

What should we know to develop an information robot?

Satoru Satake, Keita Nakatani, Kotaro Hayashi, Takyuki Kanda and Michita Imai

ATR Intelligent Robotics and Communication Laboratory, Kyoto, Japan

ABSTRACT

This paper is aimed at identifying the required knowledge for information robots. We addressed two aspects of this knowledge, ‘what should it know’ and ‘what should it do.’ The first part of this study was devoted to the former aspect. We investigated what information staff know and what people expect from information robots. We found that there are a lot of similarities. Based on this, we developed a knowledge structure about an environment to be used to provide information. The developed knowledge structure worked well. In the field study we confirmed that the robot was able to answer most of the requests (96.6%). However, regarding the latter aspect, although we initially replicated what human staff members do, the robot did not serve well. Many users hesitated to speak, and remained quiet. Here, we found that the knowledge for facilitating interaction was missing. We further designed the interaction flow to accommodate people who tend to be quiet. Finally, our field study revealed that the improved interaction flow increased the success ratio of information providing from 54.4% to 84.5%.

Subjects Robotics

Keywords Information-providing, Direction giving, Belief about robots

INTRODUCTION

Direction giving is often considered as a desired task for social robots and embodied agents (*Cassell et al., 2002; Kopp et al., 2008; Ono, Imai & Ishiguro, 2001; Okuno et al., 2009*). In our daily life, one of the roles that frequently offer direction giving is information service (*Fig. 1*). Such information booths/counters can be found in stations, airports, shopping malls, and sightseeing places.

We wondered what would be the required ‘knowledge’ to develop for a robot that engages in such an information service. Probably, most of us have experienced using information services, and many of us believe that we know what the information services are. Thus, one would argue that it is just easy to develop such a robot. One might say that “I know from common sense what the information service is. I can just implement it.” Is this true?

We started the study with two research questions:

- Is our common knowledge about the tasks of information services (i.e., what they serve) applicable to information robot?

Submitted 14 March 2015

Accepted 28 May 2015

Published 10 June 2015

Corresponding author

Satoru Satake, satoru@atr.jp

Academic editor

Yochu Jin

Additional Information and
Declarations can be found on
page 21

DOI 10.7717/peerj-cs.8

© Copyright

2015 Satake et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS



Figure 1 Information service.

- Can we create an information robot by replicating what human information staff knows and does? Or, is there any missing knowledge?

We first investigated what people would expect from an information robot, and confirmed that there are a lot of similarities with what human information staff does (section ‘Information in a Shopping Mall’). Thus, we decided to use knowledge about human information staff (what they know and what they do), and developed an information robot (section ‘System’). However, in regard to the second research question, the assumption was not true. Thus, we further investigated missing knowledge (sections ‘Preliminary Trials: Lack of ‘Knowledge’ for Interaction’ and ‘Field Experiment’).

RELATED WORKS

Information-providing robots

Robots have been deployed as tour guides. There were a couple of museum robots that navigated around the environment and provided explanations (*Thrun et al., 1999*). Robots

are also used for interactive information-providing. For instance, *Gross et al. (2009)* developed an article search robot which enables visitors to request an item and let the robot navigate to its location. Input for these robots is often with GUIs, thus there are lists of destinations/items, one of which is chosen by the user.

In contrast, in case of dialog-based system, the difficulty is to predict the set of requests users could ask. Thus, there are assumptions made for the input, such as name of locations. For instance, a virtual agent, Mack, developed by *Cassell et al. (2002)* is able to respond with the names of locations and people in offices, and provide direction giving (*Kopp et al., 2008*). But, in a real-world natural interaction, what users would ask for is not bound by such assumptions. In *Kanda et al. (2009)*, the robot provided direction-giving interaction in response to the name of locations in a shopping mall, and exceptions were handled by a human operator. That is, the system on its own did not address the questions beyond the assumptions.

Overall, in the previous studies, there was not much exploration for what people would ask/request in an information dialog with a robot. In contrast, we found that people ask various requests beyond the names of locations, and identified a required knowledge representation.

Direction-giving interaction

It is reported that a good direction consists of pairs of actions and landmarks (*Daniel et al., 2003*), such as “turn right at the post office, and . . .” To provide such explanations, there is a technique to build a knowledge about spatial relationships among shops and corridors (*Morales et al., 2011*). There are techniques to make a robot understand directions from humans (*Kollar et al., 2010*); in the study of Kollar and colleagues, the representation stores the relationship between the description of the entities in the space and the map.

In these studies, the common assumption is that a system is able to provide directions if the name of a location is asked. In contrast, our study reveals other type of requests in information dialog, and we report on the required knowledge representation.

Note that it is well known in HRI studies that gaze and pointing gestures make the interaction more natural and effective (e.g., *Sidner et al., 2004*; *Mutlu, Forlizzi & Hodgins, 2006*). The use of gesture in direction giving is also studied in conversational agents (*Kopp et al., 2008*) as well as in human-like robots (*Ono, Imai & Ishiguro, 2001*; *Okuno et al., 2009*). Our direction-giving behavior is informed by these studies.

Engagement

In our study, we noticed some visitors remained silent, even after directly approaching the robot and hearing its requests to engage. Relevant to this, there were studies about “engagement” process; that is, when people participate and feel connected in collaboration, their gaze will meet with each other and they do not quit the interaction (*Sidner, Lee & Lesh, 2003*). Rich and his colleagues developed a technique to detect engagement using people gaze (*Rich et al., 2010*). Kobayashi and his colleagues developed a technique to select a person to whom a robot should ask questions in multi-party interaction in the way a teacher appoints a student in a class for an answer (*Kobayashi et al., 2010*). Their technique

is based on the findings that people nodding and engaging in mutual gaze are more likely to answer than someone avoiding a meeting gaze. In contrast, the silent visitors in our study were people who voluntarily approached the robot. They typically behaved as if they were willing to interact with the robot but did not talk with it.

Knowledge representation

There are several computer applications (e.g., Google search or Apple's Siri) that provide information related to location. There are many similar aspects between our robot and such applications, e.g., both need connection between language and local knowledge, interpretation needs to be contextual, and answers to be provided in verbal way. Thus, similarly to these approaches, we used ontology (McGuinness & Van Harmelen, 2004) to build the knowledge representation. However, we need to build our own knowledge representation because required the knowledge structure is different, and we cannot simply apply existing software like Google search and Siri for the robot. For instance, robots can use pointing gesture (also, often robots are not equipped with display), which very much changes the way of giving direction.

INFORMATION IN A SHOPPING MALL

We investigated the daily tasks of information service employees and what visitors typically expect from robots acting as such. We found a lot of similarities. The study protocol was approved by institutional review boards of Advanced Telecommunications Research Instituted International with reference number 14-502-2.

Daily tasks of information service

We interviewed two employees working at the information desk of a shopping mall.

