and demonstrate it in a real environment. Our research questions are as follows:

– How can we develop an autonomous human-like guide robot system, which is able to detect and track a human, to estimate human attention, to approach to an appropriate

✉ Takamasa Iio
iio@sce.iit.tsukuba.ac.jp

1 University of Tsukuba/JST PRESTO, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan

2 Intelligent Robotics and Communication Labs., ATR, 2-2-2 Hikaridai Keihanna-Sci. City, Kyoto, Japan

3 Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, Japan

4 Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, Japan

Fig. 1 Robot explains exhibit at a science museum

place and provide information according to the situation, while moving around an environment?
– How do people interact with such a robot system in a real environment?

The rest of our paper is organized as follows. We review related work in Sect. 2 and in Sect. 3 propose an alternative approach called speak-and-retreat interaction. In Sect. 4, we report how our system is configured to autonomously exhibit guiding behaviors. Finally, in Sect. 5, we report how visitors interact with our robot in our field study (Fig. 1) and whether they accept it with our approach.

2 Related Works

2.1 Guide Robots

Some early works chose a tour guide approach [1, 2, 9–11], where a robot navigates around a museum or other environments and provides explanations whenever appropriate. Some robots have been equipped with a GUI, enabling users to interactively provide input about tour destinations [2, 5, 11]. TritonBot [10] tries to use speech and face recognition for interaction with visitors.

Alternatively, other challenges exist with human-like guide robots in which gaze and gestures are often used. Such robots are typically aware of close individuals. For instance, a receptionist robot identified individuals by RFID tags and interacts by orienting its face to them [4]. A direction-giving robot in a shopping mall also identified people using RFID tags and exhibited continuity behavior to repeat visitors [7].

We also want our robot to be aware of particular individuals. Compared with previous literature, it is novel because it identifies an individual, estimates the exhibit at which she is looking, and proactively approaches her to provide an explanation.

2.2 Gaze, Gesture, and Spatial Formation for Information-Providing Interaction

Previous HRI studies revealed how human-like bodies can contribute natural, smooth, and effective communication. Since appropriate gazes make interactions more engaging [12], users will be more invested in listening [13]. The importance of pointing gestures was reported [14, 15]. Proxemics, initially studied in inter-human interaction [16], have been scrutinized in human–robot interaction (e.g., [17–19]). Spatial formation is modeled so that a robot can compute where it should be when it explains an exhibit [20]. We learned from such literature and used human-like gaze, gesture, and spatial formation in the guiding behavior for our robot.

2.3 Relation-Building Behavior

Researchers have studied behaviors that contribute to establish relationships with users in repeat interactions, so called relation building or relational behaviors [21], which include social dialogs and self-disclosures. Other work reported the importance of emotion [4], storytelling [4], name referring behavior [22], and personalization [23]. Regarding nonverbal behaviors, the proximity (social distance) between a user and a robot changes over time [19]. A model adjusts the proximity and spatial configuration to express different levels of friendliness [24].

Previous studies applied relation-building behaviors in daily contexts, such as an exercise coach [21], direction giving [7], and an office supporter [23]. Although Jørgensen and Tafdrup reported that visitor's impressions in a robot tour guide scenario [25], they do not clarify effects of relation-building behaviors. In this way, the previous studies did not apply a relation-building behavior to a guide robot.

3 Design Considerations

3.1 Observation and Interviews in a Museum

Our goal is to model human guidance that serves at specific exhibits at museums. For this purpose, we observed behaviors of guides in a museum and interviewed them about how to explain exhibits.

From the observation, we saw most of guides behaved as follows: When visitors show interest, human guides approach and start their explanations. They continue to provide information and an open dialog for questions as long as the visitors seem interested.

The interview results are summarized as follows:

– Guides explain to visitors who are standing in front of an exhibit and looking at it for a while.

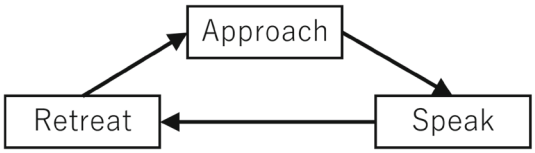


Fig. 2 Speak-and-retreat interaction

- An explanation combines several topics about an exhibit such as history, features, principles of operation, etc.
- The length of a typical explanation is in about 5–10 min.
- There are a few people asking questions during an explanation.
- Club members of the museum are often come. One-third of visitors might be the members.
- There are people who come once or twice a month.
- During an explanation, some repeaters said that they had heard the same explanation.
- Guides tend to change an explanation according to visitor's knowledge level.

From the results of the observation and the interviews, we designed speak-and-retreat interaction and relation-building behaviors.

3.2 Speak-and-Retreat Interaction

However, completely replicating such human interaction is currently impossible. The problem is the robot's capability for natural language processing. Since automatic speech recognition (ASR) remains immature for robust use (see our discussion in the introduction), it is difficult to estimate the degree of the interest of visitors and whether they are willing to listen to further explanation.

Instead, we invented an approach called speak-and-retreat interaction (Fig. 2). The robot approaches a visitor (approach phase) and says one segment of the utterances for an explanation (speak phase). After that, it immediately retreats (leaves) without yielding a conversational turn and waits for an opportunity to provide its next explanation (retreat phase). The visitors have complete freedom to look at the exhibit.

Such an approach has the following merits:

- If visitors are interested, the robot can provide a rich amount of information; the visitors can listen to as many explanations as they want.
- The visitors are not compelled to continue to listen when they are not interested. They can easily move to other exhibits when the robot is not attending to them (in the retreat phase).

We are also aware of the potential demerit:

- Visitors might be frustrated because they are not given an opportunity to talk or ask questions.
- Visitors who want to look at an exhibit in quiet and alone might be annoyed because the robot approaches to them and explains the exhibit, while they are looking at it.

We designed our robot with the above approach. In our field study we observed whether this approach encouraged interaction with visitors to such a level that they wanted to interact with it again in the future.

3.3 Relation-Building Behavior

Our robot also exhibited relation-building behaviors. According to museum guides, some people regularly visit museums. For such "regulars," guides interact quite differently than with first-time visitors. They explain more technical information and avoid previously broached topics or ideas. They sometimes engage in social dialogs. Such behaviors are consistent with the relation-building behaviors reported in the literature in Sect. 2.3. These behaviors are also evident in such other daily contexts as shops and offices and are critical for both inter-human and human–robot interaction.

Hence, we applied relation-building behavior to our human-like guide robot. It identifies individuals, engages in social dialogs, and coordinates the explanation contents in a way that avoids previously explained ideas/topics and gradually changes the target of the information from novices to people who already have some knowledge of the topic.

4 System

Figure 3 illustrates the architecture of our developed system. When a visitor enters the exhibition area, she is identified by an RFID, and her ID is associated with the tracked entity in the person-tracking module by the person identification module. The attention estimation module gauges whether a visitor is looking at an exhibit based on the position information provided by the person-tracking module. Information from the above sensory system is used in the behavior selector module in which a behavior is selected based on rule-matching. Since we extended previous architecture [7], we refer to the rules as episode rules. The system compares each episode rule with the sensory input and the history of previous interactions and executes the behavior specified in the matched episode rule. More details are explained below.

4.1 Modules

We used the following existing modules.

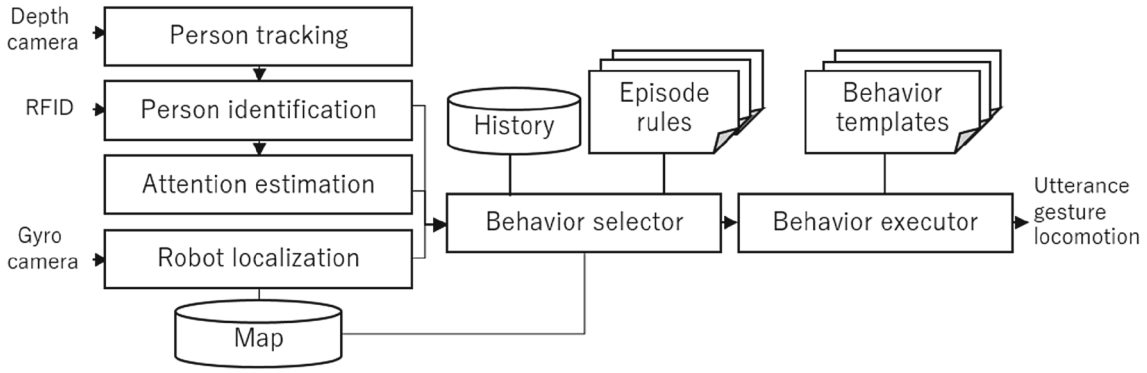


Fig. 3 System architecture

4.1.1 Robot

We used a humanoid robot, ASIMO [26], which has a bipedal locomotion mechanism with a 3-DOF head, 7-DOF arms in each arm (total of 14), and 2-DOF in each hand (total of 4). Its speech is based on speech synthesis. When necessary, the output from the speech synthesis is manually adjusted in advance to increase its naturalness.

