Social robots that naturally initiate interaction with people

CHAO SHI

MARCH 2017

Social robots that naturally initiate interaction with people

A dissertation submitted to

THE GRADUATE SCHOOL OF ENGINEERING SCIENCE

OSAKA UNIVERSITY

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ENGINEERING

BY

CHAO SHI

MARCH 2017

ABSTRACT

This thesis explores the problem of natural initiation of interaction in human-robot interaction focused on the situation that robot encounter people. When robot encounter people, robot could initiate interaction by conversation or motion.

Initiation of conversation might seem trivial for people, but it is not at all trivial for robots. The appropriate timing and good position from which to make the initial greeting are almost unconsciously decided by human. Based on analysis of human interactions, this paper proposes a model for a natural way of initiating conversation. The model mainly involves the participation state and spatial formation. When a person prepares to participate in a conversation and a particular spatial formation occurs, he/she feels that he/she is participating in the conversation; once he/she perceives his/her participation, he/she maintains particular spatial formations. We proposes a participation state model for measuring communication participation and provided a clear set of guidelines for how to structure a robot's behavior to start and maintain a conversation based on the model. Our model precisely describes the constraints and expected behaviors for the phase of initiating conversation. We implemented our proposed model in a humanoid robot and conducted both a system evaluation and a user evaluation in a shop scenario experiment. It was shown that good recognition accuracy of interaction state in a conversation was achieved with our proposed model, and the robot implemented with our proposed model was evaluated as best in terms of appropriateness of behaviors and interaction efficiency.

Abstract

For initiation of interaction by motion, there is no need to go further into conversation. Instead, it is important to express the intention of interaction to the partner. In our daily life, it is very common to see people distribute objects such as flyers to pedestrians. The givers initiate interaction with passersby by their handing motions. It would be appropriate to assume that in the future these "distributing" works would be carried out by robots. We proposed a model for a robot distributing flyers to pedestrians. The difficulty is that potential receivers are pedestrians who are not necessarily cooperative; thus, the robot needs to appropriately plan its motion, making it is easy and non-obstructive for potential receivers to receive the flyers. We observed human interactions on distributional handing in the real world, analyzed and evaluated different handing methods that people perform, and established a model for a robot to perform natural handing. The proposed model is implemented into a humanoid robot and is confirmed as effective in a field experiment.

Finally, we conducted a field study to investigate the expected use of such robot that initiate interaction with people by conversation or motion in the real world, particularly for attracting passersby which today's robots can autonomously perform with our proposed models. From interviews with ten store managers, we identified two main reasons they want to employ such social robots in their stores: robots offer cheap labor and provide unique value that humans cannot. They believe that robots are good at attracting the attention of visitors without causing or receiving stress. We also conducted three case studies in which we observed how store managers employed social robots in their stores. Each store manager requested different designs in the preparation phase. After deployment, we found that the managers were generally satisfied with the services autonomously offered by the robots, which successfully encouraged people to stop. The store managers were satisfied with the results and expressed a desire to use the robots again.

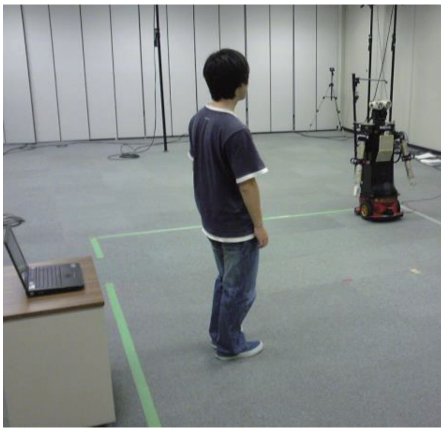
CONTENTS

1. INTRODUCTION

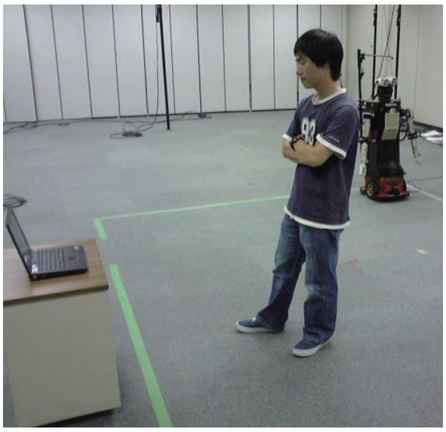
1.1. INITIATION OF CONVERSATION INTERACTION

How do you meet someone and start a conversation? Even though this might seem trivial for people, it is not at all trivial for robots. In a typical situation for humans, we stop at a certain position in relation to the target, greet the person, and find ourselves conversing. We do this almost unconsciously. As humans, we consciously think about the contents of the conversation after it has started.