First we asked an overall description of their job: they usually wait for visitors to come to the information booth. They were requested by the mall administrators to serve as 'information staff.' Only procedures for lost items were provided; for other tasks (e.g., information providing) they use their common sense.

Further, we asked them to categorize the typical requests from visitors, and how they would respond. Both reported that there are three types of requests:

Direction giving: They reported that this is the most frequent request. Visitors ask simple where-type questions, e.g., "where are the toilets?" In addition to the name of locations, people use other popular name, like "hello show," or the name of designated areas, like "smoking area." Their typical response is to provide turn-by-turn directions using utterance and pointing gesture. When visitors do not understand, they sometimes write down to a map, or on rare occasions take them to the destination.

Recommendation (inquiry): When a visitor does not know whether there are shops that meet his needs, he may query the information staff. Visitors may inquire of the characteristics of shops, such as name of items, and the category of shops. Here are some examples of questions: "Are there Japanese restaurants?"; "Are there shops that sell Osaka souvenirs?" The staff members typically verbally list the shops or events that meet their criteria. Visitors sometimes ask for a recommendation from the staff without providing

solid conditions but only using subjective words e.g., “Are there any good restaurants?” For such requests, the staff members reported that they typically try not to give a subjective preference, because their preferences may or may not match with those of the visitors. Thus, their responses for inquiry and recommendation are similar: they try to objectively reply and provide a list of shops that seem appropriate.

Lost child and lost-and-found: When children are lost, or when visitors lose items, they come to the information desk. For lost children, the staff usually makes a public announcement throughout the shopping mall. Lost items can be retrieved at the information booth when available upon confirmation of ownership.

Expectations from information robot

To investigate what people expect from information robots, we interviewed customers in the shopping mall. To find people who would be willing to help us with collecting knowledge for future robots, we prepared a situation where visitors can see a robot in the midst of interaction. Thus, we prepared a robot for information, which is controlled with Wizard-of-Oz method. We then asked people who stopped around the robot and/or interacted with it to participate in the interview. Twenty-one visitors participated in the interview.

In the interview, we asked the visitors to imagine future situations in which robots would be capable of offering information services they like, regardless of their previous observations of the robot’s capability. We then asked them to freely provide as many functions they would like information robots to have.

The interviews were recorded and transcribed for analysis. We categorized the different kind of requests expressed by the visitors, For instance, visitors reported sentences such as:

“I often look for the smoking area, thus I would like to ask the robot about it.”

This utterance was coded as expectation for *direction giving*, because we interpret it as ‘where’-type question in which visitors simply want to know the location. The followings ones were coded as expectation for *recommendation (inquiry)*:

“I’d like to know about sports and furniture shops.”

“The shop which sells the most? Well, I want the robot give me recommendations of shops.”

Such cases were classified as *recommendation (inquiry)*, because visitors need to know more information than just a location.

Then, two coders who do not know the research purpose judged whether each transcribed sentence would fit into the above defined categories, or not (which is categorized as ‘other’). The judgement of the two coders matches reasonably well, yielding kappa coefficient .857.

[Table 1](#) shows the coding result. The ratio of visitors who mention the expectation is listed in each row. They can provide multiple answers, thus the sum of the ratios exceeds 100%.

The expectation of the visitors for the information robot largely overlaps with what human information services provide. Almost all visitors (20 out of 21) mentioned that they expect direction giving and the majority (16 out of 20) reported that they expect the

Table 1 The analysis result of expectation for information.

Expectation	Ratio
Direction giving	95.2%
Recommendation (inquiry)	76.2%
Other	
Playing with children	23.8%
Lost child	4.8%

robot to offer turn-by-turn direction accompanied with pointing gesture. For instance, one spontaneously mentioned the practicality of pointing gesture in directions giving:

“Well, ‘where,’ umm, I did not understand ‘which’ direction I should go. So it would be useful if the robot could do pointing gestures,”

There were 3 visitors who expected the robot to take them to the destination, and 1 visitor who wanted the robot to explain with a map.

There were 16 people that expected a recommendation service. For instance, some mentioned “I’d like to have some recommendations for restaurants,” or “I’d like to know places where children can play around.” Others wanted to have more detailed explanations. For instance, one commented:

“I’d like to know what kind of shop it is, its atmosphere, what it sells, and so on.”

In contrast, the ‘playing with children’ category is specific to the information robot. We collected comments such as:

“Interacting with the robot was enjoyable. This is good for people who come with their children”.

“Many families only have one child. It would be nice if the robot behaved like a brother.”

Requirements

The expectations for information robots largely overlapped with what is delivered at human information services. That is, most of them expect two services: direction giving and recommendations. Thus, in this study, we focused on these two services.

Further, we investigated the required knowledge to be stored. We analyzed the utterances of the requests. We labeled them based on the type of request. For instance, we assigned a label ‘name of location’ to the utterance “I’d like to know where the event *dream world* takes place,” ‘name of item’ to the utterance “I’d like to know where can I buy *coffee*.” If multiple labels are applicable we assigned all of them. Labels are merged when possible, resulting in 6 different labels. To confirm the classification, we asked two coders who did not know the purpose of the research to classify the utterances based on the 6 labels. Their coding matches reasonably well, yielding kappa coefficient of .637.

Finally, we identified that the following information is needed:

- (1) Name of location: such as names of shops or names of events. In addition to the formal name, people use various nicknames. 78.3% of people mentioned this category.

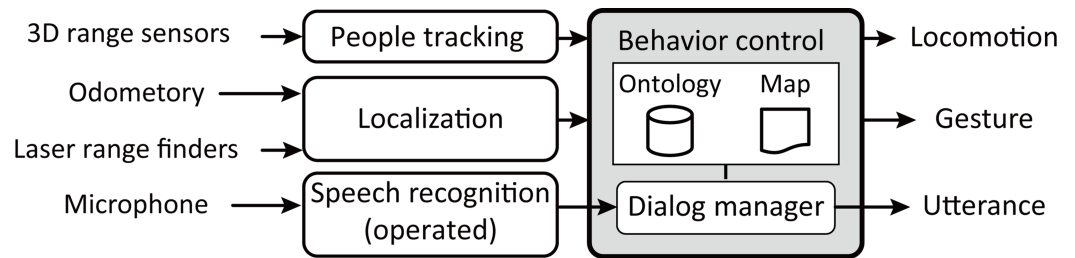


Figure 2 System architecture.

- (2) Item name: people look for specific product or entity available in shops. For instance, this category includes items such as “cell phone charger” and “coffee.” 47.8% of people mentioned this category.
- (3) Category: shops can usually be grouped into larger categories, like “restaurant,” “Japanese restaurant.” 52.2% of people mentioned this category.
- (4) Features: shops are usually recognized as some generally-known features, like “good view,” “expensive,” and “recommended.” 65.2% of people mentioned this category.
- (5) People activity: locations are sometimes referred as the activity that people do there, like “play,” “eat,” “shop.” 60.9% of people mentioned this category.
- (6) People’s state: locations are sometimes referred as the place appropriate for people’s physical condition, like “injured,” “tired,” “hungry.” For instance, some visitors said: “I would like to receive recommendation, just by saying ‘I’m hungry’ for example.”