4.1.2 Robot Localization

Localization is achieved with a landmark based method [26]. We placed 49 markers on the floor. Robot's current position is estimated from a gyro sensor and adjusted when a landmark is observed. In order to robustly detect landmarks in a real environment where the light conditions change in various ways, retroreflective markers were used as landmarks. The waist of the robot was equipped with an LED for infrared light irradiation to illuminate the makers. The robot detected the markers using the image of the floor surface taken through the infrared light transmission filter. For safety concerns, if no marker is visible for a 11.0 m walking duration, the localization is considered inaccurate, and the robot stops its navigation.

4.1.3 Environment and Person-Tracking

The study was conducted at Miraikan (National Museum of Emerging Science and Innovation) in Tokyo, Japan. We prepared a 17.5 × 8.2 m area with three exhibits (a scooter, a race car, and an engine) (Fig. 4). Range camera sensors, 37 ASUS Xtion PRO Live, were attached to the ceiling at 3.5 m. With a previous technique [27], people's head locations (x, y, z) and body orientations were provided every 33 ms. According to [27], the tracking precision as the average error of estimated person position was 74.48 mm and the accuracy was 99.94%, when two persons were in a space of approximately 8 m^2 (3.5 × 2.3 m).

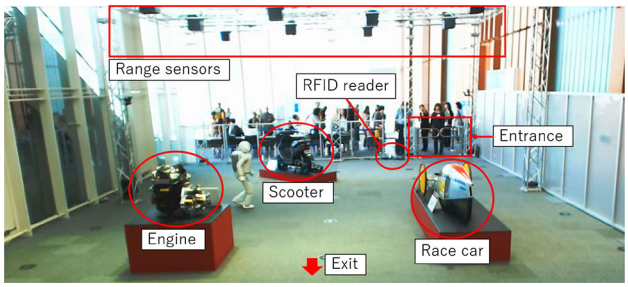


Fig. 4 Environment

4.1.4 Person Identification

We used an active-type RFID (Matrix Inc., MXRD-ST-2-100). Visitors wore RFID tags. Tag readers were placed at the entrance and the exit. When visitors passed through the entrance, the RFID was read, and their unique IDs were recognized and associated with the output from the person being tracked.

4.2 Attention Estimation

We developed a system to estimate a visitor's attention (which exhibit a person is looking at) from her location and body orientation. When a person looks at an exhibit, she tends to be near it and/or has oriented her body toward it. Thus, our attention estimator uses as features the distance between a person and an exhibit and the angle between her body orientation and the direction toward it (Fig. 5a). Note that these useful features depend on the situation. Sometimes people stop and look (Fig. 5a), and then their body orientation is the dominant element for identifying attention. In contrast, sometimes people walk around an exhibit while observing it (Fig. 5b), and at such times their body orientation is less important because it changes and is not oriented toward the exhibit. Instead, distance plays an even more important feature. We modeled such behavior in a multiple-state model and estimated the attention target from the time sequences of

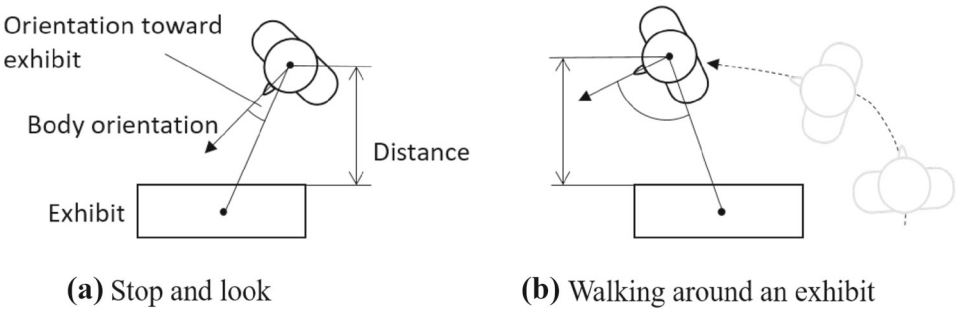


Fig. 5 Attention estimation

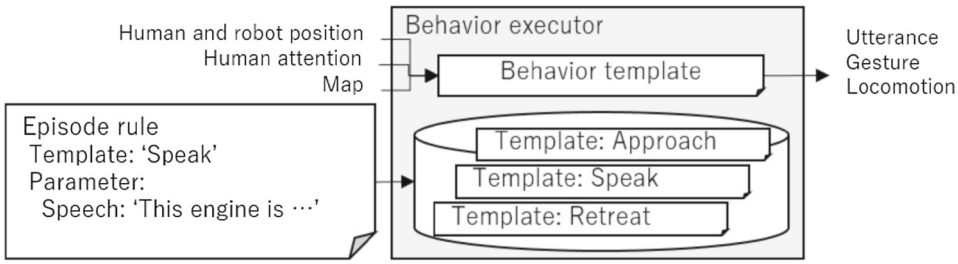


Fig. 6 Overview of behavior executor

the location and body orientation. Due to page limitations, we omit further details and will report them elsewhere.

4.3 Behavior Selector

In our architecture, just one behavior is always being executed. A behavior is a program that controls the utterances, gestures, and locomotion based on the current sensory input. The behavior selector module is a rule-based system in which the current state is periodically matched with pre-implemented rules called episode rules. Each episode rule consists of a condition, a priority, and a behavior to be executed. The conditions in the episode rules can be a combination of the sensor information and the history of previous interactions. When multiple rules are matched, a rule is chosen with higher priority. Finally, the behavior specified in the selected rule is executed when a rule is matched. If the current situation does not match with the implemented rules, nothing happens. The behavior selector does not select any behavior. When no behavior is selected, the behavior executor does nothing. Therefore, the robot does nothing until a situation that meet rules comes.

4.4 Behavior Executor

4.4.1 Overview

Figure 6 illustrates the mechanism of the behavior executor. Each behavior must specify how the robot should behave (e.g., speech, gesture, and locomotion). However, for a large-scale application, excessive elaboration is required if we make a program for each behavior. Instead, based on the encapsulation concept [28], we prepared templates for the behaviors.

Each template is designed as an abstracted behavior that is independent from concrete contents. Many specifications (e.g., utterances, gestures, and the targeted exhibit) are configured as arguments for the template. Thus, behaviors can be added or edited without touching the program by just providing arguments.

Our approach enables efficient division-of-labor. Programming experts concentrate on implementing the templates. Domain experts who know the exhibits well and recognize affective explanatory methods can work in parallel for designing behavior contents, including what a robot should say and how it should change its behavior based on the previous interaction history.

The following three templates correspond to the three states in Fig. 2 approach, speak, and retreat. Approximately every 2 min the robot approaches a visitor with behaviors taken from the approach template and provides an explanation with the speak template. Then the robot immediately backs away. Below we explain these three templates.

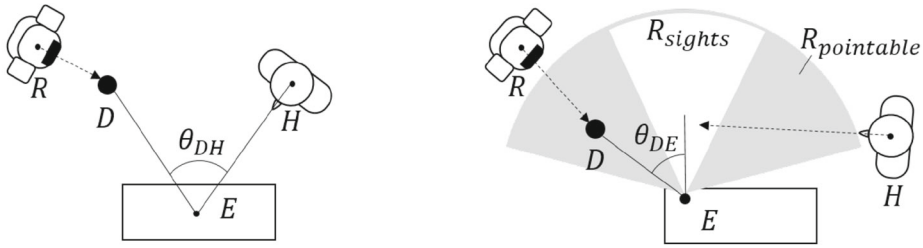
4.4.2 Approach Template

The approach template is used when a robot walks near the target person (Fig. 7). Such parameters as target person, target exhibit, target part of the target exhibit, utterance used when the robot approaches, social distance, and its speed depend on the context (that is, on the behaviors).