In contrast, it is difficult for a robot to replicate what humans unconsciously do. It needs to know every detail of the behavior, such as where and when it should stop and what should be said; however, since we do this unconsciously, intricately describing what we are doing is not easy. For instance, consider a shop situation (Figure 1.1), where a customer has an appointment with a sales-robot to get a product explanation. The customer might wait at the entrance while looking toward the direction from which the robot is coming (Figure 1.1-a). Or he/she might look at another product displayed in the shop (Figure 1.1-b). Apparently the expected behavior for the robot is different in each situation, but what is the basis for generating the expected behavior for each situation?



(a) Looking at robot



(b) Looking at a product

Figure 1.1 Situations in a shop

In this study, we focus on the *initiation of conversation* in natural human-robot interaction. Clark modeled human communication based on the notion that people in a conversation share views of whether each of them is participating in the conversation or not and, furthermore, defined their *activity roles* [1], such as a speaker, hearer, or side participant. Kendon's analysis on spatial formation, known as F-formation is in line with this view so that the participants in a conversation form a particular shape [2]. Even though HRI researchers clearly recognize the importance of the participation state and spatial formation [3-6], no study has revealed how a robot should behave in different kinds of conversation-initiation interactions depending on the situation we denote as the *initiation of conversation*. In short, the above examples of the problem in Figure 1.1 remain unsolved.

To cope with this problem, we analyzed human behavior during the *initiation of conversation*. We learned the importance of two functions in our model:

- recognition of an interlocutor's spatial formation;
- constraints on a robot's spatial formation used to maintain the participation state.

Spatial formations that people establish in the interaction are used to model people's participation in the conversation. Likewise, behaviors they perform during the conversation are used to derive guidelines for how a robot should use its knowledge and structure its behavior to initiate and maintain a conversation. By overcoming these problems, we can realize our goal in this study, i.e., providing service through initiating a conversation on the robot's own initiative, and move one step closer toward smooth integration of robots into society.

We conducted a human observation experiment and provided the results of the data analysis. We created a model of initiation of conversation based on the observation results and implemented it on our humanoid robot. We firstly conducted a system evaluation and an objective evaluation to evaluate our model in an objective way, then conducted a subjective evaluation experiment to compare our model with two baseline models, and our proposed model was evaluated as the best.

1.2. INITIATION OF MOTION INTERACTION: DISTRIBUTIONAL HANDING

A number of robots have been developed in research projects that serve people in daily public environments. For instance, Gross and his colleagues developed a robot for a shop that assists consumers [7]. In museums, robots provide information to visitors [8-10]. Other studies demonstrated the use of robots in such environments as cities [6], streets [7], offices [13], hospitals [14], senior citizen facilities [15, 16], and shopping malls [17]. We believe that robots will soon start to perform many real tasks in our daily environments.

We believe that initiating motion interaction with people such as ‘distributing’ will be one future task for robots in daily environments. People commonly distribute such objects as flyers or free samples to pedestrians, for example, coupons to customers in a shopping mall, pamphlets to visitors in a museum, or a barbershop that gives advertising flyers to pedestrians in front of a crowded train station. We expect that in the future such distributing tasks will be carried out by robots.

How can we make a robot that performs such distribution tasks? Even though this activity might seem trivial for people, it is not trivial for robots. If a robot behaves poorly, its distribution task will probably fail and disturb the activities of pedestrians. We need to identify the key factors that comprise successful distributions. In this study, we investigate the behavior of people who perform distribution tasks well. After identifying the key factors, we implement them in a humanoid robot.

In this study, we define this distribution interaction as *distributional handing* and focus in natural Human-robot interaction. We first studied *distributional handing* in human-human interaction, and then implemented it into a humanoid robot (Figure 1.2).

Figure 1.2 A robot distributing a flyer to a pedestrian

1.3. Social Robots Initiating Interaction with People in the Real World

Attracting passersby is one critical task for store workers, and many do it daily. For instance, some clerks talk to passersby who stop at their shops and invite them to visit and browse. Some loudly announce the features and characteristics of their stores (Figure 1.3, left). However, such tasks are difficult for human workers. For instance, in a shopping mall, we witnessed a young female clerk who kept announcing:

"Hello, we have two kinds of pudding."