13.0% of people mentioned this category; note that this request was not reported by the information desk staff, thus it can be considered as specific to the information robot.

Based on this analysis, we developed the knowledge representation for the information robot.

SYSTEM

Architecture

Our goal is to develop a robot that autonomously provides information services. Based on the analysis in section ‘Information in a Shopping Mall,’ we developed a knowledge representation that can be used by such a robot. Figure 2 shows the architecture of the system. Information from sensors goes through modules like *people tracking* (explained in section ‘People tracking’), *localization* (section ‘Localization’), and *speech recognition* (section ‘Speech recognition (with human operator)’). Output from these modules are used in the *behavior controller* (section ‘Behavior controller’), which contains a *dialog manager* (section ‘Dialog manager’). The environmental knowledge is stored in *ontology* (section ‘Ontology of entities in the map’) and *map* (section ‘Route perspective map’), and used by the dialog manager. We explain these modules in the later section.



Figure 3 Environment of the shopping mall.

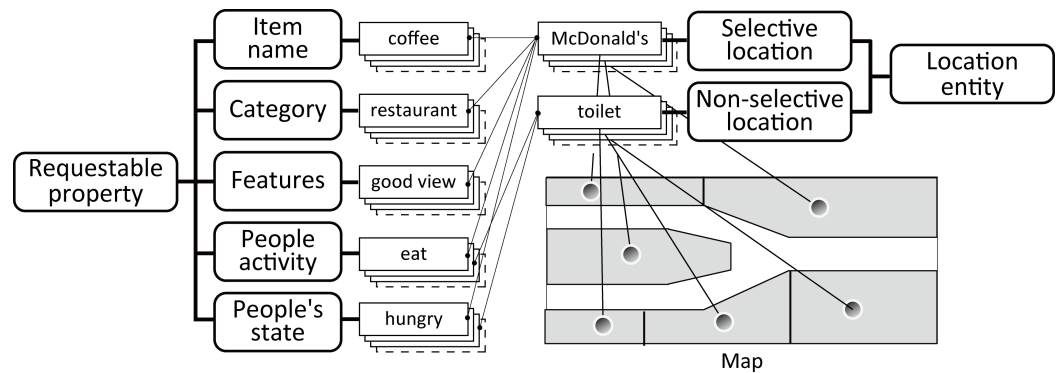


Figure 4 Knowledge representation for the environment.

Knowledge representation

There are two types of information in the knowledge representation. One is the map used for direction giving (explained in section ‘Route perspective map’). The other is shop-related data (explained in section ‘Ontology of entities in the map’).

Environment

The study was conducted in a big shopping mall located in a suburban area. It consists of three buildings (Fig. 3A), one having 12 floors, and others having 6 floors. There are 51 shops, 31 restaurants, 42 facilities, 6 event halls, 4 squares (e.g., Fig. 3B), 2 stages, and many offices. The mall is mainly busy during weekends. Almost all shops are for non-daily goods, like clothes, shoes, sports, outdoor activities. We often observe people who look for shops and locations (e.g., they look at the floor maps, and/or ask the service staff). The main hall where big events take place is located far (a 5 min of walk) from the square where we put the robot, thus people often asked where an event was taking place.

Ontology of entities in the map

We designed our knowledge representation for ‘request’ and ‘shops’ together using an ontology language, OWL (McGuinness & Van Harmelen, 2004). Figure 4 shows the designed knowledge structure, i.e., ontology. The basic element in OWL is the ‘class,’ which has ‘properties’ that store the information. There are two primary classes, ‘location entity’ and ‘requestable property’ prepared.

Table 2 Possible relationships.

Users' request	Possible relation
Item name	Is sold at/is served at/is at
Category	Belongs to
Features	Is a feature of
People activity	Is possible at
People's state	Is satisfied/healed/solved at

Location entity

We define entities like shops, facilities and events as instances of the 'location entity' class. There are three properties:

- (1) Name: we stored the official or commonly used name.
- (2) Nicknames: some shops are referred to with a nickname. We listed such nicknames people could use. For example, "Kentucky Fried Chicken" is referred as "KFC."
- (3) Location on the geometrical map: each location is associated with the geometrical map (explained in the section 'Route perspective map').

We further separate the class into two subclasses, *selective location* and *non-selective location*. When multiple locations are available, people would prefer to select one. For instance, if there are two Italian restaurants, people would choose one based on their own criteria, such as better, cheap, popular, etc. We store one extra property, 'introduction property,' in *selective location* to be used in dialog to help people selecting locations. In contrast, people would usually not care about which toilets to which they would go. Such locations are implemented as *non-selective location* class.

Requestable property

There are six types of information communicated in information dialog (section 'Requirements'). Except for name of location, they are realized as 'requestable property' class, which has subclasses 'item name,' 'category,' 'features,' 'people activity,' and 'people's state.' When a user requests information, it is turned into an instance of the 'requestable property.' Then, the location(s) having the same property will be searched. Each property item has wordings that are expected to be used in people's utterance. For instance, 'eat' (instance of people's activity subclass) is associated with wordings such as "eat," "have lunch," and "have a meal." Note that more complex requests (e.g., "Japanese" restaurant with a "good view") can be represented as multiple instances combined with 'and/or' operators, but we did not implement such complex operations because users rarely made such complex requests.

Relationships between 'location entity' and 'requestable property'

Table 2 shows possible relationships between two subclasses. For instance, some visitors could request a restaurant where they can have "pasta." To handle such requests, a "pasta" entity is prepared as an instance of 'item name' subclass which is associated with shops with

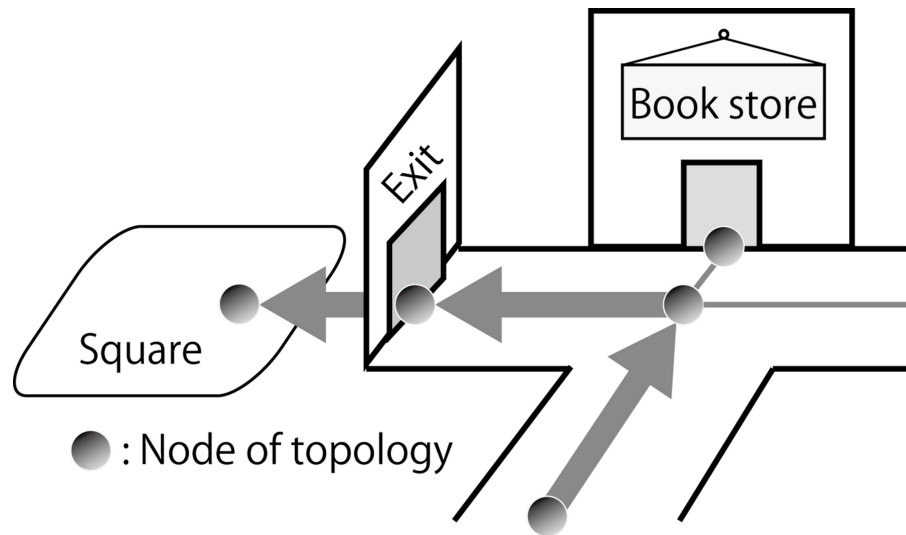


Figure 5 Illustration of the route perspective map.

the relation ‘is served at.’ Such relation is defined inside dialog management (section ‘Behavior controller’) as well. Note that an instance of ‘requestable property’ can be associated with multiple ‘location entities’ (e.g., “pasta” can be served at multiple restaurants).