On the other hand, shared by all the approach type behaviors, an algorithm can find the best position from which the robot can explain the specified exhibit in the next speak

Fig. 7 Approach behavior

Fig. 8 Computing destination for approach behavior



(a) Explaining the exhibit as a whole **(b)** Explaining part of the exhibit

Fig. 9 Speaking behavior

behavior. For computation, there are two cases: whether the robot explains the exhibit as a whole or just part of it.

When the robot explains the exhibit as a whole, it considers the spatial formation and the social distance (Fig. 8a). We computed it based on the model extended from our previous model [20] for the F-formation concept. The robot goes to a location where the person can see both the robot and the exhibit and maintains distance DE to the specified social distance, angle θ_{DH} to 90°, and a small difference in distances HE and DE.

When the robot points at a specific part of the exhibit (e.g., engine's combustion chamber, Fig. 10b), it needs to be at a location where its specified part is visible and can be indicated. Moreover, the robot needs to make space for the targeted person to approach and look. Six to nine parts are predefined for each exhibit. We defined two regions, R_{sights} and $R_{pointable}$ (Fig. 8b). R_{sights} is where it is easier for the person to see what the robot is indicating. $R_{pointable}$ is where the robot can point effectively. R_{sights} is a fan with 30–120° and $R_{pointable}$ is a fan with 100–180°, configured depending on the part's visibility. The robot chooses the location inside $R_{pointable}$ but not within R_{sights}.

After the target location is computed based on the above idea, the robot starts navigating to the target location by avoiding static and dynamic obstacles. To prevent unsafe impressions, we controlled the robot's speed so that it starts slowing down within 1.5 m of the target location.

In situations where one person approaching another, it is natural to start talking before walking has ended. When the distance is short and gazes are shared, people already feel involved in a conversation [29]. Thus, when the robot approaches (3 s before the estimated arrival), it says something like "Hi Taro, are you interested in this engine?"

4.4.3 Speak Template

The speak template is used when a robot explains an exhibit (Fig. 9). Such parameters as utterances for explanations, gestures, and explanation targets are specified. A gesture is specified by a predefined motion file.

We used two types of pointing gestures based on pointing in human-communication. People point with their index fingers to indicate an exact location and with an open hand when introducing a topic about a target at which both are already looking [30]. In our case, we used open-hand pointing (Fig. 10a) when the robot is explaining an entire exhibit and index-finger pointing (Fig. 10b) when the robot is just explaining a specific part of it. If the target visitor is at a location from which the exhibit's target part is not visible, the robot beckons the visitor closer before starting its explanation and points at the target part. During the utterance the robot's head is oriented toward the visitor's face. For joint-attention [31], its head moves toward the exhibit for 3 s when it points at a part of it.

Fig. 10 Pointing in speak behavior

(a) Pointing to a whole exhibit **(b)** Pointing to part of an exhibit

Fig. 11 Retreat behavior

4.4.4 Retreat Template

In the retreating behavior, the robot backs away from the person who got an explanation and waits until it begins the next explanation (Fig. 11). A previous report [24] argued that visibility is critical for deciding the waiting location. That is, when an attendant wishes to show warmth, she remains visible to the visitor; in contrast, for low friendliness, she remains inconspicuous. Based on this idea, the system computes the destination based on two criteria: visibility and distance.

The visibility, which is computed based on whether the target location is within the sight of the visitors, is estimated from the angle between the visitors' attention target and the location. If the angle is less than 60°, it is visible. As reported in the next subsection, we controlled the robot's friendliness: a visible location for warmth and a hidden location for low friendliness. We did not consider the visibility for average friendliness. If the robot is too close, its presence might distract the visitor; if it is too far away, approaching to listen to the next explanation might consume too much time. The robot chooses one of the locations that satisfies the above criteria.

4.5 Implemented Rules and Behaviors

4.5.1 Episode Rules

We implemented 279 episode rules for our system and classified them into 13 groups by functions (Table 1). Since the rules covered enough of the situations that might occur in the actual exhibition area, there was no situation where the robot stopped with no behavior selected. All the episode rules contain three elements: behavior, condition, and priority. Figure 12 shows a part of an episode rule of an approach to explain an exhibit where the notation is simplified for readability. In the elements, the operation of the variables related to the system behavior (system variable operation) and the conditional expressions using system variables are described. We defined 50 system variables in the episode rules. Some are shown in Table 2.

4.5.2 Example of Instantiation of Templates

The behavior element includes contents that are executed as a behavior: the behavior's name, its template, the system variable operation at its beginning and completion, and such contents as utterances, gestures, and gaze directions. For example, in Fig. 12, the behavior name is ApproachCuv001, and its template is Approach. SetUp and TearDown attributes denote the system variable operation at the behavior's beginning and completion. Figure 12 shows that the system variable template is assigned approach at its beginning. Furthermore, at its completion, the system variable last_behavior_name is assigned its own behavior name (e.g., ApproachCuv001), and the system variable template is assigned explain if the behavior is completed successfully. The content attribute describes the concrete actions to be executed by the robot. The content attributes of Fig. 12 mean that the robot greets the visitor by name ("Hi, Ichiro"), raises its right hand and 1000 ms later, and approaches, saying "That engine looks complicated, doesn't it?"

Table 1 Episode rules implemented for system

Group	Description	Num
Approach to explain	Behavior rules where robot approaches a visitor to explain exhibits	102
Approach to greet	Behavior rules where robot approaches and greets a visitor	2
Explain	Behavior rules where robot explains exhibits	102
Retreat	Behavior rules where robot leaves a visitor after its explanation is finished	2
Say hello	Behavior rules where robot greets a visitor who has entered exhibition area	1
Farewell	Behavior rules where robot says goodbye to a visitor who has left exhibition area	3
Relationship	Behavior rules where robot expresses relational behaviors to a visitor	40
Roaming	Behavior rules where robot roams around exhibition area	4
Greet	Behavior rules where robot greets a visitor who entered exhibition area	10
Stay	Behavior rules where robot stays near a visitor after an explanation	6
Summary	Behavior rules where robot summarizes an explanation that a visitor has heard before	3
Pseudo	Rules for pseudo-behaviors to operate system variables	3
Move to markers	Behavior rules where robot moves to markers when it missed them	1

```
Behavior: {
    Name:     “ApproachCvcc_main”,
    Template: “Approach”,
    SetUp:    [ “template = ‘Approach’”, … ],
    TearDown: [ “last_behavior_name = self_name’
                  if return_code == ‘SUCCESS’)”, … ],
    Contents: [
        Speech:   [ “Hi, {{ username }}.”,
                    “That engine looks complicated, huh?” ],
        Gesture:  [ “raise_right_hand”, “None” ],
        Gaze:     [ “user”, “user” ],
        Interval: [ 1000, 1000 ],
        … ]
},
Condition:
    “template == ‘Retreat’ and attention == ‘engine’ and
     (intention_to_talk == 1.0 or explanation_count_engine == 0)”,

Priority: 100
```

Fig. 12 Example of an episode rule for approach behavior

The condition element contains conditional expressions for the behavior execution. For example, the following are the conditions shown in Fig. 12. The system variable mode is roaming, and the system variable attention is engine, and either the system variable intention to talk equals 1.0 or the system variable explanation_count_engine equals 0. These conditional expressions denote that the robot is roaming and the visitor is paying attention to the engine, and the robot wants to talk with him or the visitor did not hear the explanation of the engine.

The priority is a numerical value for determining the priority order of the episode rules. When the conditions of several episode rules are matched with the system variables, the episode rule with the highest priority is chosen.

4.5.3 Behaviors for Explanation

Our field study (Sect. 5) used three historical exhibits about ecological examples of transportation. Scooter is an electrical motor moped that appeared in early commercial markets worldwide. Race car is a high mileage vehicle that won two international competitions. Engine refers to a motor called CVCC with low pollution that first cleared US government regulations in the 1970s.