"Our products make delicious presents."

She robotically repeated her message all day even when few passersby were present. Even though we felt sorry for her, we did not buy anything from that store. A few weeks later, we heard that she had quit because the work was too stressful.

As robotics technology matures, why don't we use robots for such stressful tasks? They seem within the capacity of today's autonomous robots. Robots do not need to engage in complex conversations or decision-making; they just need to react to the arrival of passersby.

However, the required capabilities for attracting visitors remain largely unknown. Since it is also unknown whether autonomous robots with current technology can satisfy user expectations, we addressed them in a study. The following are our research aims:

- To identify the expectations, requirements, and design decisions of store managers
- To evaluate whether robots can autonomously serve in designed roles with our proposed models

Figure 1.3 Enticing passersby to a store

2. Related Literature

It is assumed that social robots will eventually engage in "natural" interaction with humans, i.e., interaction like humans do with other humans. The use of human-like body properties for robots has been studied to provide greater naturalness in the interactions. Often, studies have focused on the interaction after the robot meets people. For instance, studies have been conducted on pointing gestures [18, 19] and gaze [20-23].

Similar to the concept of *initiation of conversation*, researchers have studied the phenomenon of *engagement*. Engagement is a situation where people listen carefully to an interlocutor's conversation. A model has been developed for using the gaze behavior of robots [6] and people to recognize the engagement state [24, 25].

The main difference between the *initiation of conversation* and *engagement* is that the latter addresses a phenomenon that occurs after the people and the robots have established a common belief that they are sharing a conversation. In contrast, the phenomenon of *initiation of conversation*, which our study addresses, concerns the situation before or just at the moment when they establish this common belief of mutually sharing a conversation.

Within the research on human communication, studies are sparse on how humans initiate conversation beyond the basic facts that they select interaction partners and recognize and approach each other [26], stop at a certain distance [27], start the conversation with a greeting [28, 29], share a recognition of each other's state of participation [1], and arrange themselves in a suitable spatial formation [2]. Recent studies have started to reveal more detailed interaction, including the knowledge of detection of service initiation signals used in bars [30] and the finding that side participants stand close to the participants and often become the next participant [31]. But this new knowledge remains limited.

In HRI, spatial formation has been studied in relation to initiating conversation. Michalowski et al. revealed the relation between the robot's environment and the person's engagement toward the conversation, and they suggested that to improve the interaction it's important to put a stronger emphasis on movement in the estimation of social engagement and to vary the timing of interactive behaviors [4]. Hüttenrauch et al. used a Wizard-of-Oz study and found that people follow an F-formation in their interactions with robots, just as with humans [32]. Kuzuoka et al. studied the effect of body orientation and gaze in controlling F-formation and found that with these movements, a robot could lead the interaction partner to adjust his/her position and orientation while considering the proper F-formation [3]. Studies have also generated more natural robot behavior, such as the approach direction and distances to a seated person [33, 34] and the path to approach and catch up with a walking person [35, 36], the standing position for presenting a product [37], the proper distances for passing behavior [38] and following behavior [39], and the selection criteria for choosing an interaction partner [40]. A few studies have attempted to promote people's participation by encouraging behavior [5, 41] and detecting the requested behavior [42]. However, since these studies were aimed at encouraging people's participation, they only showed the one-sided behavior of the robot, not how robots should behave while considering the people's real-time status in the *initiation of conversation*. In our research, we proposed a model that could make the robot recognize the participation state of the people and then act accordingly to make them both participate in a conversation and maintain it.

Handing capability is built in a number of techniques in robotics. Recent progress in mobile manipulation [43, 44] is clearly relevant. Some studies have also investigated how to make ‘grasping’ socially acceptable [45].

Specific to handing interaction, early studies concentrated on generating natural handing motions that imitated humans [46-48]. Huber et al. showed that a minimum jerk model makes arm motions appear more natural and shortened the subjects’ response times [49]. Cakmak et al. designed a handing-over motion to convey the moment when the person accepts the object [50] and how to learn a preferable robot configuration for the task [51]. Sisbot et al. showed how to navigate a robot [52] and manipulate objects near humans [53]. Koay et al. presented their results from a human-robot interaction study that investigated the issues of participant preferences in terms of a robot-approaching method and handing behavior in the context of a robot handing an object to a seated person [52]. The use of perspective for joint manipulation has also been addressed [54]. However, most of these researches focused on behaviors for handing an object to a specified person who was stopped at a fixed position. For handing objects to walking pedestrians, the following necessary knowledge is very different from conventional researches, such as choosing a pedestrian as the target, approaching the person, the timing of the approach, and how the robot should extend its arm to provide the object to the target. This knowledge remains unknown.