Finally, we prepared the data for the shopping mall (section ‘Environment’). There are 201 location entities (84 shops, 75 service facilities, 39 events, and 3 buildings) with in total 3,345 nicknames. There are 530 requestable properties (501 items, 163 categories, 44 features, 63 people activities, 22 people’s states) prepared as well.

Route perspective map

Informed by [Morales et al. \(2011\)](#), we manually prepared a route perspective map (illustrated in [Fig. 5](#)), which consists of pairs of landmarks and actions. Using the map, the system generates turn-by-turn directions giving, such as “go straight, turn left at the book store, go out the door with exit sign . . .” The map includes the following information:

- (1) **Topological map:** Nodes are located at decision points in the map. Transition through corridor or between different floors, such as stairs, escalators, and elevators, are expressed as movements between nodes. Entrances of shops, facilities, and events (i.e., *location entity*) are also represented as nodes.
- (2) **Landmarks:** If available, visible landmarks are manually associated for each route as denoted in [Morales et al. \(2011\)](#), e.g., famous shop names with salient signboards, elevators, and escalators.
- (3) **Actions:** In [Morales et al. \(2011\)](#), actions were only turning behaviors, which were computed from a topological map. In contrast, as there are many floors and multiple buildings, we added actions like “enter the next building,” “go to the 3rd floor.”

Behavior controller

When a person stops by the robot (within 2.5 m for 3.0 s), or is detected as approaching at 2.5 m from it, it starts a dialog. The robot orients its body and gaze to the user. When there is no user, the robot shows liveliness by slightly moving head and arms.

During the dialog, its head and body is oriented toward the user, except for the moment when it performs a pointing gesture which is often used when giving directions. When it points at a direction, its head direction is oriented toward the pointed direction for the first three seconds of pointing in order to draw the user's attention toward the pointed direction. The robot ends the dialog when the user leaves the robot's side (3 m away), or when the dialog management module decides to end the dialog.

Dialog manager

We developed a rule-based mechanism for dialog management. Assuming that there is an input coming from the speech recognition module (explained in 'Speech recognition (with human operator)'), the input is turned into text and matched with name/nickname properties of location entities and with instances of *requestable properties* (explained in 'Ontology of entities in the map'). If a requestable property matched, it is compared with *location entities*.

When only non-selective locations are matched, it chooses the nearest one. In case the user asked for a location with a specific *name of location*, there should be only one location to be matched. In these cases, the system provides *direction-giving* dialog, in which turn-by-turn directions to the location are generated.

Otherwise, it initiates a *recommendation* dialog. It verbally lists the locations that match with the *requestable property* instance one by one. For each location, it explains the location using the text in its introduction property. For instance, it utters "Ramen is served at a ramen restaurant named *Kaika-ya*. They serve a ramen with tuna soup. May I explain the directions to go there?" As human staff does, we carefully avoid telling subjective preferences, but only provided objective facts.

In addition, it reacts to the words for greeting. When an input matches with words like "hello," it returns a greeting utterance. When an input matches with leave-taking words like "bye," it returns leave-taking words and ends the dialog.

When no location is matched, the system explains that "(requested item) is not in this shopping mall. I only know about this mall."

Other modules

Robot

We used a robot characterized by its human-like physical expressions. It is 120 cm high and 40 cm in diameter on a mobile platform. It has a 3-DOF head and 4-DOF arms. There are two 30 m range laser sensors attached. We used the robot with a maximum speed of 550 mm/sec and 50 °/s for rotations. The accelerations are set to 300 mm/s² and 50 °/s². To clearly communicate its role, we put an 'information staff' sign in Japanese on the chest of the robot (Fig. 1).

People tracking

We use a people tracking method described in [Brscic et al. \(2013\)](#), which provides an estimation of the location of pedestrians every 33 ms. It covers the square we used. There are 49 3-D range sensors attached on the ceiling (combination of Panasonic D-Imager, ASUS Xtion, and Velodyne HDL-32E).

Localization

For robot localization, we use a particle filter with a ray tracing approach on a grid map ([Fox, Burgard & Thrun, 1999](#)). The grid map is built from odometry and laser scanner data. This module is called every 350 ms and updates the robot's position.

Speech recognition (with human operator)

We developed fully-autonomous system using ASR (automatic speech recognition), but in order to better test the overall framework we used a human operator instead of ASR.

Automatic speech recognition (ASR). We used an ASR software, ATRASR ([Matsuda et al., 2006](#)). It uses a language model based on FSA (Finite State Automaton). We constructed the language model mainly using the terms appeared in the ontology.

With preliminary trials using the Wizard-of-Oz approach, we analyzed the way visitors speak to the robot. In total, 470 requests collected over 3 days of preliminary trials. From the analysis of the requests, we found that they mainly follow three ways of speaking, as follows:

- Noun/adjectives only:

People only spoke words like a *name (nickname) of location, category, or item name*, such as “restaurant,” and “coffee.” Sometimes, for *features and people's activity*, they add such terms like “place for” (eat/lunch/play). Some ontology items are adjectives, such as “tired.” People sometimes only spoke such adjectives.

- “Where is” question:

The above noun is used in “where is” question, such as “where is *Kaika-ya* (the name of restaurant) ?”

- “I would like to” sentence:

People also use the form of “I would like to” + “verb” + “noun” in requesting sentences, such as “I would like to buy coffee.”

For all *names, nicknames, and requestable properties*, we automatically generated grammatical structures for ASR. Further, we added the following grammars. First, some basic verbs like “go” can be used in “I would like to” type sentences but were not included in the ontology (as they by themselves does not represent any specific request), which we manually added (8 verbs). Second, we added filler words, such as “well,” “ah,” that appear in advance to questions (12 words). Third, to eliminate noises from environments, like sounds from people's walking, whistle from ships, we added some fillers (66 fillers). Overall, we prepared the lexicon whose size is 1,469 with 4,938 links.