Table 2 Example of system variables

System variable	Description
Robot status	
Template	Behavior’s template
last_behavior_name	Behavior name being executed by robot
return_code	Return code when robot completed behavior
intention_to_talk	Degree to which robot wants to talk to a visitor. Value increases depending on time visitor spends at an exhibit and decreases after robots explanations
Visitor status	
Attention	Point to which visitor is paying attention
explanation_count_X	Number of times visitor received explanation of exhibit X from robot
Stage	Variable representing relationship between the visitor and the robot. Value is mainly related to executing relational behaviors

We created 102 behaviors for the explanations and implemented them as instances of the speak template. Each explanation is designed to last roughly 1 min, and typically at the end the robot encouraged the participants to find a feature of the exhibit that they might want to learn more about. For instance, the robot might explain the race car as follows: “With this machine, the participants competed in mileage races with just 1 l of fuel. This machine broke the world record several times. Notice the exhibit has many objects that improved its mileage.”

```
Behavior: {
  Name:     “Retreat_normal”,
  Template: “Retreat”,
  …
},
Condition:
 “(last_behavior_name == ‘ExplainCvcc_main’ or
   last_behavior_name == ‘ExplainCvcc_cylinder’ or
   last_behavior_name == ‘ExplainCvcc_chamber’ or
   ...) and
  intention_to_talk < 1.0”
```

```
Behavior: {
  Name:     “Retreat_normal”,
  Template: “Retreat”,
  …
},
Condition:
 “template == ‘Speak’ and
  intention_to_talk < 1.0”
```

Fig. 13 Use of template in episode rules: Left: rule without template name; hence names of behavior instances are listed. Right: rule with template name

4.5.4 Implementation of Speak-and-Retreat Interaction

Our behavior templates (Sect. 4.4.2–4) simplify implementing the speak-and-retreat interaction pattern. We have over 100 instances for explanation behaviors (Table 1). For the speak-and-retreat pattern, we executed retreat behaviors after completing each of the explanation behaviors. If we need to specify transition rules from/to these retreat behaviors, a large number of names can be cited in the conditions: last_behavior_name == ExplainEngine_main, and last_behavior_name == ExplainEngine_cylinder … (Fig. 13, left).

Instead, due to the templates, we can refer to the behaviors using their templates, like template == speak. Typical speak-and-retreat patterns can be implemented with such rules as shown in Fig. 13, right. Thus, by referring to the names of the behavior templates, we can easily implement the episode rules.

4.5.5 Relation-Building Behaviors

We deployed a stage-based model [21], where the system manages the growing relationship between the robot and visitors as a progress of a stage. For this study, we created three stages: new, acquaintance, and friend. The stage proceeds one by one when a visitor leaves the environment or if a visitor listened to one or more explanations. Hence, acquaintance and friend mean twice and three times return visitors, respectively. The robot's behavior varies along with the stage as follows:

Verbal expressions: Based on a previous work [21] we implemented 40 relation-building behaviors for the acquaintance and friend stages using the speak template. The behaviors show a praising, self-disclosure, or empathetic utterance before an explanation behavior. Below are some examples:

– “Hanako, you seem to have learned a lot from the exhibits” (praise).
– “Now I only guide at Miraikan, but I eventually hope to work at many other places and guide lots of people” (self-disclosure).
– “This scooter is so cool. Hanako and I have similar taste” (empathy).

We also made eight greeting behaviors for acquaintances and a frequency stage using the speak template. In the behaviors, the robot talks about the person's last visit when she enters the area. For example, it might say, “Hanako, welcome back. Nice to see you again.” Such utterances represent the robot's continuity.

Non-verbal expressions: We applied a friendliness model of non-verbal behavior [24] to the approach behaviors. For new visitors, the robot slowly approaches at a relatively remote social distance (2.2 m). For friend visitors, it shows warmth, approaches more quickly, and chooses a shorter social distance (1.0 m). Based on the stages, we implemented these differences in approach behavior in the approach template because this property is shared by all the approaches. In addition to the approach to explain behavior, we made two behaviors for the approach to greet. The robot exhibits them for visitors in the acquaintance or frequent stages when they enter the exhibition area.

4.6 Execution Example

Figure 14 shows how our system works and Fig. 15 shows simplified examples of the relevant episode rules. The visitor stood at the engine exhibit while the robot was roaming around after retreating (Fig. 14, top). Then he approached the race car (Fig. 14, middle); here, since the attention estimation module estimated that the visitor was focusing on the race car, it updated the attention variable to the race car. Between the two episode rules, the right episode rule fired, because its condition matched the system variables, e.g., attention is now on the race car and the robot's behavior template is in retreat. Since this episode rule was chosen, the behavior executor updated the system variables and executed the behavior (bottom, Fig. 14). Thus, the robot approached this visitor. Likewise, the system conducted attention estimation, updated the variables, matched the episode rules, and executed behaviors.

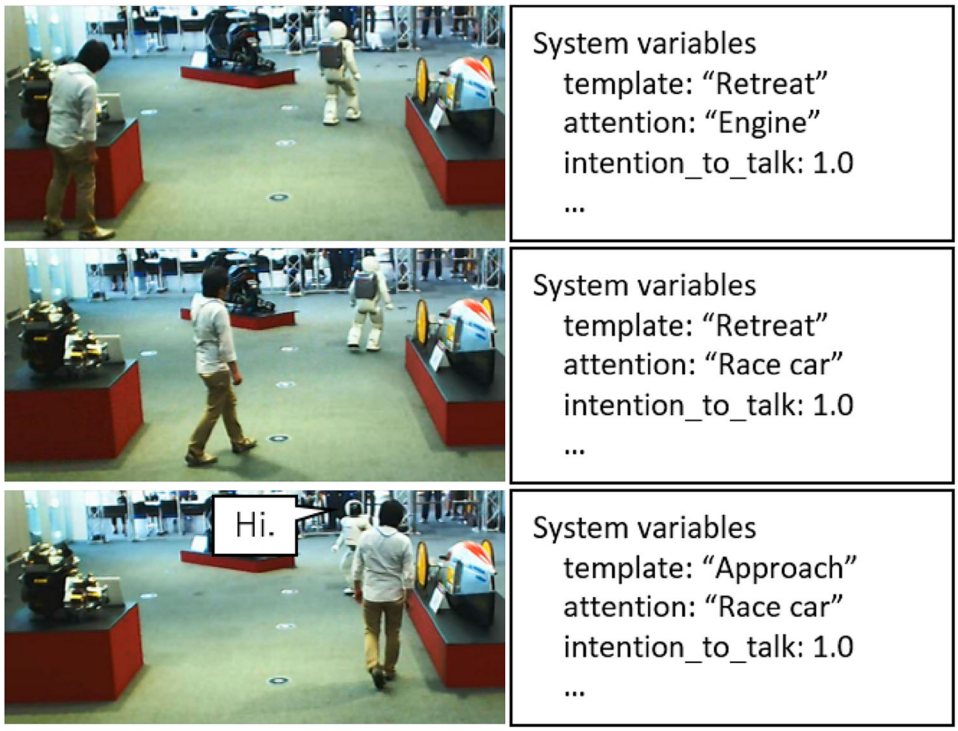


Fig. 14 Scenes where robot approaches a visitor and system variables in scenes

```
Behavior: {
  Name:     "ApproachCvcc_main",
  Template: "Approach",
  SetUp:    [ "template = 'Approach'", … ],
  TearDown: [ "last_behavior_name = …, … ],
  Contents: [ Speech: [ "Hi.",…],…], …
},
Condition:
  "template == 'Retreat' and
   attention == 'engine' and
   intention_to_talk == 1.0",
```

```
Behavior: {
  Name:     "ApproachEcorun_main",
  Template: "Approach",
  SetUp:    [ "template = 'Approach'", … ],
  TearDown: [ "last_behavior_name = …, … ],
Contents: [ Speech: [ "This…",…], …
},
Condition:
  "template == 'Retreat' and
   attention == 'race_car' and
   intention_to_talk == 1.0",
```

Fig. 15 Examples of episode rules

5 Field Trial in a Museum

5.1 Procedure

We used our developed robot at Miraikan for 18 days from 10:00 to 17:00 with a one-hour break. Museum visitors could freely walk around the exhibition space. Before starting trials, we gave museum visitors who visited our booth instructions for safety of the robot and privacy concerns. Written informed consent, which was approved by our institution's ethics committee for studies involving human participants, was obtained from all participants. Some of the participants also consented to being visible in non-anonymized pictures. 231 visitors signed up: 131 males, 100 females, ranging from 7 to 76 years old, average age 29.83, s.d. 17.98.

