What should we know to develop an information robot?

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

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 knowledge for facilitating interaction is 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 %.

What Should We Know to Develop an Information Robot?

Satoru Satake, Keita Nakatani, Kotaro Hayashi, Takayuki Kanda, Michita Imai

ATR

090-7258-7247

Fax: 099-7258-7247

satoru@atr.jp

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 knowledge for facilitating interaction is 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 %.

Keywords *Information-providing, Direction-giving, Belief about robots*

1 Introduction

Direction-giving is often considered as a desired task for social robots and embodied agents [1-4]. In our daily life, one of the roles that frequently offer direction-giving is information service (Figure 1). Such information booth/counter can be found in stations, airports, shopping malls, and sightseeing places.

Satoru Satake satoru@atr.jp, Keita Nakatani k-nakatani@atr.jp, Kotaro Hayashi,

hayashik@atr.jp, Takayuki Kanda kanda@atr.jp, Michita Imai michita@ailab.ics.keio.ac.jp

31 We wonder what would be the required 'knowledge' to develop a robot that engages in such an

32 information service. Probably, most of us have experience using information service, and many

33 of us believe that we know what the information service is. Thus, one would argue that it is just

34 easy to develop such a robot. One might say that "I know from common sense what the

35 information service is. I can just implement it." Is this true?

37   We started the study with two research questions:

38 ● Is our common knowledge about the tasks of information service (i.e. what they serve)

39   applicable to information robot?

40 ● Can we create an information robot by replicating what human information staff knows

41   and does? Or, is there any missing knowledge?

42   We first investigated what people would expect from an information robot, and confirmed

43 that there are a lot of similarities with what human information staff does (section 3). Thus, we

44 decided to use knowledge about human information staff (what they know and what they do),

45 and developed an information robot (section 4). However, about the second research question,

46 the assumption was not true. Thus, we further investigated missing knowledge (section 5 and 6).

47 ## 2 RELATED WORKS

48 ### 2.1 Information-Providing Robots

49 Robots have been deployed as tour guides. There were a couple of museum robots that navigated

50 around the environment and provided explanations [5]. Robots are also used for interactive

51 information-providing. For instance, Gross et al., developed an article search robot, which

52 enables visitors to request an item and let the robot navigate to its location. [6]. Input for these

53 robots is often with GUIs, thus there are lists of destinations / items, from which users choose

54 one.

55   In contrast, in case of dialog-based system, the difficulty is to predict the set of requests

56 users could ask. Thus, there are assumptions made for the input, like name of locations. For

57 instance, a virtual agent, Mack, developed by Cassell et al., is able to respond the name of

58 locations and people in office, and provide direction-giving [1] [2]. But, in a real-world natural

59 interaction, what users would ask is not bounded by such assumptions. In [7], the robot provided

60   direction-giving interaction in response to the name of locations in a shopping mall, and

61   exceptions were handled by a human operator. That is, the system yet did not by itself address

62   the questions beyond the assumptions.

63   Overall, in the previous studies, it was little explored what people would ask/request in

64   information dialog with a robot. In contrast, we found that people ask various requests beyond

65   name of locations, and identified a required knowledge representation.

2.2 Direction-giving Interaction

66

67   It is reported that a good direction consists of pairs of actions and landmarks [8] , such as "turn

68   right at the post office, and …". To provide such explanations, there is a technique to build a

69   knowledge about spatial relationships among shops and corridors [9]. There are techniques to

70   make a robot understand directions from humans [10]; in their study, the representation stores

71   the relationship between description of the entities in the space and the map.

72   In these studies, the common assumption is that a system is able to provide directions if the

73   name of a location is asked. In contrast, our study reveals other type of requests in information

74   dialog, and we report on the needed knowledge representation.

75   Note that it is well known in HRI studies that gaze and pointing gesture make the

76   interaction more natural and effective (e.g. [11, 12]). The use of gesture in direction-giving is

77   also studied in conversational agent [2] as well as in human-like robots [3, 4]. Our direction-

78   giving behavior is informed by these studies.

2.3 Engagement

79

80   In our study, we noticed some visitors remained silent, even after directly approaching the robot,

81   and hearing its requests to engage. Relevant to this, there were studies about "engagement"

82   process, that is, when people participate and feel connected in collaboration, their gaze well meet

83   with each other and they do not quit the interaction [13]. Rich and his colleagues developed a

84   technique to detect engagement using people gaze [14]. Kobayashi and his colleagues developed

85   a technique to select a person to whom a robot should ask questions in multi-party interaction, in

86   the way a teacher appoints a student in a class for an answer [15]. Their technique is based on the

87   findings that people nodding and engaging in mutual gaze are more likely to answer than

88   someone avoiding meeting gaze. In contrast, in our study, the silent visitors were people who

89 voluntarily approached the robot. They typically behaved as if they were willing to interact with

90 the robot but did not talk with it.

91 **2.4 Knowledge Representation**

92 There are several computer applications e.g. Google search or Apple's Siri that provides

93 information related to locations. There are many similar aspects between our robot and such

94 applications, e.g. both need connection between language and local knowledge, interpretation

95 needs to be contextual, and answers to be provided in verbal way. Thus, similarly to these

96 approaches, we used ontology [16] to build the knowledge representation. However, we need to

97 build our own knowledge representation because required knowledge structure is different, and

98 we cannot simply apply existing software like Google search and Siri for the robot. For instance,

99 robots can use pointing gesture (also, often robots are not equipped with display), which very

100 much change the way of giving direction.

101

102 # 3 INFORMATION IN A SHOPPING MALL

103 We investigated the daily tasks of information service employees and what visitors typically

104 expect from robots acting as such. We found a lot of similarities. The study protocol was

105 approved by institutional review boards of Advanced Telecommunications Research Instituted

106 International with reference number 14-502-2.

107 **3.1 Daily Tasks of Information Service**

108 We interviewed two employees working at the information desk of a shopping mall (Figure 1,

109 left).

110   First we asked an overall description of their job: they usually wait for visitors to come at

111 the information booth. They were requested by the mall administrators to serve as 'information

112 staff'. Only procedures for lost items were provided. For other tasks (e.g. information providing)

113 they use their common sense.

114   Further, we asked them to categorize the typical requests from visitors, and how they would

115 respond. Both reported that there are three types of requests:

116   **Direction giving:** They reported that this is the most frequent request. Visitors ask simple

117   where-type questions, e.g. "where are the toilets?" In addition to the name of locations, people

118   use other popular name, like "hello show", or the name of designated areas, like "smoking area".

119   Their typical response is to provide turn-by-turn directions using utterance and pointing gesture.

120   When visitors do not understand, they sometimes write down to a map, or rarely they take them

121   to the destination.

122   **Recommendation (inquiry):** When a visitor does not know whether there are shops that

123   meet his needs, he may ask it to the information staff. Visitors may inquire characteristics of

124   shops, such as name of items, and category of shops. Here are some examples of questions: "Are

125   there Japanese restaurants?" "Are there shops that sell Osaka souvenirs?" The staff members

126   typically verbally list the shops or events that meet their criteria. Visitors sometimes ask for a

127   recommendation from the staff without providing solid conditions but only with subjective

128   words e.g. "Are there any good restaurants?" For such request, the staff members reported that

129   they typically try not to give a subjective preference, because their preferences may or may not

130   match with visitors'. Thus, their response for inquiry and recommendation are similar: they try to

131   objectively reply and provide a list of shops that seem appropriate.

132   **Lost child and lost-and-found:** When children are lost, or they lost some items, they come

133   to the information desk. For lost child, the staff usually makes a public announcement throughout

134   the shopping mall. Lost items can be retrieved at the information booth when available upon

135   confirmation of the owner.