In addition to the handing interaction, a few studies have also addressed the process of initiating interaction by considering proxemics [55] and inviting behavior [56]. These studies focused on ways for a robot to exhibit intention to initiate interaction, but in these cases the robots were stationary, which makes a quite different situation from our work: distribution to pedestrians.

Some studies focused on how robots should approach humans. Dautenhahn et al. studied what kind of approaching behavior by a robot was preferred by users and concluded that when they are seated, approaching from the side is better than from the front [57]. Satake et al. proposed an algorithm for proactive approaching in which a robot meets people from the front [58]. Shi et al. modeled people's behavior to initiate conversation in shopping situations, where the mutual spatial configuration was modeled rather than the approaching direction [59]. These studies show that approach methods for robots are different across each different situation, i.e., the state of the target person, the robot's interaction purpose, etc. This result suggests that we should specifically investigate a proper approaching method for distributing to pedestrians.

Previous studies have identified the promising contexts (roles, tasks, and situations) in which social robots can successfully serve. For instance, in a museum a robot attracted visitors' attention and explained exhibits [60]. Other robots were also successful in museum/exhibition contexts [61-64]. Some studies revealed that robots can perform other tasks, including receptionist [65], snack delivery [66], health management [67], and education [68, 69].

Some studies specifically addressed visitors or passersby of stores. A robot led visitors to request products/items in a store [70], and another robot improved the atmosphere of a transit area and a shopping space [71]. On the street, a robot successfully collected information from passersby [72]. In a shopping mall, a semi-autonomous robot successfully provided directions to stores and recommended stores [17] and distributed discount coupons [73].

Many research works have investigated social acceptance. For instance, Weiss et al. investigated the social acceptance of robots from the observations of people's reactions [74]. Heerink et al. developed a model of social acceptance [75]. In their model, the perceived ease of use by people as well as their perceived enjoyment was considered the source of their intention to use robots. Acceptance is considered the consequence of the satisfaction of the diverse needs of users and their expectations [76]. Many established methods have evaluated social acceptance from the perspective of people who interacted with robots.

Some studies explored factors beyond the perceptions of interlocutors. For instance, Salvini et al. argued that acceptability must be considered beyond user level and included views from bystanders as well as legal and ethical perspectives [77]. Mutlu et al. identified the importance of organizational factors through an ethnological study [78].

For robots to work in such environments as stores, the perceptions of administrators (in our case, store managers) must also be addressed. They are the decision makers who will determine design choices. They will judge whether to employ robots. However, the perceptions of administrators has been overlooked in previous studies. Thus, our study is novel because it approaches social acceptance from a managerial perspective.

3. Initiation of Conversation Interaction

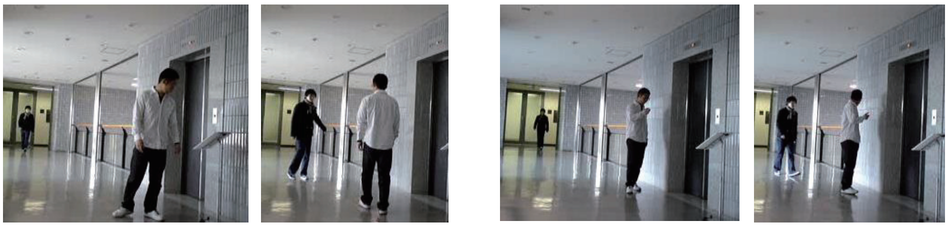
3.1. Modeling Initiation of Conversation

To find the regular patterns in people's behavior at the moment of the *initiation of conversation*, we observed the interaction of two people when they started a conversation. We focused on their spatial formation and gaze, both of which have been discussed in the literature as important factors for human communication [79].

3.1.1. Data Collection

We collected data in two different settings, *shop* and *meeting* scenarios, to find the consistencies and differences across different purposes and environments. In each scenario, one person initiated conversation with the other. We assumed that whether a participant plans to explain an object or lead another to a location in the store after the initial greeting influences how that person behaves in the *initiation of conversation*. Based on this assumption, we divided each scenario into two situations.