Due to safety concerns, the exhibition space was reserved for registered participants who received such safety instructions as the appropriate distance that must be maintained from the robot. Only one visitor was allowed to enter at a time. In the exhibition area, they could browse freely around the exhibits and exit anytime. To provide more opportunities, each visit was limited to ten minutes. Every time a visitor left, we conducted a brief interview with him/her. The visitors were not paid.

5.2 System Performance

Next we describe the following two aspects of our system's performance: autonomy and exhibit selection.

Autonomy The system basically operated autonomously. Operators helped when the system's localization failed (about once per hour), the person-tracking failed, or when the interacting person's ID was lost (about ten times per hour).

Exhibit selection The robot proactively approached the participants to explain the exhibits they were looking at. We evaluated how well the system estimated the target exhibit. Two independent coders who did not know the study's

Fig. 16 Child looks at embedded wheel

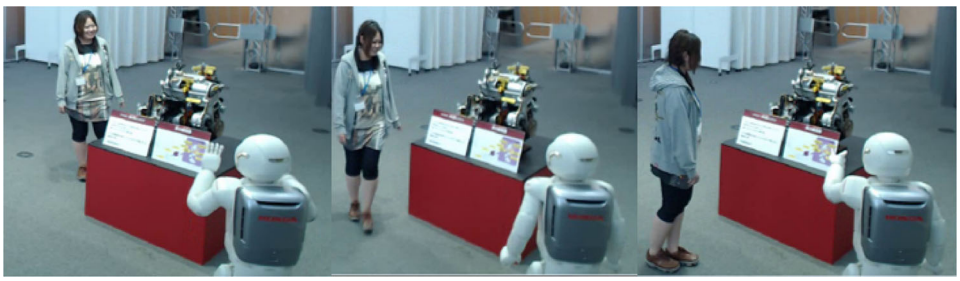


Fig. 17 Woman approached when invited by robot

purpose observed video and evaluated which exhibit the participants were looking at for each moment the robot provided explanations, and we checked the matching ratio. The result is satisfactory, showing 95.3% matching between the robot selections and the coder judgments. The major reason for failure was due to the complexity of human behavior. When a participant looked at a remote exhibit near another exhibit, the system often failed to correctly estimate her body orientation and misrecognized which exhibit she was looking at.

Although operators helped the system at each time the system failed, the intervention by was in short time. On average, when localization fails, it took about 20 s to move the robot to a nearby marker, and the ID reassignment of a lost person was about 3 s. Thus, the time that the operator helped the system was less than a minute in an hour. From the fact that 98% of all uptime was working autonomously, we believe that the system worked as designed in a reasonably autonomous way.

5.3 Visitor Behaviors

We analyzed how visitors behaved during their first-time visits (repeat visitors provided similar but brief comments on this). First, we analyzed the duration of their stays. Staying was restricted to ten minutes. A majority of the participants (139 of 231 visitors) stayed for the entire ten minutes: an average of 8 min and 58 s (s.d. 95 s) and listened to 5.23 explanations (s.d. 1.51). Explanations were provided approximately every 2 min (speak state in Fig. 2) and roughly lasted a minute. During the remaining time, the robot was in the retreat or approach state, while the participants either looked at the exhibit being explained or moved to other exhibits. The environment where we had a field trial had only three static exhibits, in other words it was a little bit boring environment. If the robot provided an uninteresting explanation in such an environment, the participants would go out the exhibit area after hearing an explanation at each exhibit (i.e. three times). Nevertheless, in fact participants heard the explanation five or more times. Therefore, we believe that the interaction with the robot was sufficiently engaging.

Participants attentively listened to the explanations provided by the robot. They tended to stay near the exhibit when the robot explained it. When it talked and pointed at a part of the exhibit, participants often peered at it (Fig. 16). When a participant was too far from the location to see which part was being explained, the robot stood where it could see the part and invited the participant to come closer. Participants usually followed such a request (Fig. 17).

We coded the overall behavior of the participants and evaluated whether they listened to the explanations. They were classified as did not listen if a participant left the exhibit (moved to another or left the environment). They were classified as atypical if a participant did not listen to one or more explanations, tested/teased the robot, or did any atypical behavior that a typical exhibit visitor would avoid. One coder unfamiliar with our study purpose coded all of the data, and 10% were coded by a second coder. Their coding results matched well, yielding a Kappa coefficient of 0.895. The coding result shows that 179 visitors (77%) listened to all of the explanations provided by the robot without any atypical behaviors. Figure 18 shows an example of such a typical visitor.

Among the remaining visitors, 13% stopped listening to the explanation once or more (nevertheless, they typically listened to most of them). 8% exhibited testing or teasing behavior toward the robot, such as impeding its navigation (Fig. 19). 7% behaved atypically to the exhibits, like standing in the middle of two exhibits, which caused the system to misrecognize the visitors' attention.

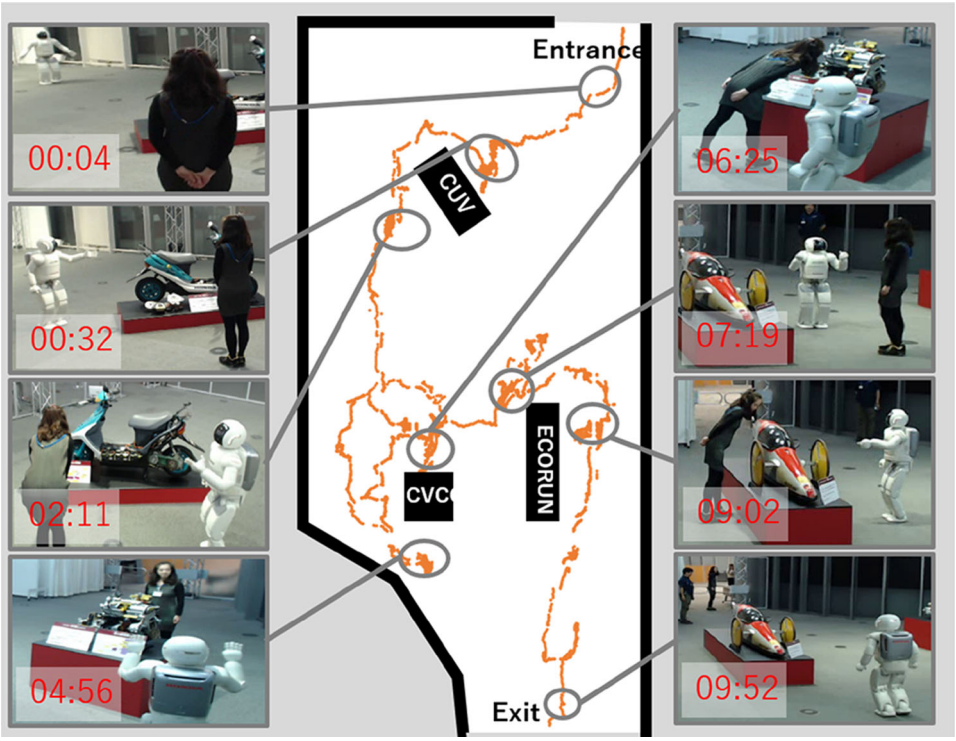


Fig. 18 Typical overall behavior

Fig. 19 Testing behavior

5.4 Repeat Visitors

Thirty-two participants visited twice, eight visited three times, three visited four times, and one visited five times. Some even visited on other days from their previous visits (23 out of 44 visits). We did not find any differences in their staying patterns from their first visits. They listened attentively to the robot's explanations and typically stayed for the entire ten minutes. For instance, visitors on their 2nd visits stayed an average of 9 min and 26 s and listened to 5.52 explanations, and visitors on their 3rd time stayed an average of 9 min and 29 s and listened to 6.75 explanations. Note that they still listened to the robot for such a long time on their repeat visits, even though the exhibits had not changed at all.