136   **3.2 Expectations from Information Robot**

137   To investigate what people expect from information robot, we interviewed customers in the

138   shopping mall. To find people who are willing to help us collecting knowledge for future robots,

139   we prepared a situation where visitors can see a robot in interaction. Thus, we prepared a robot

140   for information, which is controlled with wizard-of-oz method. Then, we asked people who

141   stopped around the robot and/or interacted with it to participate to the interview. 21 visitors

142   participated.

143   In the interview, we asked the visitors to imagine future situation where robots will be

144   capable to offer information service as they like, regardless of the robot's capability they

145   observed previously. Then, we asked them to freely provide as many functions they would like

146   information robots to have.

147   The interviews were recorded, and transcribed for analysis. We categorized the different

148   kind of requests expressed by the visitors, For instance, visitors reported sentences such as,

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

150   This utterance was coded as expectation for *direction giving*, because we interpret it as

151   'where'-type question in which visitors simply want to know the location. The followings ones

152   were coded as expectation for *recommendation (inquiry)*:

153   "I'd like to know about sports and furniture shops"

154   "The shop which sells the most? Well, I want the robot give me recommendations of shops"

155   Such cases were classified as *recommendation (inquiry)*, because visitors need to know more

156   information than just a location.

157   Then, two coders who do not know the research purpose judged whether each transcribed

158   sentence would fit into the above defined categories, or not (which is categorized as 'other').

159   Two coders' judgment matches reasonably well yielding kappa coefficient .857.

160

161   Table 1 shows the coding result. The ratio of visitors who mention the expectation is listed

162   in each row. They can provide multiple answers, thus the sum of the ratios exceeds 100%.

163   The visitors' expectation for the information robot largely overlaps with what human

164   information daily do. Almost all visitors (20 out of 21) mentioned that they expect direction

165   giving and the majority (16 out of 20) reported that they expect the robot to offer turn-by-turn

166   direction accompanied with pointing gesture. For instance, one spontaneously mentioned the

167   practicality of pointing gesture in directions giving:

168   "Well, 'where', umm, I did not understand 'which' direction I should go. So it would be

169   useful if the robot could do pointing gestures,"

170   There were 3 visitors who expected the robot to take them to the destination, and 1 visitor who

171   wanted the robot to explain with a map.

172   There were 16 people that expected a recommendation service. For instance, some

173   mentioned "I'd like to have some recommendations for restaurants," or "I'd like to know places

174   where children can play around". Others wanted to have more detailed explanations. For

175   instance, one commented:

176   "I'd like to know what kind of shop it is, its atmosphere, what it sells, and so on."

177   In contrast, the 'playing with children' category, is specific to the information robot. We

178 collected comments such as:

179   "Interacting with the robot was enjoyable. This is good for people who come with their

180   children"

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

182 **3.3 Requirements**

183 People's expectation toward information robots largely overlapped with what is served at human

184 information. That is, most of them expect two services: direction giving and recommendations.

185 Thus, in this study, we focused on these two services.

186   Further we investigated the required knowledge to be stored. We analyzed the utterances of

187 the requests. We labeled them based on the type of request. For instance, we assigned a label

188 'name of location' to the utterance "I'd like to know where the event *dream world* takes place",

189 'name of item' to the utterance "I'd like to know where can I buy *coffee*". If multiple labels are

190 applicable we assigned all of them. Labels are merged when possible, resulting in 6 different

191 labels. To confirm the classification, we asked two coders who do not know the research purpose

192 to classify the utterances based on the 6 labels. Their coding matches reasonably well, yielding

193 kappa coefficient of .637.

194   Finally, we identified that the following information is needed:

195 1) Name of location: such as names of shops or names of events. In addition to the formal

196   name, people use various nicknames. 78.3% of people mentioned this category.

197 2) Item name: people look for specific product or entity available in shops. For instance,

198   this category includes items such as "cell phone charger" and "coffee". 47.8% of people

199   mentioned this category.

200 3) Category: shops can usually be grouped into larger categories, like "restaurant",

201   "Japanese restaurant". 52.2% of people mentioned this category.

202 4) Features: shops are usually recognized as some generally-known features, like "good

203   view", "expensive", and "recommended." 65.2% of people mentioned this category.

204 5) People activity: locations are sometimes referred as the activity that people do there,

205   like "play", "eat", "shop". 60.9% of people mentioned this category.

206   6)   People's state: locations are sometimes referred as the place appropriate for people's

207   physical condition, like "injured", "tired", "hungry". For instance, some visitors

208   mention that,

209   "I would like to receive recommendation, just by saying 'I'm hungry' for

210   example".

211   13.0% of people mentioned this category; note that, such request was not reported by

212   the information desk staff thus it can be considered as specific to the information robot.

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

214 robot.

4 System

4.1 Architecture

217 Our goal is to develop a robot that autonomously provides information service. For this purpose,

218 based on the analysis in section 3, we developed a knowledge representation that can be used by

219 such robot. Figure 2 shows the architecture of the system. Information from sensors goes through

220 modules like *people tracking* (explained in section 4.5.2), *localization* (section 4.5.3), and s*peech*

221 *recognition* (section 4.5.4). Output from these modules are used in the *behavior controller*

222 (section 4.3), which contains a *dialog manager* (section 4.4). The environmental knowledge is

223 stored in *ontology* (section 4.2.2) and *map* (section 4.2.3), and used by the dialog manager. We

224 explain these modules in the later section.

225

4.2 Knowledge representation

227 There are two types of information in the knowledge representation. One is the map used for

228 direction giving (explained in section 4.2.3). The other is shop-related data (explained in section

229 4.2.2).

4.2.1 Environment

231 The study was conducted in a big shopping mall located in a suburban area. It consists of three

232 buildings (Figure 3 left), one having 12 floors, and others having 6 floors. There are 51 shops, 31

233 restaurants, 42 facilities, 6 event halls, 4 squares (e.g. Figure 3 right), 2 stages, and many offices.

234 The mall is mainly busy during weekends. Almost all shops are for non-daily goods, like clothes,

235 shoes, sports, outdoor activities. We often observe people who look for shops and locations (e.g.

236 they look at the floor maps, and/or ask the service staff). The main hall where big events take

237 place is located far (it takes 5 minutes of walk) from the square where we put the robot, thus

238 people often ask where an event is.

239

240

241 *4.2.2 Ontology of entities in the map*

242 We designed our knowledge representation for 'request' and 'shops' together using an ontology

243 language, OWL [16]. Figure 4 shows the designed knowledge structure, i.e. ontology. The basic

244 element in OWL is the 'class', which has 'properties' that store the information. There are two

245 primary classes, 'location entity' and 'requestable property' prepared.

246

247 **Location Entity**

248 We define entities like shops, facilities and events as instances of the 'location entity' class.

249 There are three properties:

250   1) Name: we stored the official or commonly used name.

251   2) Nicknames: some shops are referred to with a nickname. We listed such nicknames

252   people could use. For example, "Kentucky Fried Chicken" is referred as "KFC."

253   3) Location on the geometrical map: each location is associated with the geometrical map

254   (explained in the section 4.2.3).

255 We further separate the class into two subclasses, *selective location* and *non-selective*

256 *location*. When multiple locations are available, people would prefer to select one. For instance,

257 if there are two Italian restaurants, people would choose one based on their own criteria, like

258 better, cheap, popular, etc. We store one extra property, 'introduction property,' in *selective*

259 *location*, to be used in dialog to help people selecting locations. In contrast, people would usually

260 not care about which toilets they would go to. Such locations are implemented as *non-selective*

261 location class.

262 **Requestable property**

263   There are six types of information communicated in information dialog (section 3.3).

264   Except for name of location, they are realized as 'requestable property' class, which has

265   subclasses 'item name', 'category', 'features', 'people activity', and 'people's state'. When a

266   user requests some information, it is turned into an instance of the 'requestable property'. Then,

267   the location(s) having the same property will be searched. Each property item has wordings that

268   are expected to be used in people's utterance. For instance, 'eat' (instance of people's activity

269   subclass) is associated with wordings such as "eat", "have lunch", and "have a meal". Note that

270   more complex requests (e.g. "Japanese" restaurant with a "good view") can be represented as

271   multiple instances combined with 'and/or' operators, but we did not implement such complex

272   operations because users rarely made such complex requests.