The ASR outputs the matched *names*, *nicknames*, or *requestable properties*, which are used in the dialog manager to determine the answer to be provided. In case the ASR detects the recognition to be less reliable (because the input does not match well with its language model), the dialog manager prompts the user to say again with utterances like “could you repeat please?” The ASR is deactivated while the robot is speaking.

We evaluated the system performance using this ASR implementation. We put the robot on a square of the mall (Fig. 3B), and let the visitors freely use it. With our preliminary test, with 22 users, there are 81 requests, for which the robot was only able to correctly respond in 19.8% of the cases. (In a similar study only 21.3% of successful recognition was achieved [Matsuda et al., 2006.](#))

There were 4 types of errors: error in sound detection error (due to other ambient sounds, the system failed to detect the start of utterance) (17.3%), ASR resulted in low reliability score (30.9%), utterance did not match with the prepared grammar/vocabulary (2.47%), and mis-recognition in ASR (29.6%). In case the mis-recognition occurred, often the system seemed to be interfered by ambient noise, which was matched with some vocabulary in the lexicon.

In contrast, in case ASR successfully detected the *names*, *nicknames*, or *requestable properties*, the system provided appropriate answers. Overall, this preliminary test revealed that the system is capable of handling users’ utterances when the ASR is successful, while we would yet need to wait ASR technologies to be ready for real world environments.

Wizard-of-Oz. The system is ready for autonomous speech recognition. But, for this study, to focus on other parts of interaction rather than working for errors in speech recognition, we used a human operator only to support speech recognition.

We strictly limit the task of the operator, and have him work like the dumb ASR software described in the previous section. We did not allow the operator to add his knowledge. Just like the output from the ASR, the operator only typed the words spoken by the user. For instance, to our knowledge, if a user asks for a “Place for lunch” but such wording is not in the system vocabulary, in previous studies Wizard-of-Oz operators replaced such words to the ones system can handle, like “restaurant”; by doing so, the system can work with a very limited vocabulary and knowledge. Instead, with our system, a novice person who does not know the environment (e.g., list of shops) can easily serve as an operator.

PRELIMINARY TRIALS: LACK OF ‘KNOWLEDGE’ FOR INTERACTION

We conducted a preliminary study with the system reported in the previous section. We initially intended to supplement missing data and evaluate its performance. We found the system itself worked well (we will report in section ‘Evaluation of system performance’); however, interaction failed in other parts we did not think about. That is, some visitors responded in an unexpected way. In short, until this study was conducted, we focused on the ‘information’ aspect, which we found to be satisfyingly prepared, but we found a problem in ‘interaction.’



Figure 6 The interaction failed because the visitor did not speak.



Figure 7 The visitor kept silent after prompted.

Here, we report two typical cases of failures. From these cases of failures, with a trial-and-error approach, we seek the reason why interaction fails and seek for better pattern of interaction for the problem. Finally, we generate hypotheses about missing knowledge in interaction (to be reported in the next section).

Case 1: Interaction did not start

The initial version of the robot imitated the interaction of human information staff. It waited for the arrival of the visitors, and waited for them to make a request. This is what a human staff member would do. The signboard showing ‘information staff’ on the chest of the robot was very visible, so we expected that every visitor would have common expectations as those investigated in section ‘Expectations from information robot.’

However, frequently people would stay in front of the robot without saying anything. [Figure 6](#) shows one of such cases. A man stopped in front of the robot, and the robot was ready to receive a request, orienting its body and head toward him; but, without talking to it, he moved to a side of the robot, and the robot followed. He moved back, and it followed again. Finally, he left after 30 s of silence.

Case 2: Passive visitors

Further, we noticed that the conversation got stuck when it asked for a request, even though the user initially spoke to the robot. For instance, [Fig. 7](#) shows a visitor who engaged in greeting, but came to be silent when prompted to ask request. She left after 5 s of silence after being prompted.

We interpreted that such people do not have concrete requests in their mind, thus they were stuck when asked to offer requests.

FIELD EXPERIMENT

For each case of problems found in the preliminary trial (reported in the previous section), we generated a hypothesis, and conducted an experiment to confirm our idea to supplement such weakness. The study protocol was approved by institutional review boards of Advanced Telecommunications Research Instituted International with reference number 14-502-2.

Experiment 1

Hypothesis

We initially replicated the way human staff interact with visitors. That is, we make it clear that the robot is serving as information staff. Assuming that visitors have the common expectations the purpose of information staff we let the robot wait for a visitor to make a request, and to prompt to request if not asked. However, this assumption was not always correct. Visitors may not share or may be unsure about their expectations of the ‘information robot’ role. If this is the case, we can probably moderate the problem by letting the robot first explain its role (direction giving and recommendation). Thus, we made the following prediction:

Prediction 1: If the robot proactively explains its role as information staff, people will more frequently request information from it.

Participants

The study was conducted during weekends. The participants were visitors of the shopping mall who are typically group of friends and families who come to the mall for leisure. The mall is big and the layout is complicated, thus people are often in real need of getting directions from someone.

When a robot is placed on the mall, people sometimes stopped at the robot. We assumed that such people who stopped at the robot as the participants.

Condition

There are two conditions compared.

- **With self-introduction:** when a person stops, the robot starts self-introduction. It says, “Hello, I can provide directions and recommendations.” Then, it prompts him/her to request “May I provide you some information?”
- **Without self-introduction:** when a person stops, the robot waits him/her to request without speaking to the user.

In both conditions, when a visitor requests it immediately moves into the information dialog. After 20 s of silence, the robot closed the interaction saying “bye-bye.”

Procedure

The robot was placed at a square of the mall (Fig. 3B). We choose this location because visitors often arrive from the nearby escalator, and need direction giving around this

location. The study was conducted during daytime on weekends. We prepared six pairs of 25-minutes time slots. For each pair, two conditions were assigned. Between the slots, we put 5-minutes break, so that visitors are not influenced by the adjacent time slot.

The visitors of the mall were able to freely interact with the robot. There was a signboard showing ‘information staff’ on the frontal side of the robot, which was clearly visible to the visitors. Beyond that, there were no restrictions nor instructions provided to visitors. There was a person ensuring safety, but he stayed behind a column so that his presence was hardly noticeable from pedestrians. In such circumstances, we observed the pedestrians’ natural reaction to the robot.

Measurement

Considering the role of the information staff, we define the success of the interaction as follows:

Success: The case where the robot was able to receive a request and offered appropriate information/service.

We coded the success from the recorded video. Note that we only evaluated people who stopped in front of the robot (more than 3 s) and faced towards it; we consider that letting people stop is beyond the scope of this paper. If the same person interacted multiple times, only the first one was evaluated. Further, we only evaluated one participant per group (i.e., only the first member of the group, who stopped and faced the robot, was counted as our participant), so that the experiment would not suffer from other members’ prior interactions.