5.5 Interview Results

5.5.1 Research Methods

We conducted semi structured interviews to all participants. The interviews were carried out by third parties who did not know our research purpose. They asked the participants following an interview guide we prepared for. The interview guide is described as follows:

1. Which is better, comparing robot's explanation and human's explanation? Please answer "robot", "human" or "undecided".
2. Why do you think so? (If the participant does not mention pros and cons) Please tell me the good and bad points respectively.
3. Do you want to use these robots if they were in another exhibition space? Please answer "yes", "no" or "undecided".
4. Why do you think so?
5. (Showing pictures) Please choose a picture that best describes a relationship you have felt with the robot.
6. (To only repeaters) How did you feel the robot you visited this time compared to the robot you visited last time?

The first and second items are questions about a comparison of a robot guide with a human guide. The third and fourth items are based on the intention to use concept [32]. We analyzed these interview results from each visitor's first visit. The fifth item is an Inclusion of Others in the Self (IOS)

Table 3 Why participants preferred human/robot

Preference	Reasons	Rate (%)
Human	Interactive	48
	Flexible	19
	Easy-to-listen	13
	Emotional	6
Robot	Enjoyable/attractive	16
	Accuracy (free from errors and fatigue)	15
	Obligation-free	14
	Novelty	14
	Easy-to-listen	6
Both	Equivalence	7

pictorial scale [33]. The fifth and sixth items are used for analyzing any differences of a relationship with the robot that the repeat visitors perceived from their previous visit. The interviews were recorded by IC recorders.

The analysis of the participants' free comments was conducted in the following procedure: First, in each question, we divide each comment into sentences and make some categories from the meaning of each sentence. Then, two-third-party coders classified the comments to the category. By looking at the distribution of the classification, we considered what factors are important in the participants' comments.

5.5.2 Comparison with a Human Guide

One hundred and six people (46.3%) preferred a human guide, 57 (24.9%) preferred the robot guide, and 66 (28.8%) reported undecided. We prepared categories based on their reasons, and two independent coders categorized them. Table 3 shows the categorized coding results in which the percentages are from the entire population. Note that an answer can be classified into multiple categories. Their classification matched reasonably well, yielding an average Kappa coefficient of 0.808.

The table shows the four primary reasons that support their preferences for a human guide:

- "A person can answer questions about things that I did not understand by listening to the explanations. I prefer a person" (categorized as interactive).
- "People are more flexible with facial expressions and situations" (flexible).
- "ASIMO is cute, but its speech was sometimes hard to understand" (easy-to-listen).
- "A person can communicate emotion and more detailed messages like 'this part is difficult' through by eye contact and facial expressions (but the robot cannot do that). That is why I think a person is better" (emotional).

Those who preferred the robot provided the following comments:

- "As long a robot's batteries are charged, it can continue to operate. It will not get tired, so its explanations will not become rough. When a person gets tired, he may explain haphazardly. But robots cannot get tired" (accuracy).
- "Human guides expect responses from visitors. I feel rather obligated to follow such expectations. In contrast, a one-way robot is easier" (obligation-free).
- "It is easier to listen to a robot. It speaks at the same pace, unlike a person who sometimes speaks too fast, sometimes too slow, which I do not like" (easy-to-listen).

Some participants commented on the similarities of humans and robots:

- "A person can explain how to do something. A robot can do the same, and it can continue its explanation until I understand" (equivalence).

5.5.3 Intention to Use

An overwhelming majority of participants 214 (94.74%) answered 'yes' to the question about intention to use (i.e., wanting to interact with it again). Six (2.63%) answered 'no' and six (2.63%) were undecided. We further analyzed the reasons for those who answered 'yes' by coding their answers by two independent coders (first coder coded all the data, and the second only coded 10% for confirmatory coding). The classifications matched reasonably well, yielding an average Kappa coefficient of 0.691.

Table 4 shows the coding result with the percentages from the entire population. Like above, answers can be classified into multiple categories. Thus, the sum probably exceeds 100%, and the sub-categories exceed the totals for each category.

Thirty-two percent of the participants attributed their praise to the robot's explanatory capability:

- "Because I could understand the exhibits today" (useful for understanding).
- "Compared with people, their answers are more accurate. A robot does not get tired, either. It can always provide a similar amount of explanation in a similar way" (accuracy).
- "I may or may not listen. It depends on how well the robot is explaining. If a person tries to explain something to me, it is more difficult to refuse. Sometimes, I would like to just half-listen" (obligation-free).

Twenty-nine percent of the participants commented on their interactions with the robot:

Table 4 Why participants wanted to interact again

Reasons	Rate (%)
Explanatory capability	32
Useful for understanding	22
Obligation-free (compared with humans)	8
Enjoyable/attractive	4
Accurate (free of errors and fatigue)	3
Body language, such as gestures	2
Interaction	29
Enjoyable	18
Relationship with robot	11
Relation-building behavior	4
Novelty	50
Interest to state-of-art robotics	26
Novelty of robots presence	23
Novelty of interacting with robots	5
No reason provided	10

Table 5 Perceived differences from previous interaction

Perceived differences	Rate (%)
Familiarity to robot	80
Robot is friendlier	46
Robot remembered me	46
Social distance seemed shorter	39
Good explanations	80
No duplicated explanations	49
Explanations based on previous ones	39
Easier to understand	17

– "I would like to shorten the psychological distance to the robot. Well, shorten might be a new concept to the robot. I would like to be able to touch it, of course, I do not really mean physical touching, but just having a relationship" (relationship with robot).
– "When it said my name and approached me, I thought 'Wow, it is so cute!' I was flattered and surprised, yet happy that it approached me" (relation-building behavior).

About half of the participants praised its novelty:

– "Because robots are relatively rare, I would probably visit many times" (novelty of presence of robots).

Participants who were not positive provided such comments as "interacting with it was weird," "I was afraid of breaking it," "I do not think it provides any merits," and "I prefer a human guide."

5.5.4 Relationship with this Robot from Repeat Visitors

We compared the ratings of the IOS scale among each visit of the repeat visitors with those who came two or more times (44 visitors). A pair-wise t test revealed that their ratings on the second visit ($M = 5.47$, $SD = 1.29$) were significantly higher than on their first visit ($M = 4.88$, $SD = 1.32$, $t(42) = -6.180$, $p = 0.001$, $r = 0.69$). Thus, after the second visit, the visitors felt closer to the robot.

We analyzed the interview results for the differences with the previous visit and analyzed the interviews from their second and subsequent visits. Their answers were coded by two independent coders whose classifications matched reasonably well, yielding an average Kappa coefficient of 0.634. Table 5 shows the coding result where the percentages are from the entire population. Answers can be classified into multiple categories.

Eighty percent of the participants commented on familiarity to the robot:

– "ASIMO is so cute! When I heard 'welcome again' and 'I'm happy' from it, I realized that it was behaving more friendly, and I felt more comfortable with it" (robot is more friendly).
– "I was happy when the robot remembered me. I felt connected to it. I noticed that ASIMO remembered its explanations and seemed to behave according to each individual" (robot remembers me).
– "My sense of closeness increased, which felt like friendship" (social distance is closer).

Eighty percent of the participants commented on the explanations provided by the robot:

– "On my second visit, the robot's explanations built on my 1st visit. It added more details and contents. It seemed intelligent" (explanations based on previous ones).
– "It pointed at specific locations of the parts. And it was easier to understand and gave more detailed explanations. I mean, it piqued my interest. It matched what I wanted to know on this 2nd visit" (easier to understand).
– "I did not know that the idea of air propulsion is invented over 70 years ago. Since ASIMO explained that idea so differently, it was much easier to understand. Smoother than before and better. The explanation was easier to understand than on previous visits" (easier to understand).

Overall, the majority of repeat visitors noticed and praised the robot's capability to handle repeated interactions. They typically focused on one or two features they recognized without necessarily talking about every difference they perceived. Thus, the interview results captured the features that

were most notable for the participants, but the lack of a mention does not necessarily indicate a lack of contributions.

Visitors who came three times or more responded similarly, and some left comments that reflected a growing relationship with the robot:

– "I wanted to see ASIMO because it acted like a friend."
– "ASIMO spoke to me in a very friendly manner. I felt that my impressions were being communicated to it."

6 Discussion and Conclusion

6.1 Interpretation of Results

One of the innovation points is that we have built a robot system that can walk around the environment, recognize to which exhibition a visitor is paying attention, and give explanations according to the situation at an appropriate position. Most of previous research on museum robot guides reported tour guide robots taking a route with visitors [1, 2, 9–11]. To the best of our knowledge, there is no report of autonomous robot system that can proactively approach and provide explanations to visitors who are looking around exhibits. Therefore, we believe that the robot system is quite unique in the world. This paper contributes to show the modules required to build the robot system and how to integrate the modules.