273   **Relationships between 'location entity' and 'requestable property'**

274   Table 2 shows possible relationships between two subclasses. For instance, some visitors

275   could request a restaurant where they can have "pasta". To handle such requests, a "pasta" entity

276   is prepared as an instance of 'item name' subclass which is associated with shops with the

277   relation 'is served at". Such relation is defined inside dialog management (section 4.3) as well.

278   Note that an instance of 'requestable property' can be associated with multiple 'location entities'

279   (e.g. "pasta" can be served at multiple restaurants).

280

281   Finally, we prepared the data for the shopping mall (section 4.2.1). There are 201 location

282   entities (84 shops, 75 service facilities, 39 events, and 3 buildings) with in total 3345 nicknames.

283   There are 530 requestable properties (501 items, 163 categories, 44 features, 63 people activities,

284   22 people's states) prepared as well.

285   *4.2.3 Route Perspective Map*

286   Informed by [9], we manually prepared a route perspective map (illustrated in fig. 5), which

287   consists of pairs of landmarks and actions. Using the map, the system generates turn-by-turn

288   directions giving, such as "go straight, turn left at the book store, go out the door with exit sign,

289   …" The map includes the following information:

290   1)   **Topological map**: Nodes are located at decision points in the map. Transition through

291   corridor or between different floors, such as stairs, escalators, and elevators are

292   expressed as movements between nodes. Entrances of shops, facilities, and events (i.e.

293   *location entity*) are also represented as nodes.

294   2)   **Landmarks**: If available, visible landmarks are manually associated for each route as

295   denoted in [9], e.g. famous shop names with salient signboards, elevators, and

296   escalators.

297   3)   **Actions:** In [9], actions were only turning behaviors, which were computed from a

298   topological map. In contrast, as there are many floors and multiple buildings, we added

299   actions like "enter the next building", "go to the 3rd floor".

300

4.3 Behavior Controller

302   When a person stops by the robot (within 2.5 m for 3.0 seconds), or is detected as approaching at

303   2.5 m from it, it starts a dialog. The robot orients its body and gaze to the user. When there is no

304   user, the robot shows liveliness by slightly moving head and arms.

305   During the dialog, its head and body is oriented toward the user, except for the moment

306   when it performs a pointing gesture which is often used when giving directions. When it points

307   at a direction, its head direction is oriented toward the pointed direction for the first three

308   seconds of pointing in order to draw the user's attention toward the pointed direction. The robot

309   ends the dialog when the user leaves the robot's side (3 m away), or when the dialog

310   management module decides to end the dialog.

4.4 Dialog Manager

312   We developed a rule-based mechanism for dialog management. Assuming that there is an input

313   coming from the speech recognition module (explained in 4.5.4), the input is turned into text and

314   matched with name/nickname properties of location entities and with instances of *requestable*

315   *properties* (explained in 4.2.2). If a requestable property matched, it is compared with *location*

316   *entities*.

317   When only non-selective locations are matched, it chooses the nearest one. In case the user

318   asked for a location with a specific *name of location*, there should be only one location to be

319   matched. In these cases, the system provides *direction-giving* dialog, in which turn-by-turn

320   directions to the location are generated.

321   Otherwise, it initiates a *recommendation* dialog. It verbally lists the locations that match

322 with the *requestable property* instance one by one. For each location, it explains the location

323 using the text in its introduction property. For instance, it utters "*Ramen* is served at a ramen

324 restaurant named *Kaika-ya*. They serve a ramen with tuna soup. May I explain the directions to

325 go there?" As human staff does, we carefully avoid telling subjective preferences, but only

326 provided objective facts.

327   In addition, it reacts to the words for greeting. When an input matches with words like

328 "hello", it returns a greeting utterance. When an input matches with leave-taking words like

329 "bye", it returns leave-taking words and ends the dialog.

330   When no location is matched, the system explains that "(requested item) is not in this

331 shopping mall. I only know about this mall".

332 **4.5 Other Modules**

333 *4.5.1 Robot*

334 We used a robot characterized by its human-like physical expressions. It is 120 cm high and 40

335 cm in diameter on a mobile platform. It has a 3-DOF head and 4-DOF arms. There are two 30 m

336 range laser sensors attached. We used the robot with a maximum speed of 550 mm/sec and 50

337 degree/sec for rotations. The accelerations are set to 300 mm/sec2 and 50 degree/sec2. To clearly

338 communicate its role, we put an 'information staff' sign in Japanese on the chest of the robot

339 (Figure 1, right).

340 *4.5.2 People Tracking*

341 We use a people tracking method described in [17], which provides an estimation of pedestrians'

342 locations every 33 ms. It covers the square we used. There are 49 3-D range sensors attached on

343 the ceiling (combination of Panasonic D-Imager, ASUS Xtion, and Velodyne HDL-32E).

344 *4.5.3 Localization*

345 For robot localization, we use a particle filter with a ray tracing approach on a grid map [18]. The

346 grid map is built from odometry and laser scanner data. This module is called every 350 msec

347 and updates the robot's position.

348 *4.5.4 Speech Recognition (with Human Operator)*

349 We developed fully-autonomous system using ASR (automatic speech recognition), but finally

350 to better test the overall framework we used a human operator instead of ASR.

351 *4.5.4.1 Automatic Speech Recognition (ASR)*

352 We used an ASR software, ATRASR [19]. It uses a language model based on FSA (Finite State

353 Automaton). We constructed the language model mainly using the terms appeared in the

354 ontology.

355   With preliminary trials using Wizard-of-oz approach, we analyzed the way visitors speak to

356 the robot. In total, 470 requests collected over 3 days of preliminary trials. From the analysis of

357 the requests, we found that they mainly follow three ways of speaking, as follows:

358 ● Noun/adjectives only:

359   People only spoke words like a *name (nickname) of location, category,* or *item name,* such

360   as "restaurant," and "coffee". Sometimes, for *features* and *people's activity*, they add such

361   terms like "place for" (eat/lunch/play). Some ontology items are adjectives, such as "tired".

362   People sometimes only spoke such adjectives.

363 ● "Where is" question:

364   The above noun is used in "where is" question, such as "where is *Kaika-ya (*the name of

365   restaurant*) ?"

366 ● "I would like to" sentence:

367   People also use the form of "I would like to" + "verb" + "noun" in requesting sentences,

368   such as "I would like to buy coffee".

369 For all *names, nicknames*, and *requestable properties,* we automatically generated grammatical

370 structures for ASR. Further, we added the following grammars. First, some basic verbs like "go"

371 can be used in "I would like to" type sentences but were not included in the ontology (as they by

372 themselves does not represent any specific request), which we manually added (8 verbs). Second,

373 we added filler words, such as "well", "ah", that appear in advance to questions (12 words).

374 Third, to eliminate noises from environments, like sounds from people's walking, whistle from

375 ships, we added some fillers (66 fillers). Overall, we prepared the lexicon whose size is 1469

376 with 4938 links.

377   The ASR outputs the matched *names, nicknames*, or *requestable properties*, which are used in

378   the dialog manager to determine the answer to be provided. In case the ASR detects the

379   recognition to be less reliable (because the input does not match well with its language model),

380   the dialog manager prompts the user to say again with utterances like "could you repeat please?"

381   The ASR is deactivated while the robot is speaking.

382   We evaluated the system performance using this ASR implementation. We put the robot on a

383   square of the mall (Fig. 3 right), and let the visitors freely use it. With our preliminary test, with

384   22 users, there are 81 requests, for which the robot was only able to correctly respond in 19.8%

385   of the cases. (In a similar study only 21.3% of successful recognition was achieved [19]).