Result

In total, there were 238 interactions evaluated, which were coded by two coders who did not know the study hypothesis. One coded the whole data and the second one did confirmatory coding for 10% of the data. Their coding results matches well (kappa coefficient .962).

Figure 8 shows the result of the study. There were 69.0% of the successful interactions in the *with self-introduction* condition, while 54.4% in the *without self-introduction condition*. Typical failure was, like the one shown in **Fig. 6**, when visitors stayed in front of the robot but remained silent even if they were prompted to talk to the robot. Some visitors left in the middle of the conversation, and some explicitly said they did not need service (6 cases in *with self-introduction condition*).

We applied a Chi-square test to evaluate the ratio of success against failures. There is a significant difference between the conditions ($\chi^2(1) = 4.755, p < .05, \phi_c = .141$).

Thus, prediction 1 was confirmed. When the robot provides self-introduction, the interactions ended with success more frequently. We interpret that even though the robot serves an ‘information’ role, people should share a common expectation. Unless it explains its role, some people might fail in using it.

Discussion

It is plausible that there are two sources of failure addressed. One is the belief that the robot can talk to them; another is the expectation that it offers information. We mainly argued

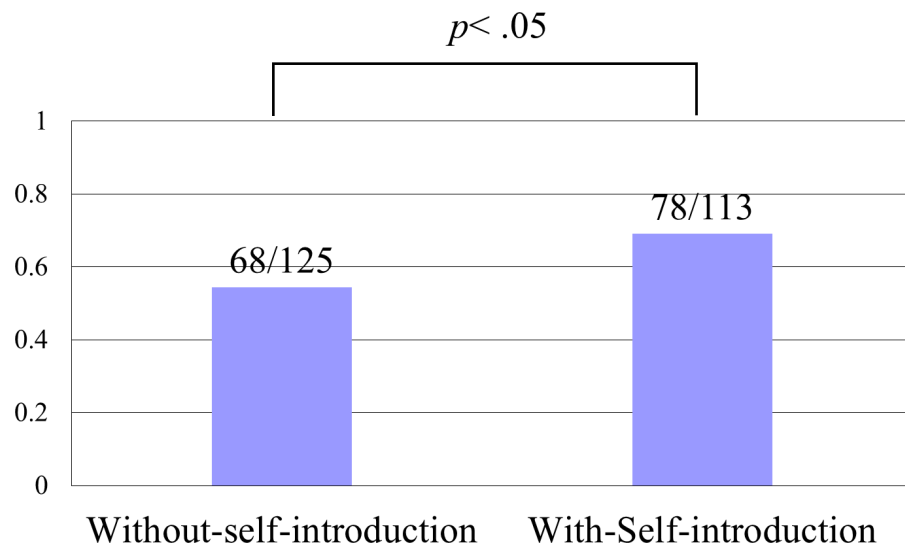


Figure 8 Result of the experiment 1.

the second point, but it simultaneously offered help for the first element. Thus, one would argue that it is better to compare with a robot that only speaks to users but does not provide self-introduction.

However, it was not easy to prepare such a condition when the robot only shows the capability that it can talk in the context of information service. For instance, if it only greeted people, visitors might expect it to engage in variety of interactions, but in reality the robot can only react for the ‘information’ role. Thus, although the effect would be due to both elements, we conducted the study in such a way. It remains as an open question what is the best length of self-introduction. We could make it short and only imply its task by saying something like “May I help you?” We consider that to our observation, people did not get bored due to length of the self-introduction and thus it could be considered as reasonable.

Experiment 2

Hypothesis

In the experiment 1, we found that self-introduction moderated the problem of failure; yet, interaction failed for about 30% of the visitors. We hypothesized that there are visitors who initiated interaction out of curiosity, without a concrete request in mind. Such people would be stuck when a robot prompts them for a request in a direct way. We hypothesized that we can moderate this problem, if the robot turns its offer into a question that they can easily answer. Thus, we made the following prediction:

Prediction 2: If the robot prompts a user for a request in a way of questions they can easily answer, people will more frequently make requests to the information robot.

Participants

The same procedure was used as in experiment 1.

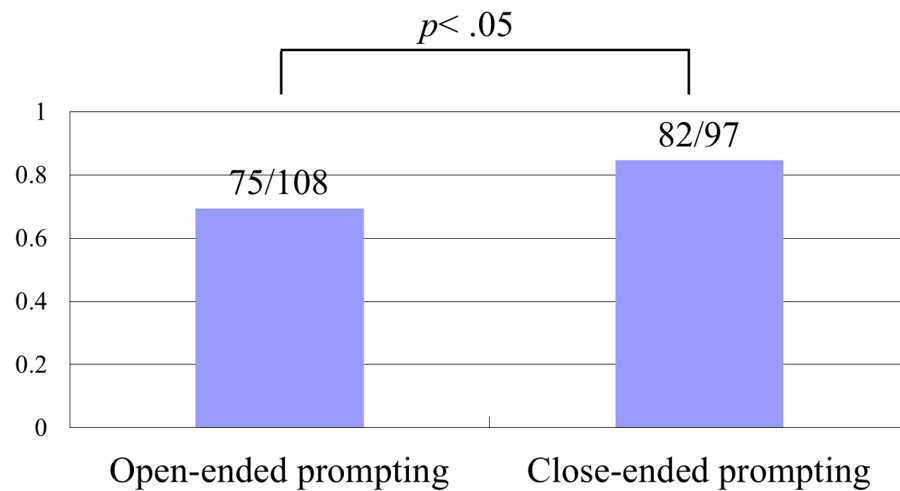


Figure 9 Result of the experiment 2.

Condition

There are two conditions compared. In both conditions, when a person stops, the robot starts with a self-introduction, saying “Hello, I can provide directions and recommendations.” This is identical to the wording used in experiment 1. After a short pause, the robot utters “I will give recommendations based on the locations you are going to,” and prompts the user to ask. The prompting utterance differs depending on the following condition:

- **Open-ended prompting:** It prompts the user by saying “*What kind of recommendation do you wish?*”
- **Close-ended prompting:** It prompts the user by saying “*Where are you going?*”

In both conditions, whenever a visitor requests something to the robot, it immediately moves into the information dialog. If the user keeps silent for 8 s, it once repeats the prompting utterance. If there were 20 s of silence after the prompting utterance, the robot closed the interaction saying “bye-bye.”

Procedure

The same procedure was used as in experiment 1. We prepared seven pairs of 25-minutes time slots.

Measurement

The same measurement was used as in experiment 1.

Result

In total, there were 205 interactions evaluated, which were coded by two coders who do not know the study hypothesis. One coded the all data and second one did confirmatory coding for 10% of the data. Their coding results matches well (kappa coefficient .936).