(a) Shop scenario



(b) Meeting scenario

Figure 3.1 Examples of initial positions in two scenarios

Shop scenario: This interaction was conducted in a 5 x 5-m room in which four objects were placed (Figure 3.1-a). One person behaved as a *visitor* waiting in the shop, and the other person acted as a *host* (a clerk) who greets the *visitor* and either (1) offers a service or (2) explains products.

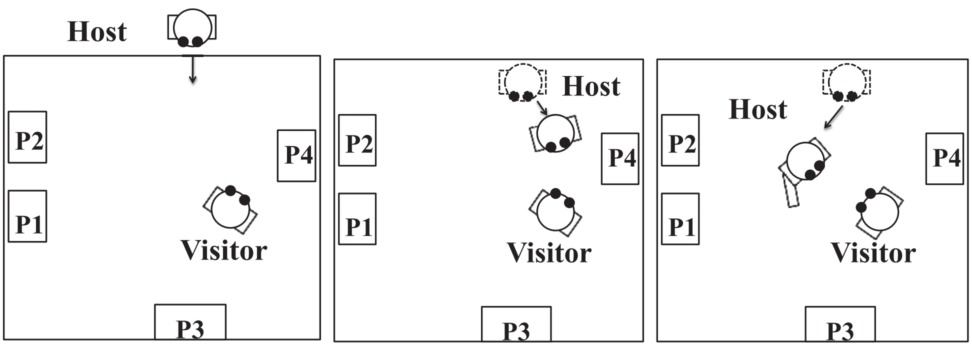
Meeting scenario: This interaction was conducted in the lobby (4 x 10 m) of a research institute (Figure 3.1-b). One person acted as a *visitor*, and the other behaved as a *host* who meets the *visitor* and either (1) offers help or (2) leads the *visitor* to another location.

We set the initial position of the *host* out of sight of the *visitor*, and then the *host* entered the environment to initiate conversation. The experimenter provided either of two *plans:* the *host* only needs to greet the *visitor* in *without plan* or explain a product (or lead the *visitor*) in *with plan*. With this setting, we observed how they behaved both verbally and non-verbally to initiate a conversation.

Twenty Japanese undergraduate students (ten pairs, eleven men and nine women) were paid for their participation in this data collection. We had confirmed that the two participants in a pair did not know each other before the experiment. The participants could make sure about the environment (ex., the products put in the shop) before the interaction so that they could provide information to the visitor easily. They repeated each scenario ten times (after five trials, they switched roles, so each acted in one role five times for each scenario). We asked the *visitor* to position himself/herself differently every time so that we could collect diverse data. Beyond these instructions, the participants were allowed to behave freely.

Although we specified the roles that the participants acted, the behaviors in the whole interaction were done freely by the participants. We did not determine their detailed behaviors; we only planned their roles and asked them to behave while considering these roles (we asked participants to not repeat the most recent behavior). Thus, the situations that both the *host* and the *visitor* faced were often different. By analyzing the detailed behaviors that the participants had both unconsciously and consciously carried out, we wanted to find out the regular patterns of people's interaction when initiating a conversation.

The interaction data was collected with one video camera. We set the camera at the place from where its field of view could cover the whole interaction of the two people. We have put some marks on the floor to help with the data analysis such as retrieving distance and angle parameters.



(a) initial setting **(b) without a subsequent plan** **(c) with a subsequent plan**

Figure 3.2 Influence of *subsequent plan* in *initiate position*

3.1.2. DATA ANALYSIS

Participants took diverse spatial formations and behaviors when they initiated conversations. For example, the *host* sometimes directly approached and greeted the *visitor,* saying, "Welcome, may I help you?" in the central area (Figure 3.2-b); in other cases the *host* moved to the side of the *visitor* and only spoke first when he/she reached a position near the *visitor* (Figure 3.2-c). To retrieve the systematic patterns in such *initiations of conversation*, we observed the position and timing of the *host's* performance: (1) how to initiate conversation (*initiation behavior*), (2) where to initiate conversation (*initiation position*), (3) where to talk (*talking position*), and (4) how to talk (*utterances*).