386   There were 4 types of errors: error in sound detection error (due to other ambient sounds, the

387   system failed to detect the start of utterance) (17.3%), ASR resulted in low reliability score

388   (30.9%), utterance did not match with the prepared grammar/vocabulary (2.47%), and mis-

389   recognition in ASR (29.6%). In case the mis-recognition occurred, often the system seemed to be

390   interfered by ambient noise, which was matched with some vocabulary in the lexicon.

391   In contrast, in case ASR successfully detected the *names, nicknames*, or *requestable*

392   *properties*, the system provided appropriate answers. Overall, this preliminary test revealed that

393   the system is capable of handling users' utterances when the ASR is successful, while we would

394   yet need to wait ASR technologies to be ready for real world environments.

395

396 *4.5.4.2 Wizard-of-Oz*

397 The system is ready for autonomous speech recognition. But, for this study, to focus on other

398 parts of interaction rather than working for errors in speech recognition, we used a human

399 operator to only support speech recognition.

400   We strictly limit the task of the operator, and have him work like dumb ASR software

401 described in the previous section. We did not allow the operator to add his knowledge. Like the

402 output from the ASR, the operator only typed the words included in what the user said. For

403 instance, to our knowledge, if a user asks for a "Place for lunch" but such wording is not in the

404 system vocabulary, in previous studies Wizard-of-oz operators replaced such words to the ones

405 system can handle, like "restaurant"; by doing so, system can work with very limited vocabulary

406 and knowledge. Instead, with our system, a novice person who does not know the environment
407 (e.g. list of shops) can easily serve as an operator.

5 Preliminary trials: Lack of 'Knowledge' for Interaction

408

409 We conducted a preliminary study with the system reported in the previous section. We initially
410 intended to supplement missing data and evaluate its performance. We found the system itself
411 worked well (we will report in section 6.3); however, interaction failed in other parts we did not
412 think about. That is, some visitors responded in an unexpected way. In short, until here, we
413 focused on the 'information' aspect, which we found to be satisfyingly prepared, but we found a
414 problem in 'interaction'.

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

419

420 **Case 1: Interaction did not start**

421   The initial version of the robot imitated the interaction of human information staff. It waited
422 for the arrival of the visitors, and waited for them to make a request. This is what a human staff
423 member does. The signboard showing 'information staff' on the chest of the robot was very
424 visible, so we expected that every visitor would have common expectations as those investigated
425 in section 3.2.

426   However, frequently there are people who stay in front of the robot without saying
427 anything. Figure 6 shows one of such cases. A man stopped in front of the robot, and the robot
428 was ready to receive a request, orienting its body and head toward him; but, without talking to it,
429 he moved to a side of the robot, and the robot followed. He moved back, and it followed again.
430 Finally, he left after 30 seconds of silence.

431

432 **Case 2: Passive visitors**

433   Further, we noticed that the conversation got stuck when it asked for a request, even though
434 the user initially spoke to the robot. For instance, Figure 7 shows a visitor who engaged in

435 greeting, but came to be silent when prompted to ask request. She left after 5 seconds of silence

436 after being prompted.

437   We interpreted that such people do not have concrete requests in their mind, thus they were

438 stuck when asked to offer requests.

439

6 Field Experiment

441 For each case of problems found in the preliminary trial (reported in the previous section), we

442 generated a hypothesis, and conducted an experiment to confirm our idea to supplement such

443 weakness. The study protocol was approved by institutional review boards of Advanced

444 Telecommunications Research Instituted International with reference number 14-502-2.

6.1 Experiment 1

6.1.1 Hypothesis

447 We initially replicated the way of interaction of human staff. That is, we make it clear that the

448 robot is serving as information staff. Assuming that visitors would have the common

449 expectations of what an information staff is, we let the robot wait for a visitor to make a request,

450 and to prompt to request if not asked. However, this assumption can be not always true. Visitors

451 may not share or are unsure about their expectations of the 'information robot' role. If this is the

452 case, we can probably moderate the problem by letting the robot first explain its role (direction

453 giving and recommendation). Thus, we made the following prediction:

454   **Prediction 1:** If the robot proactively explains its role as information staff, people will more

455   frequently request information from it.

6.1.2 Participants

457 The study was conducted during weekends. The participants were visitors of the shopping mall,

458 who are typically group of friends and families who come to the mall for leisure. The mall is big,

459 and the layout is complicated, thus people are often in real needs of getting directions from

460 someone.

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

463 *6.1.3 Condition*

464 There are two conditions compared.

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

470   In both conditions, when a visitor requests it immediately moves into the information
471 dialog. After 20 seconds of silence, the robot closed the interaction saying "bye-bye".

472 *6.1.4 Procedure*

473 The robot was placed at a square of the mall (Figure 4 right). We choose this location because
474 visitors often arrive from the nearby escalator, and need direction giving around this location.
475 The study was conducted during daytime on weekends. We prepared six pairs of 25-minutes
476 time slots. For each pair, two conditions were assigned. Between the slots, we put 5-minutes
477 break, so that visitors are not influenced by the adjacent time slot.

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

483 *6.1.5 Measurement*

484 Considering the role of the information staff, we define the success of the interaction as follows:
485 Success: The case where the robot was able to receive a request and offered appropriate
486   information/service.
487   We coded the success from the recorded video. Note that we only evaluated people who
488 stopped in front of the robot (more than 3 seconds) and faced toward the robot; we consider that

489 letting people stop is beyond the scope of this paper. If the same person interacted multiple

490 times, only the first one was evaluated. Further, we only evaluated one participant per group (i.e.

491 only the first member of the group, who stopped and faced the robot, was counted as our

492 participant), so that the experiment would not suffer from other members' prior interaction.

6.1.6 Result

494 In total, there were 238 interactions evaluated, which were coded by two coders who do not

495 know the study hypothesis. One coded the whole data and second one did confirmatory coding

496 for 10% of the data. Their coding results matches well (kappa coefficient .962).

497   Figure 8 shows the result of the study. There were 69.0% of the successful interactions in

498 the *with self-introduction* condition, while 54.4% in the *without self-introduction condition*.

499 Typical failure was, like the one shown in figure 6, where visitors stayed in front of the robot but

500 remained silent even if they were prompted to talk to the robot. Some visitors left in the middle

501 of the conversation, and some explicitly said they did not need service (6 cases in *with self-*

502 *introduction condition*).

503

504   We applied a Chi-square test to evaluate the ratio of success against failures. There is a

505 significant difference between the conditions ($\chi2(1)=4.755$, $p<.05$, $\varphi c=.141$).

506   Thus, the prediction 1 was confirmed. When the robot provides self-introduction, the

507 interactions ended with success more frequently. We interpret that even though the robot serves

508 an 'information' role, people should share a common expectation. Unless it explains its role,

509 some people might fail in using it.

6.1.7 Discussion

511 It is plausible that there are two sources of failure addressed. One is the belief that the robot can

512 talk to them; another is the expectation that it offers information. We mainly argued the second

513 point, but it simultaneously offered help for the first element. Thus, one would argue that it is

514 better to compare with a robot that only speaks to users but does not provide self-introduction.

515   However, it was not easy to prepare such a condition, where the robot only shows the

516 capability that it can talk in the context of information service. For instance, if it only greeted

517 people, visitors might expect it to engage in variety of interactions, but in reality the robot can

518 only react for the 'information' role. Thus, although the effect would be due to both elements, we

519 conducted the study in such a way. It remains as an open question what is the best length of self-

520 introduction. We could make it short and only imply its task by saying something like "may I

521 help you?" We consider that to our observation, people did not get bored due to length of the

522 self-introduction and thus it could be considered as reasonable.