Furthermore, it is also an innovation point that we demonstrated the robot system in the real museum and the robot system was generally accepted by common visitors. Perhaps many HRI researchers maybe think that it is difficult to get robots to work well in the real world due to limitations of sensors and recognition technologies for now. They also may think that, even if such limitations were solved, people would get bored to interaction with a robot soon. In contrast to such assumptions, in our robot system, participants continued to interact with the robot for about 9 min. Most visitors wanted to interact with the robot again. We are encouraged that half mentioned either its explanatory or interactive capabilities. Many commented that they got useful, surprising, and/or interesting information from the robot. Although a robot's novelty will eventually wear off, the other reasons will remain. Thus, even with the current technology, deployment space exists for autonomous guide robots.

When hypothetically compared with a human guide, we were surprised that 24.5% of participants chose the robot guide even with its current capability. They commented that the robot guide is obligation-free and does not get tired, emphasizing two potential strengths of robot guides. In the future when robots are more capable of engaging in flexible natural-language dialogs, perhaps a majority of people will prefer robot guides. We believe that these results suggest a promising future for them.

Our speak-and-retreat interaction design was successful. Although its interactivity is inadequate if compared with a human guide, visitors nevertheless deemed the robot's explanation to be useful and wanted to interact with it again. Thus our speak-and-retreat interaction contributed a new way of designing human-like interaction.

Even though the exhibits did not change, we identified many repeat visitors who remained in the space until their allotted ten minutes were exhausted and reported that their relationship with the robot seemed closer. Unfortunately, since we could not separate the effect from our design, perhaps the effect would happen because they simply met the robot again. Our participants reported many aspects of the robot's relation-building capability as perceived differences, and this capability motivated them to interact with it.

6.2 Generalizability and Limitations

Based on the nature of field studies, this study has many specific characteristics, such as culture, robot appearance, and the exhibits. ASIMO might encourage visitors to participate and raise their expectations before the interactions. On the other hand, we believe that the result after the interactions mainly reflects their interaction with the robot, e.g., since a large majority wanted to interact with it again, with positive feedback about some characteristics. Furthermore, the speak-and-retreat interaction and the robot system we developed are adaptable to other robots that can move and show pointing and gaze, such as HRP series [34] that are bipedal same as ASIMO or Pepper [35] and Robovie [36] that move on wheels. Even if we used another robot in the field trial, we would obtain similar results to the current results. This is because many participants mentioned not the appearance of the robot but the goodness of explanations and relational behaviors of the robot. These opinions are not derived from the characteristics of ASIMO, but the functions of the robot. Therefore, we believe that the results are reproducible even if other robots are used.

Our trials attracted attentions of not only participants but also other visitors. Therefore, onlookers may have some effects on participants' behaviors towards the robot. For example, cognition of being seen by others might have biased participants to behave socially, in other words politely. In the future, if a robot becomes a natural existence and people become unconcerned with other's eyes to themselves interacting with a robot, the politeness of their behavior might decrease. As a result, it might increase impolite behaviors such as ignoring.

We probably need to adjust our system to allow its use system elsewhere. For instance, the computation for the location's explanations assumes somewhat large exhibits; if they

were smaller, we must more precisely consider the visibility for a 3D space, for example. Our system is currently designed for just a single user. But we need to consider groups of visitors, and hence the rule system requires extensions, including whether a member in the group has already listened to previous explanations. Different robots will have different types of hands, and so we must change the robot's pointing behaviors and address their comprehensiveness.

In this study we used RFID tags in which personal information was entered to identify visitors. Since RFID tags are becoming cheaper and cheaper, there would be no problem to use them in actual museums. However, as a more practical solution, we could use smartphones that visitors have. For example, the following application would be practical: Visitors enter personal information such as names and ages on the application on their smartphones. That information is stored in a database on a cloud, and the application generates a QR code to identify the visitors. The visitors can be personally identified by holding the QR code over code readers in an exhibition.

Although this approach is practical, it is still difficult to correspond to groups of visitors because everyone in the group has to make the code reader read the QR code. In order to deal with a group without boring operations, we should use face images. If visitors' face images are associated with their personal information when they enter the personal information, the personal identification would become possible by face recognition. Given the development of current face image recognition accuracy, we believe that such an application may be sufficiently feasible in the future.

6.3 Benefits of Using Robots Instead of Human Guides

Regarding benefits of using mobile humanoid robots such as ASIMO instead of humans, such robots are currently very costly and has safety issues. However, as participants in the experiment pointed out, robots have better points than humans. For example, robots can memorize a lot of descriptions of exhibits, visitors' information and visitors' behavior history. This will enable robots to provide information personalized to visitors. Furthermore, robot guide is obligation-free more than humans. In future, we consider that there is a possibility that robots explain exhibits instead of humans if the robots become cheaper and safer.

6.4 Principle of Interaction for Human-Like Guide Robots in Museums

We show principles of interaction for human-like guide robots in museums that we obtained through this study. The principles are the following four things:

To approach to a visitor if he or she is paying attention to an exhibit If a visitor is paying attention to an exhibit, the visitor is likely to be interested in the exhibit. Therefore, the robot needs to approach to the visitor for giving an explanation, as human guides do so (see the Sect. 3.1). In this study, the robot estimated which exhibit the visitor paid attention to by recognizing visitor's position and attention. Then, following the episode rules, the robot performed an appropriate approach behavior.

To take a position suitable for an explanation Usual exhibits have several parts to be explained. For example, a car has tires, engines, headlights and so on. When explaining the car, the robot would be unable to explain headlights from back of the car. Thus, when the robot explains a part of an exhibit, the robot should take a position where both the robot and a visitor can look at the part. In this study, we defined explainable positions of each exhibit. The robot could move to positions suitable for explanations by incorporating the positions into the calculation formula of the approach behavior.

To explain considering history of interaction with a visitor It is extremely important not to repeat an explanation once given to a visitor. In addition, to explain an exhibit based on the previous explanation will lead to a deep understanding of the exhibit. For example, after the robot explained the outline of an exhibit, if the visitor was still looking at the exhibit, the robot should explain a detailed topic such as technical information and principles of operation. In this study, the robot could explain considering history of interaction with each visitor by the behavior selector module.

To present a relationship with a visitor (especially a repeater) Avid fans such as club members of a museum visit the museum several times a year. The robot should not treat these visitors like a first meeting. For repeaters, showing behavior different from ordinary visitors leads to their satisfaction. In this study, for repeaters, the robot approached quickly, closed positioning at approach, and showed self-disclosure and compliment. The repeaters who participated in our field trial were happy to understand the relation-building behavior.

6.5 Future Works

In order to realize a more practical robot guide system, we have to consider the following things: recognition of visitor's intention and treatment of multiple visitors.

It is important to recognize whether a visitor who are looking at an exhibit really wants an explanation from a guide. If it was achieved, the robot would not disturb anyone who wanted to see the exhibit in quiet and alone. In addition, if the robot was able to recognize whether a visitor wanted more explanation or wanted the robot to stop explaining, the robot could guide more flexibly.

Regarding the treatment of multiple visitors, from the view of a function of the system, the current robot can also handle groups by extending the condition part of the episode rules. However, since the situation where groups of visitors look around becomes complex [37], we need to model the situation and define episode rules for appropriate behaviors in the situation. For example, even with regard to the order of explanation, it is unclear that the robot should approach to which of the following visitors: a visitor who is looking at an exhibit for a long time, a repeater, or a family including children.

To solve those problems, we will make a more human-like practical guide robot.

Acknowledgements We thank Honda R\&D Co., Ltd. and the Miraikan staff for their help.