Figure 9 shows the result of the study. There were 84.5% of successful interactions in *close-ended prompting* condition and 69.4% in *open-ended prompting* condition. Similar to the experiment 1 failure cases, some visitors kept silent when prompted, some visitors left

in the middle, and some explicitly said they did not need the service (3 cases in *close-ended prompting* condition).

We applied a Chi-square test to evaluate the ratio of success against failures. There is a significant difference between the conditions ($\chi^2(1) = 5.678, p < .05, \phi_c = .166$).

The prediction 2 was consequently confirmed. When the robot's prompting was close-ended, the interaction was more frequently successful than open-ended prompting. We interpret that as predicted many visitors did not have requests in mind and got stuck when asked to request; instead, if the robot offered a prompting utterance that invited the user to talk about what they know (e.g., their destination), it will more easily continue the dialog and offer information requested by the user.

Discussion

There are some open questions remaining. One would argue that those who kept silent are people who did not want to 'hear' the information, thus they did not respond to 'hear' questions in close-ended prompting. It is possible that they did not have that much will to spontaneously ask the robot to provide information; nevertheless, in *open-ended prompting* condition, people who were coded as success stayed until the robot finished providing information. One would also argue that the robot could anyway give information even if visitors kept silent. This is possible, and maybe the robot should do so for the remaining 15.5% of people. Our assumption is that it is probably better if they hear information they requested, rather than randomly chosen information. We could not fully clarify why the remaining 15.5% of people who kept quiet in *close-ended* condition. We tried to interview such people, but they did not want to be interviewed.

Evaluation of system performance

Throughout the experiment 1 and 2, the robot was controlled with the system reported in 'System'. In total, there were 435 requests made for the information robot. We analyzed how they were handled, and evaluated whether the robot's responses were correct.

66.8% of the case requests were a *name of location* and 4.0% were a *nickname*. In the other cases, these requests were turned into requestable properties: there were 4.4% *item name*, 14.6% *category*, 7.2% *feature*, 2.5% *people activity*, and 0.4% *people's state*. In 78.6% of cases, the system provided direction-giving service, and 21.4% recommendation service.

The appropriateness was evaluated by coders who do not know the study hypothesis. They judged based on the following criteria:

Correct: the information the user requested is included and correct in the response from the robot.

For instance, when a user asked "Are there Japanese restaurants?" the coder judged whether the robot provided the information about any Japanese restaurant (if any), and whether the provided information is correct. There coding results show moderate matching (kappa coefficient was .481).

There were 96.6% of cases judged as correct. Incorrect cases were caused by the lack of nickname (8 cases), users who left before information was provided (3 cases), operator's mistype (3 cases), and complex requests which the system was unable to handle

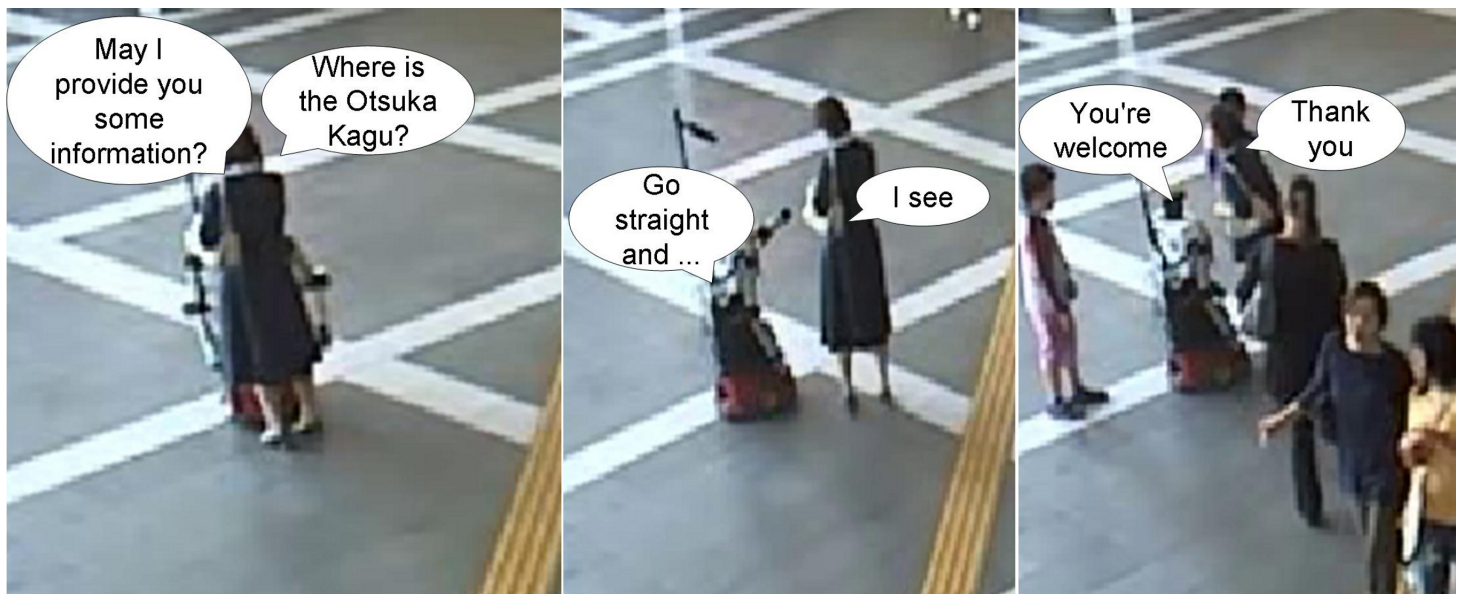


Figure 10 A scene of correct and successful interaction.

(1 case). Overall, we believe that the system was able to cover the requests from users reasonably well.

Figure 10 shows one of example of scene of interaction where the robot provided correct information. She asked a ‘where’ question using the name of a furniture shop, which was matched with the *location entity* instance of the furniture shop. Thus, the robot provided the direction to the shop while pointing the direction. She listened to the direction while looking at the robot. When the robot pointed, she looked at the pointed direction. Finally, she said “Thank you!” to the robot, and walked to the pointed direction.

Figure 11 shows a scene where visitors’ requests were based on their physical state. They only said, “I’m hungry.” The robot was able to associate it to restaurants, so it recommended *ramen* restaurant. They requested it to provide directions to the restaurant, and the robot pointed the direction and explained the route.

Overall, the system worked reasonably well.

LIMITATION

The content of knowledge can be local to the specific environment, robot, language, culture, and so on. The common sense about what the information service is would differ across cultures. Thus, if our study results were to be applied somewhere else, although we believe that most of the framework and structure of knowledge is pertinent, we would probably need to carefully adjust the knowledge. For instance, it is plausible that people in other cultures would inquire information with a different form. Knowledge about interaction would also differ. People in other cultures can be more or less open, active, hesitate, and/or curious, thus the effectiveness of such strategy can be different.