523 **6.2 Experiment 2**

524 *6.2.1 Hypothesis*

525 In the experiment 1, we found that self-introduction moderated the problem of failure; yet,

526 interaction failed for about 30% of the visitors. We hypothesized that there are visitors who

527 initiated interaction out of curiosity, without a concrete request in mind. Such people would be

528 stuck when a robot prompts them for a request in a direct way. We hypothesized that we can

529 moderate this problem, if the robot turns its offer into a question that they can easily answer.

530 Thus, we made the following prediction:

531 **Prediction 2:** If the robot prompts a user for a request in a way of questions they can easily

532   answer, people will more frequently make requests to the information robot.

533 *6.2.2 Participants*

534 The same procedure was used as in experiment 1.

535 *6.2.3 Condition*

536 There are two conditions compared. In both conditions, when a person stops, the robot starts with

537 a self-introduction, saying "Hello, I can provide directions and recommendations." This is

538 identical to the wording used in experiment 1. After a short pause, the robot utters "I will give

539 recommendations based on the locations you are going to", and prompts the user to ask. The

540 prompting utterance differs depending on the following condition:

541 - ***Open-ended prompting:*** It prompts the user saying "*What kind of recommendation do you*

542   *wish*?"

543 - ***Close-ended prompting:*** It prompts the user saying "*Where are you going*?"

544   In both conditions, whenever a visitor requests something to the robot, it immediately

545 moves into the information dialog. If the user keeps silent for 8 seconds, it once repeats the

546 prompting utterance. If there were 20 seconds of silence after the prompting utterance, the robot

547 closed the interaction saying "bye-bye".

548 *6.2.4 Procedure*

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

550 slots.

551 *6.2.5 Measurement*

552 The same measurement was used as in experiment 1.

553 *6.2.6 Result*

554 In total, there were 205 interactions evaluated, which were coded by two coders who do not

555 know the study hypothesis. One coded the all data and second one did confirmatory coding for

556 10% of the data. Their coding results matches well (kappa coefficient .936).

557   Figure 9 shows the result of the study. There were 84.5% of successful interactions in *close-*

558 *ended prompting* condition and 69.4% in *open-ended prompting* condition. Similar to the

559 experiment 1, in failure cases, some visitors kept silent when prompted, some visitors left in the

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

561 condition).

562   We applied a Chi-square test to evaluate the ratio of success against failures. There is a

563 significant difference between the conditions ($\chi^2(1)=5.678$, $p<.05$, $\varphi_c=.166$).

564   The prediction 2 was consequently confirmed. When the robot's prompting was close-

565 ended, the interaction was more frequently successful than open-ended prompting. We interpret

566 that as predicted many visitors did not have requests in mind and got stuck when asked to

567 request; instead, if the robot offered a prompting utterance that invited the user to talk about what

568 they know (e.g. their destination), it will more easily continue the dialog and offer information

569 requested by the user.

570

571 *6.2.6 Discussion*

572 There are some open questions remaining. One would argue that those who kept silent are people

573 who did not want to 'hear' the information, thus they did not respond to 'hear' questions in

574 close-ended prompting. It is possible that they did not have that much will to spontaneously ask

575 the robot to provide information; nevertheless, in *open-ended prompting* condition, people who

576 were coded as success stayed until the robot finished providing information. One would also

577 argue that the robot could anyway give information even if visitors kept silent. This is possible,

578 and maybe the robot should do so for the remaining 15.5% of people. Our assumption is that it is

579 probably better if they hear information they requested, rather than randomly chosen

580 information. We could not fully clarify why the remaining 15.5% of people who kept quiet in

581 *close-ended* condition. We tried to interview such people, but they did not want to be

582 interviewed.

583 **6.3 Evaluation of System Performance**

584 Throughout the experiment 1 and 2, the robot was controlled with the system reported in section

585 4. In total, there were 435 requests made for the information robot. We analyzed how they were

586 handled, and evaluated whether the robot's responses were correct.

587   66.8% of the case requests were a *name of location* and 4.0% were a *nickname.* In the other

588 cases, these requests were turned into requestable properties: there were 4.4% *item name*, 14.6%

589 *category*, 7.2% *feature,* 2.5% *people activity,* and 0.4% *people's state.* In 78.6% of cases, the

590 system provided direction-giving service, and 21.4% recommendation service.

591   The appropriateness was evaluated by coders who do not know the study hypothesis. They

592 judged based on the following criteria:

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

594   the robot.

595

596   For instance, when a user asked "are there Japanese restaurants?" the coder judged whether

597 the robot provided the information about any Japanese restaurant (if any), and whether the

598 provided information is correct. There coding results show moderate matching (kappa coefficient

599 was .481).

600   There were 96.6% of cases judged as correct. Incorrect cases were caused by the lack of

601   nickname (8 cases), users who left before information was provided (3 cases), operator's mistype

602   (3 cases), and complex requests which the system was unable to handle (1 case). Overall, we

603   believe that the system was able to cover the requests from users reasonably well.

604   Figure 10 shows one of example of scene of interaction where the robot provided correct

605   information. She asked a 'where' question using the name of a furniture shop, which was

606   matched with the *location entity* instance of the furniture shop. Thus, the robot provided the

607   direction to the shop while pointing the direction. She listened to the direction while looking at

608   the robot. When the robot pointed, she looked at the pointed direction. Finally, she said "thank

609   you!" to the robot, and walked to the pointed direction.

610

611   Figure 11 shows a scene where visitors' requests were based on their physical state. They

612   only said, "I'm hungry". The robot was able to associate it to restaurants, so it recommended

613   *ramen* restaurant. They requested it to provide directions to the restaurant, and the robot pointed

614   the direction end explained the route.

615   Overall, the system worked reasonably well.

616

617 7 Limitation

618   The content of knowledge can be local to the specific environment, robot, language, culture, and

619   so on. The common sense about what the information service is would differ across cultures.

620   Thus, if our study results were to be applied somewhere else, although we believe that most of

621   the framework and structure of knowledge is pertinent, we would probably need to carefully

622   adjust the knowledge. For instance, it is plausible that people in other cultures would inquire

623   information with a different form. Knowledge about interaction would also differ. People in

624   other cultures can be more or less open, active, hesitate, and/or curious, thus the effectiveness of

625   such strategy can be different.

8 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 the way they interact 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 a lot.

9 ACKNOWLEDGMENTS

10 REFERENCES

[1]   J. Cassell, T. Stocky, T. Bickmore, Y. Gao, Y. Nakano, K. Ryokai, D. Tversky, C. Vaucelle, and H. Vilh jálmsson, Mack: Media Lab Autonomous Conversational Kiosk, in *Proc. of Imagina* vol. 2, ed, 2002, pp. 12-15.
[2]   S. Kopp, P. A. Tepper, K. Ferriman, K. Striegnitz and J. Cassell, Trading Spaces: How Humans and Humanoids Use Speech and Gesture to Give Directions, in *Conversational Informatics: An Engineering Approach*,   T. Nishida ed., pp. 133-160,   2008.
[3]   T. Ono, M. Imai and H. Ishiguro, A Model of Embodied Communications with Gestures between Humans and Robots,   *Annual Meeting of the Cognitive Science Society (CogSci2001)*,   pp. 732-737, 2001.
[4]   Y. Okuno, T. Kanda, M. Imai, H. Ishiguro and N. Hagita, Providing Route Directions: Design of Robot's Utterance, Gesture, and Timing,   *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2009)*,   pp. 53-60, 2009.
[5]   S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, Minerva: A Second-Generation Museum Tour-Guide Robot,   *IEEE Int. Conf. on Robotics and Automation (ICRA1999)*,   pp. 1999-2005, 1999.