Funding This study was funded by JST CREST Grant Number JPMJCR17A2, Japan, and JST, PRESTO Grant Number JPMJPR1851, Japan, and by JSPS KAKENHI Grant Numbers 18H03311.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Burgard W et al (1998) The interactive museum tour-guide robot, national conf. on artificial intelligence (AAAI1998), pp 11–18
2. Thrun S et al (1999) Minerva: a second-generation museum tour-guide robot. In: IEEE international conference on robotics and automation (ICRA1999), pp 1999–2005
3. Siegwart R et al (2003) Robox at Exp.02: a large scale installation of personal robots. Robot Auton Syst 42:203–222
4. Gockley R et al (2005) Designing robots for long-term social interaction. In: IEEE/RSJ international conference on intelligent robots and systems (IROS2005), pp 1338–1343
5. Gross H-M et al (2008) Shopbot: progress in developing an interactive mobile shopping assistant for everyday use. In: IEEE international conference on systems, man, and cybernetics (SMC2008), pp 3471–3478
6. Shiomi M et al (2008) A semi-autonomous communication robot—a field trial at a train station. In: ACM/IEEE international conference on human–robot interaction (HRI2008), pp 303–310
7. Kanda T, Shiomi M, Miyashita Z, Ishiguro H, Hagita N (2010) A communication robot in a shopping mall. IEEE Trans Robot 26:897–913
8. Bennewitz M, Faber F, Joho D, Schreiber M, Behnke S (2005) Towards a humanoid museum guide robot that interacts with multiple persons. In: IEEE-RAS international conference on humanoid robots (Humanoids'05), pp 418–423
9. Pateraki M, Trahanias P (2017) Deployment of robotic guides in museum contexts. In: Ioannides M, Magnenat-Thalmann N, Papagiannakis G (eds) Mixed reality and gamification for cultural Heritage. Springer, Cham, pp 449–472
10. Wang S, Christensen HI (2018) Tritonbot: first lessons learned from deployment of a long-term autonomy tour guide robot. In: 2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN), IEEE, pp 158–165
11. Pang WC, Wong CY, Seet G (2018) Exploring the use of robots for museum settings and for learning heritage languages and cultures at the Chinese heritage centre. PRESENCE: Teleoperators Virtual Environ 26(4):420–435
12. Sidner CL, Kidd CD, Lee C, Lesh N (2004) Where to look: a study of human–robot engagement. In: International conference on intelligent user interfaces (IUI 2004), pp 78–84
13. Mutlu B, Shiwa T, Kanda T, Ishiguro H, Hagita N (2009) Footing in human–robot conversations: how robots might shape participant roles using gaze cues. In: ACM/IEEE international conference on human–robot interaction (HRI2009), pp 61–68
14. Kuzuoka H, Oyama S, Yamazaki K, Suzuki K, Mitsuishi M (2000) Gestureman: a mobile robot that embodies a remote instructor's actions. In: ACM conference on computer-supported cooperative work (CSCW2000), pp 155–162
15. Scassellati B (2002) Theory of mind for a humanoid robot. Auton Robots 12:13–24
16. Hall ET (1966) The hidden dimension. Doubleday, New York
17. Michalowski MP, Sabanovic S, Simmons R (2006) A spatial model of engagement for a social robot. In: IEEE international workshop on advanced motion control, pp 762–767
18. Feil-Seifer D, Matarić M (2012) Distance-based computational models for facilitating robot interaction with children. J Human–Robot Interact 1:55–77
19. Walters ML, Oskoei MA, Syrdal DS, Dautenhahn K (2011) A long-term human–robot proxemic study. In: IEEE international symposium on robot and human interactive communication (RO-MAN2011), pp 137–142
20. Yamaoka F, Kanda T, Ishiguro H, Hagita N (2010) A model of proximity control for information-presenting robots. IEEE Trans Robot 26:187–195
21. Bickmore TW, Picard RW (2005) Establishing and maintaining long-term human–computer relationships. ACM Trans Comput–Human Interact 12:293–327
22. Kanda T, Hirano T, Eaton D, Ishiguro H (2004) Interactive robots as social partners and peer tutors for children: a field trial. Human—Comput Interact 19:61–84
23. Lee MK et al (2012) Personalization in Hri: a longitudinal field experiment. In: ACM/IEEE international conference on human–robot interaction (HRI2012), pp 319–326
24. Huang C-M, Iio T, Satake S, Kanda T (2014) Modeling and controlling friendliness for an interactive museum robot, robotics: science and systems conference (RSS2014)
25. Hasse Jørgensen S, Tafdrup O (2017) Technological fantasies of nao—remarks about alterity relations. Transformations 29:88–103
26. Shigemi S, Kawaguchi Y, Yoshiike T, Kawabe K, Ogawa N (2006) Development of New Asimo. Honda R\&D Tech Rev 18:38
27. Brscic D, Kanda T, Ikeda T, Miyashita T (2013) Person tracking in large public spaces using 3d range sensors. IEEE Trans Human—Mach Syst 43:522–534
28. Glas DF, Satake S, Kanda T, Hagita N (2011) An interaction design framework for social robots, robotics: science and systems conference (RSS2011)
29. Shi C, Shimada M, Kanda T, Ishiguro H, Hagita N (2011) Spatial formation model for initiating conversation, robotics: science and systems conference (RSS2011)

30. Kendon A, Verasnte L (2003) Pointing by hand in "Neapolitan", in pointing where language, culture, and cognition meet, Kita S (ed)
31. Imai M, Ono T, Ishiguro H (2003) Physical relation and expression: joint attention for human–robot interaction. IEEE Trans Ind Electron 50:636–643
32. Heerink M, Kröse B, Wielinga B, Evers V (2008) Enjoyment, intention to use and actual use of a conversational robot by elderly people. In: ACM/IEEE international conference on human–robot interaction (HRI2008), pp 113–120
33. Aron A, Aron EN, Smollan D (1992) Inclusion of other in the self scale and the structure of interpersonal closeness. J Pers Soc Psychol 63:596
34. Kaneko K, Kanehiro F, Morisawa M, Akachi K, Miyamori G, Hayashi A, Kanehira N (2011) Humanoid robot hrp-4-humanoid robotics platform with lightweight and slim body. In: 2011 IEEE/RSJ international conference on intelligent robots and systems, IEEE, pp 4400–4407
35. Pandey AK, Gelin R (2018) A mass-produced sociable humanoid robot: pepper: the first machine of its kind. IEEE Robot Autom Mag 25(3):40–48
36. Ishiguro H, Ono T, Imai M, Maeda T, Kanda T, Nakatsu R (2001) Robovie: an interactive humanoid robot. Ind Robot: Int J 28(6):498–504
37. Trejo K, Angulo C, Satoh SI, Bono M (2018) Towards robots reasoning about group behavior of museum visitors: leader detection and group tracking. J Ambient Intell Smart Environ 10(1):3–19

Takamasa Iio received the Ph.D degree from Doshisha University, Kyoto, Japan, in 2012. He is an assistant professor at University or Tsukuba, Ibaraki, Japan. His research interests include social robotics, group conversation between humans and multiple robots and social behaviors of robots.

Satoru Satake received the B.Eng., M.Eng., and Ph.D. degrees from Keio University, Yokohama, Japan, in 2003, 2005, and 2009, respectively. He is currently a Researcher with the Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan. His current research interests include human-robot interaction, social robots, and field trial.

Takayuki Kanda is a professor in Informatics at Kyoto University, Japan. He is also a Visiting Group Leader at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. He received his B. Eng, M. Eng, and Ph. D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively. He is one of the starting members of Communication Robots project at ATR. He has developed a communication robot, Robovie, and applied it in daily situations, such as peertutor at elementary school and a museum exhibit guide. His research interests include human-robot interaction, interactive humanoid robots, and field trials.

Kotaro Hayashi received an M. Eng and Ph.D. degrees in engineering from the Nara Institute of Science and Technology, Nara, Japan in 2007, 2012. Since 2009, he has been a doctoral student at the Graduate School of Information Science, Nara Institute of Science and Technology, Nara, Japan and is currently a assistant professor in the Toyohashi University of Technology.His current research interests human-robot interaction, interac- tive humanoid robots, Human factors and field trials. He is the Robotics Society of Japan.

Florent Ferreri received his B.S. and M.S. degrees in Engineering specialized in network security from Telecom Sud-Paris (formerly INT), Evry, France, in 2009. He was a Research Engineer at the Advanced Telecommunications Research Institute International (ATR) until 2018 where he worked on network robot systems, ubiquitous sensing, autonomous navigation, and robot behavior design in human–robot interaction. He is currently working on perception in the context of autonomous driving at Ascent Robotics.

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