Figure 11 Request made based on visitors' state.

CONCLUSION

We investigated the knowledge relevant to information robot. First, we confirmed that what visitors expect for an information robot well overlapped with what human information staff do. We developed a knowledge representation for information robot. Our field study confirmed the knowledge representation was useful. When users requested, the robot was able to provide information with 96.6% of success. However, it also revealed that many people did not behave in the same way as they did with human staff. Our initial version of interaction flow only allowed 55.4% of success in providing information, while visitors in failure kept silent during the interaction. Through our field experiments, we found that some people need the robot to provide self-introduction about its role, and some people need close-ended prompting, i.e., letting users talk about what they know to make a request, instead of letting them generate a request. Finally, the robot was able to provide information for 84.5% of visitors. What we changed might be subtle, yet it changed the results quite a bit.

ACKNOWLEDGEMENTS

We thank the staff of the ATC shopping mall for their support.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Japan Science and Technology Agency (JST) and its CREST program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Japan Science and Technology Agency (JST).
CREST.

Competing Interests

Takyuki Kanda is an Academic Editor for PeerJ Computer Science. Satoru Satake, Keita Nakatani, Kotaro Hayashi, and Takyuki Kanda are employees of Intelligent Robotics and Communication Laboratories (IRC), Advanced Telecommunications Research Institute International (ATR).

Author Contributions

- Satoru Satake conceived and designed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, performed the computation work.
- Keita Nakatani performed the experiments, analyzed the data, prepared figures and/or tables.
- Kotaro Hayashi performed the experiments, prepared figures and/or tables.
- Takyuki Kanda conceived and designed the experiments, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.
- Michita Imai contributed reagents/materials/analysis tools.

Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The study protocol was approved by institutional review boards of Advanced Telecommunications Research Institute International with reference number 14-502-2.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.8#supplemental-information>.

REFERENCES

- Brcsic D, Kanda T, Ikeda T, Miyashita T. 2013.** Person tracking in large public spaces using 3d range sensors. *IEEE Transaction on Human-Machine Systems* **43**:522–534
DOI [10.1109/THMS.2013.2283945](https://doi.org/10.1109/THMS.2013.2283945).
- Cassell J, Stocky T, Bickmore T, Gao Y, Nakano Y, Ryokai K, Tversky D, Vaucelle C, Vilhjálmsson H. 2002.** Mack: media lab autonomous conversational kiosk. In: *Proceedings of IMAGINA*, vol. 2, 12–15. Available at <http://alumni.media.mit.edu/~tstocky/pubs/Cassell.Stocky-Imagina02.pdf>.
- Daniel M-P, Tom A, Manghi E, Denis M. 2003.** Testing the value of route directions through navigational performance. *Spatial Cognition & Computation* **3**:269–289
DOI [10.1207/s15427633scc0304_2](https://doi.org/10.1207/s15427633scc0304_2).
- Fox D, Burgard W, Thrun S. 1999.** Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research* **11**:391–427.

- Gross H-M, Boehme H, Schroeter C, Mueller S, Koenig A, Einhorn E, Martin C, Merten M, Bley A. 2009.** Toomas: interactive shopping guide robots in everyday use—final implementation and experiences from long-term field trials. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS2009)*. Piscataway: IEEE, 2005–2012.
- Kanda T, Shiomi M, Miyashita Z, Ishiguro H, Hagita N. 2009.** An affective guide robot in a shopping mall. In: *ACM/IEEE international conference on human–robot interaction (HRI2009)*. Piscataway: IEEE, 173–180.
- Kobayashi Y, Shibata T, Hoshi Y, Kuno Y, Okada M, Yamazaki K. 2010.** Choosing answerers by observing gaze responses for museum guide robots. In: *ACM/IEEE international conference on human–robot interaction (HRI2010)*. Piscataway: IEEE, 109–110.
- Kollar T, Tellex S, Roy D, Roy N. 2010.** Toward understanding natural language directions. In: *ACM/IEEE international conference on human–robot interaction (HRI2010)*. Piscataway: IEEE, 259–266.
- Kopp S, Tepper PA, Ferriman K, Striegnitz K, Cassell J. 2008.** Trading spaces: how humans and humanoids use speech and gesture to give directions. In: Nishida T, ed. *Conversational informatics: an engineering approach*. Chichester: John Wiley & Sons, 133–160.
- Matsuda S, Jitsuhiro T, Markov K, Nakamura S. 2006.** Atr parallel decoding based speech recognition system robust to noise and speaking styles. *IEICE Transactions on Information and Systems* **E89-D**:989–997 DOI 10.1093/ietisy/e89-d.3.989.
- McGuinness DL, Van Harmelen F (eds.) 2004.** *Owl web ontology language overview, W3C Recommendation*. Available at <http://www.w3.org/TR/owl-features/>.
- Morales Y, Satake S, Kanda T, Hagita N. 2011.** Modeling environments from a route perspective. In: *ACM/IEEE international conference on human robot interaction (HRI2011)*. Piscataway: IEEE, 441–448.
- Mutlu B, Forlizzi J, Hodgins J. 2006.** A storytelling robot: modeling and evaluation of human-like gaze behavior. In: *IEEE-RAS international conference on humanoid robots (Humanoids'06)*. Piscataway: IEEE, 518–523.
- Okuno Y, Kanda T, Imai M, Ishiguro H, Hagita N. 2009.** Providing route directions: design of robot's utterance, gesture, and timing. In: *ACM/IEEE international conference on human–robot interaction (HRI2009)*. Piscataway: IEEE, 53–60.
- Ono T, Imai M, Ishiguro H. 2001.** A model of embodied communications with gestures between humans and robots. In: *Annual meeting of the cognitive science society (CogSci2001)*. Palo Alto: AAI, 732–737. Available at <http://conferences.inf.ed.ac.uk/cogsci2001/pdf-files/0732.pdf>.
- Rich C, Ponsler B, Holroyd A, Sidner CL. 2010.** Recognizing engagement in human–robot interaction. In: *ACM/IEEE international conference on human–robot interaction (HRI2010)*. Piscataway: IEEE, 375–382.
- 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)*. New York: ACM, 78–84.
- Sidner CL, Lee C, Lesh N. 2003.** Engagement by looking: behaviors for robots when collaborating with people. In: Kruiff-Korbayova, Kosny, eds. *DiaBruck: the proceedings of the seventh workshop on the semantics and pragmatics of dialogue*. Saarbrücken: Saarland University, 123–130.
- Thrun S, Bennewitz M, Burgard W, Cremers AB, Dellaert F, Fox D, Hahnel D, Rosenberg C, Roy N, Schulte J, Schulz D. 1999.** Minerva: a second-generation museum tour-guide robot. In: *IEEE international conference on robotics and automation (ICRA1999)*. Piscataway: IEEE, 1999–2005.