658 [6]   H.-M. Gross, H. Boehme, C. Schroeter, S. Mueller, A. Koenig, E. Einhorn, C. Martin, M.
659   Merten, and A. Bley, Toomas: Interactive Shopping Guide Robots in Everyday Use -
660   Final Implementation and Experiences from Long-Term Field Trials, *IEEE/RSJ Int.*
661   *Conf. on Intelligent Robots and Systems (IROS2009)*, pp. 2005-2012, 2009.
662 [7]   T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro and N. Hagita, An Affective Guide
663   Robot in a Shopping Mall, *ACM/IEEE Int. Conf. on Human-Robot Interaction*
664   *(HRI2009)*, pp. 173-180, 2009.
665 [8]   M.-P. Daniel, A. Tom, E. Manghi and M. Denis, Testing the Value of Route Directions
666   through Navigational Performance, *Spatial Cognition & Computation*, vol. 3, pp. 269-
667   289, 2003.
668 [9]   Y. Morales, S. Satake, T. Kanda and N. Hagita, Modeling Environments from a Route
669   Perspective, *ACM/IEEE Int. Conf. on Human Robot Interaction (HRI2011)*, pp. 441-
670   448, 2011.
671 [10] T. Kollar, S. Tellex, D. Roy and N. Roy, Toward Understanding Natural Language
672   Directions, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp. 259-
673   266, 2010.
674 [11] C. L. Sidner, C. D. Kidd, C. Lee and N. Lesh, Where to Look: A Study of Human-Robot
675   Engagement, *Int. Conf. on Intelligent User Interfaces (IUI 2004)*, pp. 78-84, 2004.
676 [12] B. Mutlu, J. Forlizzi and J. Hodgins, A Storytelling Robot: Modeling and Evaluation of
677   Human-Like Gaze Behavior, *IEEE-RAS Int. Conf. on Humanoid Robots*
678   *(Humanoids'06)*, pp. 518-523, 2006.
679 [13] C. L. Sidner, C. Lee and N. Lesh, Engagement by Looking: Behaviors for Robots When
680   Collaborating with People, ed, 2003.
681 [14] C. Rich, B. Ponsler, A. Holroyd and C. L. Sidner, Recognizing Engagement in Human-
682   Robot Interaction, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp.
683   375-382, 2010.
684 [15] Y. Kobayashi, et al., Choosing Answerers by Observing Gaze Responses for Museum
685   Guide Robots, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp.
686   109-110, 2010.
687 [16] D. L. McGuinness and F. Van Harmelen, Owl Web Ontology Language Overview, *W3C*
688   *recommendation*, 2004.
689 [17] D. Brscic, T. Kanda, T. Ikeda and T. Miyashita, Person Tracking in Large Public Spaces
690   Using 3d Range Sensors, *IEEE Transaction on Human-Machine Systems*, vol. 43, pp.
691   522 - 534, 2013.
692 [18] D. Fox, W. Burgard and S. Thrun, Markov Localization for Mobile Robots in Dynamic
693   Environments, *Journal of Artificial Intelligence Research*, vol. 11, pp. 391-427, 1999.
694 [19] S. Matsuda, T. Jitsuhiro, K. Markov and S. Nakamura, Atr Parallel Decoding Based
695   Speech Recognition System Robust to Noise and Speaking Styles, *IEICE*
696   *TRANSACTIONS on Information and Systems*, vol. E89-D, pp. 989-997, 2006.
697
698

The analysis result of expectation for information

2   **Table 1** The analysis result of expectation for information

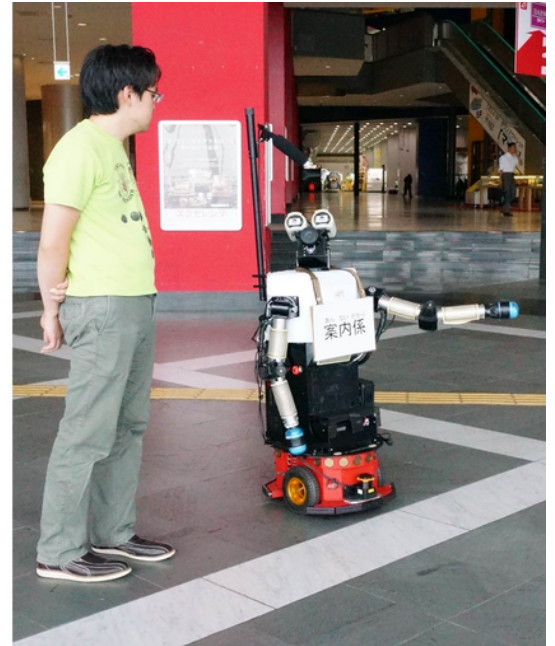
Expectation	Ratio
Direction giving	95.2%
Recommendation (inquiry)	76.2%
Other	
Playing with children	23.8%
Lost child	4.8%

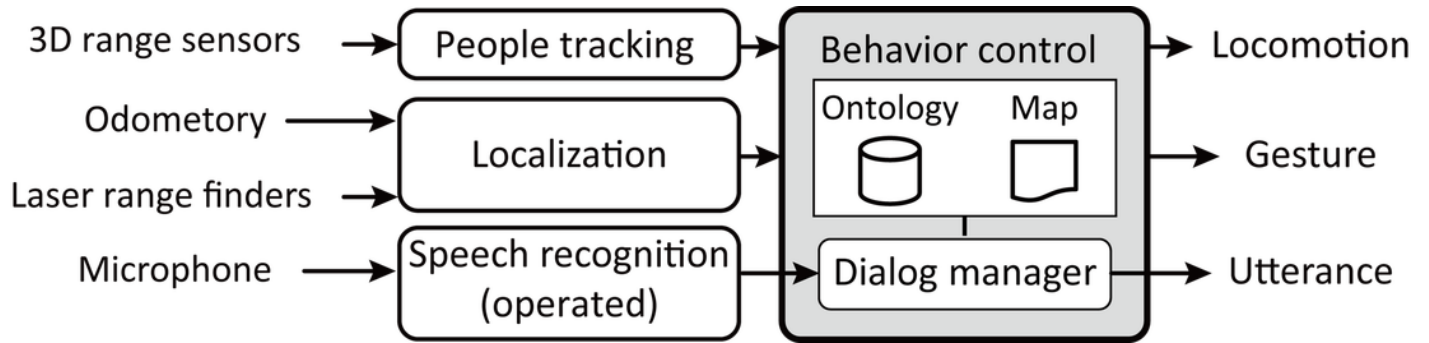
Possible relationships

2   **Table 2** Possible relationships

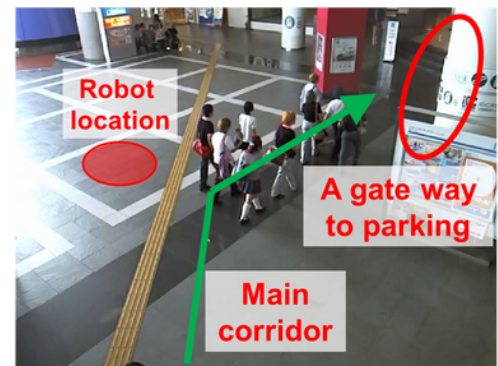
Users' request	Possible relation
Item name	is sold at / is served at / is at
Category	belongs to
Features	is a feature of
People activity	is possible at
People's state	is satisfied/healed/solved at