3.1.2.1. Patterns of Initiation Behavior

In our preliminary analysis of how the *hosts* behaved, we found that their choice of *initiation behavior* was influenced by two factors: *visibility* and *plan*. For example, most *hosts* directly approached the *visitors* when the *visitors* noticed them or when the *hosts* did not have a *plan*. On the other hand, most *hosts* approached the place where both the *visitor* and the next target (e.g., product or a route to the next location) are visible when the *hosts* had a *subsequent plan* and the *visitors* did not notice the host. From these observations, we coded all situations to scrutinize the differences in the *host*'s behavior patterns. We used Cohen's Kappa, an index of inter-rater reliability that is commonly used to measure the level of agreement between two sets of dichotomous ratings or scores [80]. We asked two coders who have no knowledge about robotics and HRI to analyze the collected data. They did not participate in the data collection experiment and did not know about the purpose of the collected data. They were only told to analyze the data based on their own cognition. First, the two coders classified *visibility* into two cases: the *visitor* noticed the *host* (*noticed*) and the *visitor* did not notice the *host* (*unnoticed*). Moreover, we analyzed the *initiation behavior*, which coders classified into two cases: *approach to visitor* and *approach to a place where both visitor and target are visible*.

Cohen's Kappa coefficient from the two coders' classifications was 0.87 for *visibility* and 0.84 for *initiation behavior*, indicating that their classifications were highly consistent. After the classifications, to analyze the consistent trajectories for modeling, the two coders discussed and reached a consensus on their classification results for the entire coding process.

Table 3.1 Analysis of initiation behavior

Scenario	Plan	Visibility	Initiation behavior: Approaching visitor	Initiation behavior: Approach to a place where both visitor and target are visible
Shop (100 cases)	With plan (50 cases)	Noticed (18/50)	18 (100%)	0 (0%)
		Unnoticed (32/50)	3 (9.3%)	29 (90.7%)
	Without plan (50 cases)	Noticed (16/50)	16 (100%)	0 (0%)
		Unnoticed (34/50)	34 (100%)	0 (0%)
Meeting (100 cases)	With plan (50 cases)	Noticed (24/50)	21 (87.5%)	3 (12.5%)
		Unnoticed (26/50)	8 (30.7%)	18 (69.3%)
	Without plan (50 cases)	Noticed (29/50)	29 (100%)	0 (0%)
		Unnoticed (21/50)	21 (100%)	0 (0%)

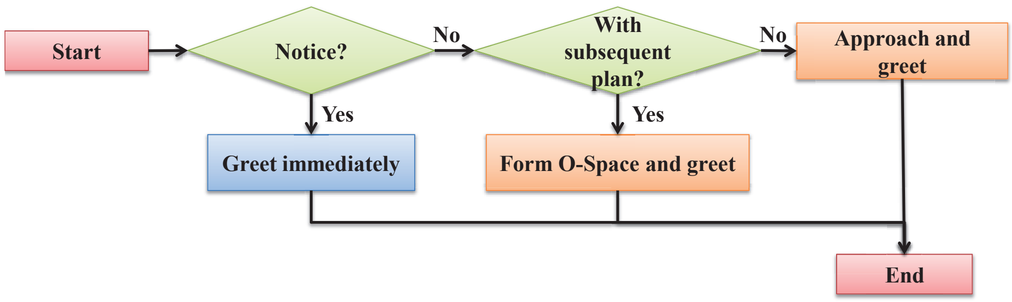
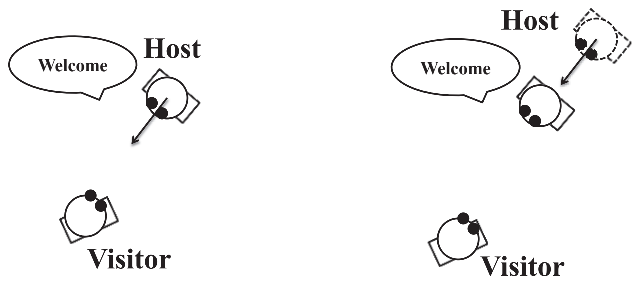


Figure 3.3 Choice of *initiate timing* and *position*

The coding results are shown in Table 3.1, which confirms our observation. We found that when the *visitor* did not notice the *host*'s arrival when the *host* had a *subsequent plan*, most *hosts* tended to choose a *behavior* by considering their *subsequent plans* regardless of their scenario. In addition, at this time the host formed a spatial formation with the visitor while considering the target product, in a way similar to using O-space [37]. O-space is a convex empty space surrounded by the people involved in a social interaction, where every participant looks inward into it to share attention to the same product, and no external person is allowed in this region. The *hosts* always moved toward the *visitors* to greet them when they did not have *subsequent plans* in both scenarios; even if the *hosts* did have *subsequent plans*, most moved to the *visitors* when they were noticed by the *visitors*. As shown in Figure 3.3, in summary, we found that the choice of *initiation behavior* was influenced by whether the *hosts* had a further plan to explain something to the *visitor*. However, this choice is also influenced by *visibility*. If the *visitor* noticed the *host* within a certain distance, the *host* moved to the *visitor* to initiate the conversation.