1

Information service

2

System architecture

3

Environment of the shopping mall

4

Knowledge representation for the environment

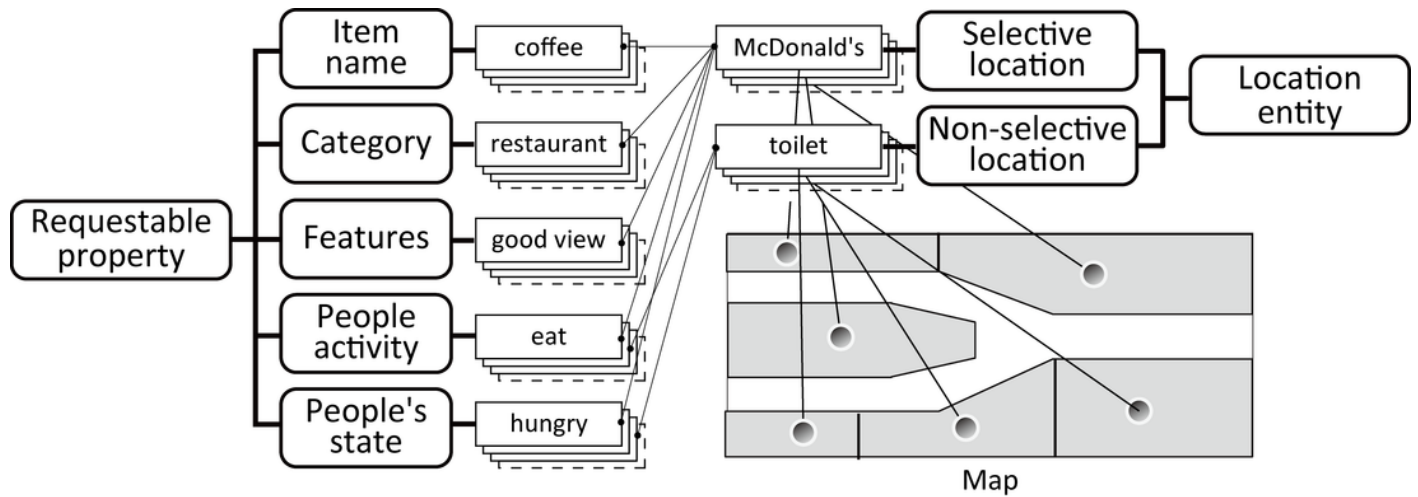
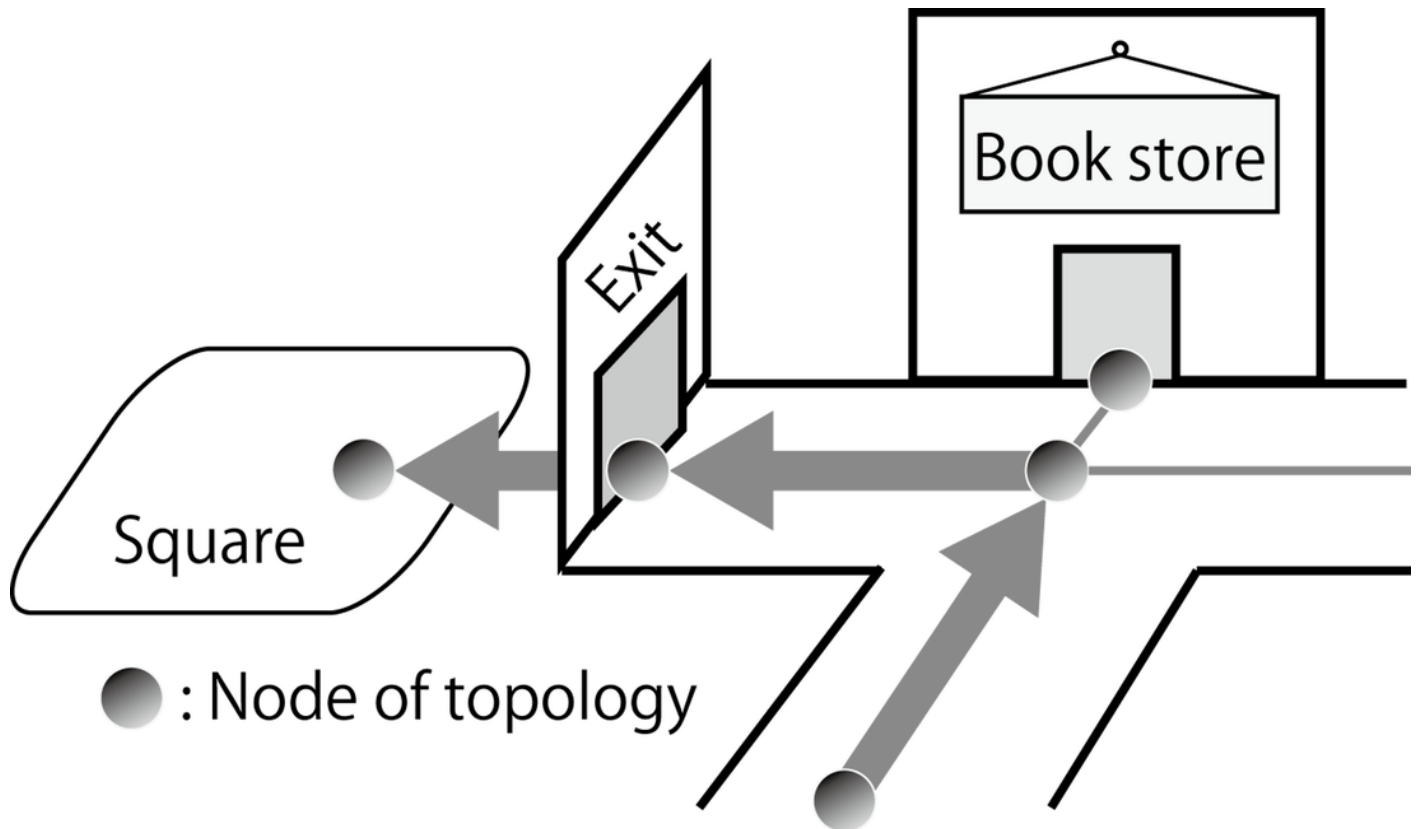



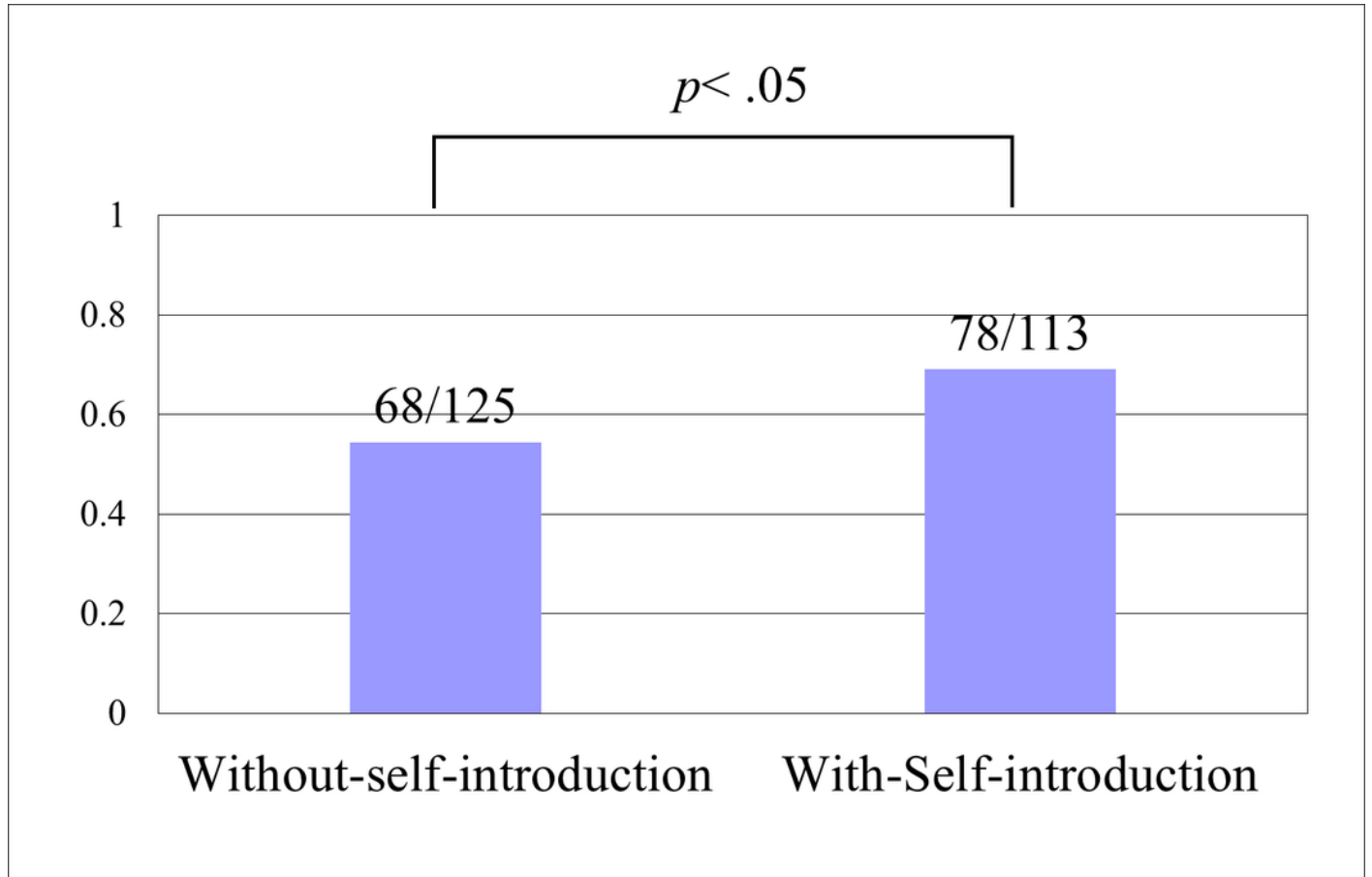
Illustration of the route perspective map

6

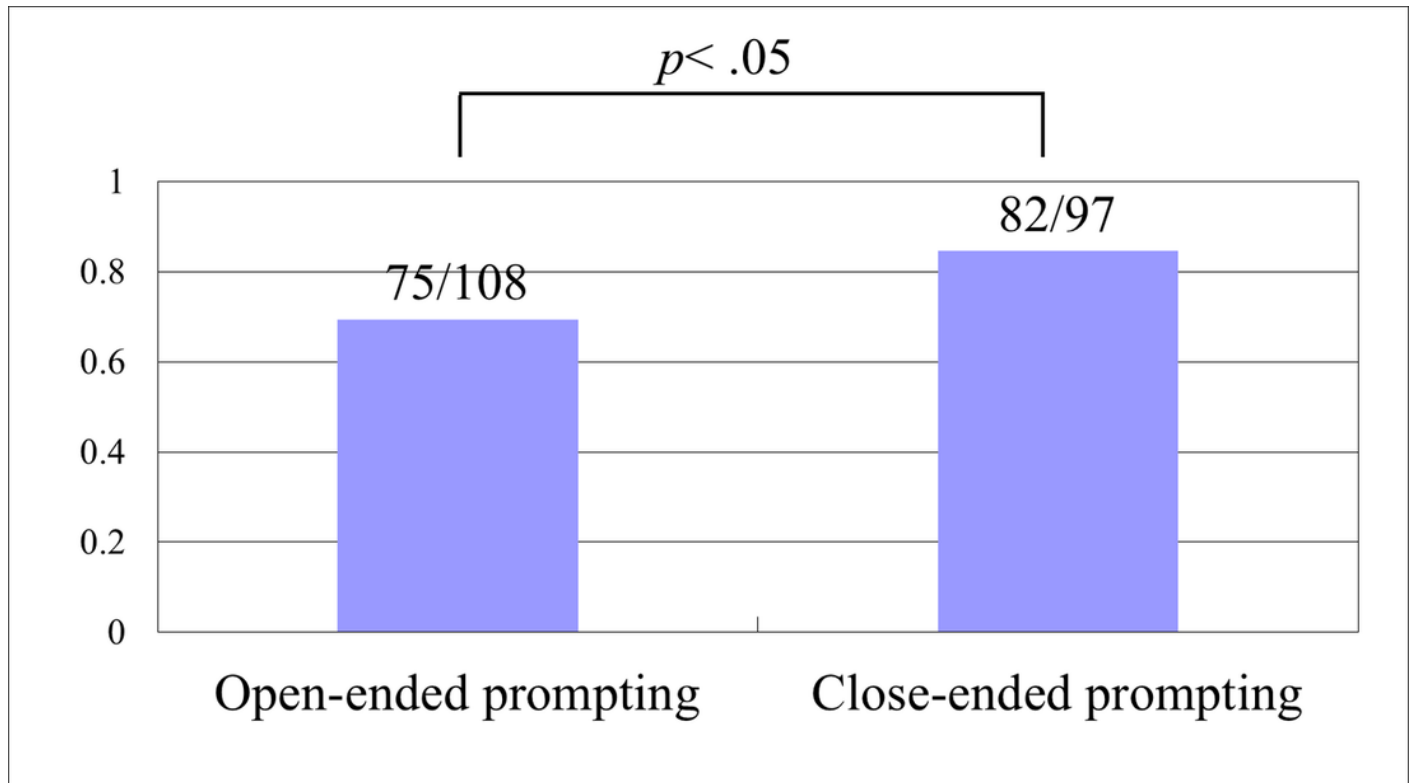
The interaction failed because the visitor did not speak

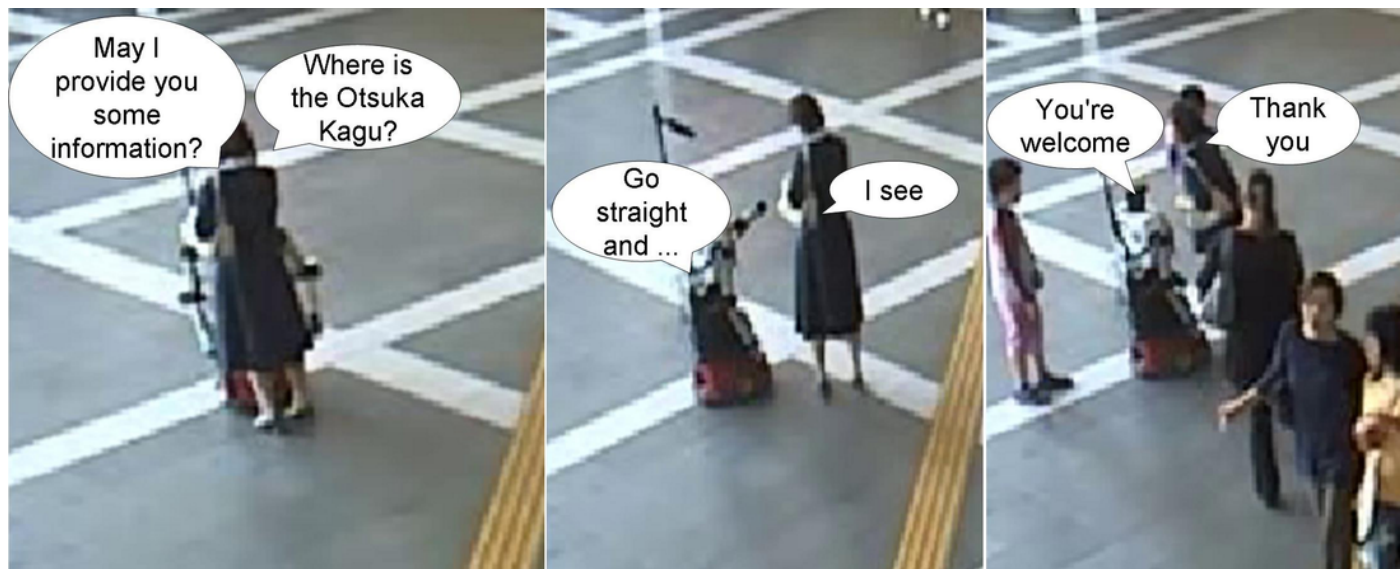
7

The visitor kept silent after prompted

8

Result of the experiment 1

9

Result of the experiment 2

10

A scene of correct and successful interaction

11

Request made based on visitors' state