3.1.2.2. Initiation Position

In our preliminary analysis of the timing of the initiation of the *hosts*, we found that their position was influenced by the *greeting pattern* and the *position relationships*. For example, when the *visitors* were noticed by the *hosts*, the *hosts* immediately greeted the *visitors* as they approached, but some *hosts* greeted the *visitors* after approaching the *visitors* when they were far away. Moreover, if the *visitors* were not noticed by the *hosts*, the *hosts* approached the *visitors* differently, depending on their initial position relationships.



(a) Greet immediately **(b) Greet after apporaching**

Figure 3.4 Detailed analysis of *initiation position* in *notice* category

From these observations, we coded the *host*'s *greeting patterns* to scrutinize the differences in their behavior patterns. Again, the two coders classified the *greeting patterns* into two cases separately for both *noticed* and *unnoticed* case: the *host* greets *visitors* immediately (Figure 3.4-a), the *host* greets *visitors* after approaching them (Figure 3.4-b); the *host* approaches from the frontal direction and then greets, and the *host* approaches from the non-frontal direction and then greets.

Cohen's Kappa coefficient from the two coders' classification was 0.93 for *noticed* and 0.84 for *unnoticed* for *greeting patterns,* indicating that their classification was highly consistent. After classification, to analyze the consistent trajectories for modeling, the two coders discussed and reached a consensus on their classification results for the entire coding process.

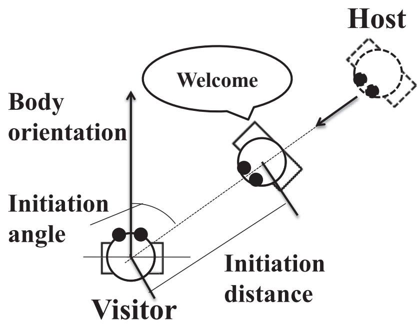


Figure 3.5 Initiation distance and initiation angle

We further analyzed the position relationships between the *host* and *visitor*. First, we measured the distance (*initiation distance*) and angle (*initiation angle*) (Figure 3.5) when the *host* attracted the attention of the *visitor* by saying, “Excuse me” or “Welcome,” because the position relationship in this timing is essential to understanding how the *host* initiates participation.

In the *noticed* category, we found that the *initiation distance* is different depending on the scenario and greeting patterns. In the *shop scenario*, the average for *initiation distance* was 2.2 +/- 0.2 m and 2.5 +/- 0.3 m for “greet immediately” and “greet after approaching.” In the *meet scenario* the average of *initiation distance* was 3.3 +/- 1.5 m and 6.2 +/- 1.0 m for “greet immediately” and “greet after approaching.”

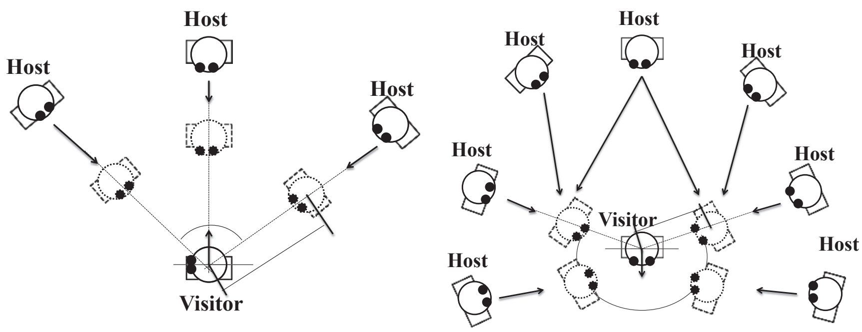
Our interpretation is that the *host* immediately greets the *visitor* when the distance from the *visitor* is lower than a certain distance, but the *host* does not immediately greet the *visitor* when the distance from him/her is greater than a certain distance when the *visitor* notices the *host*. Note that the *initiation angle* is not measured in the *noticed* category because the *visitor* and the *host* face each other.

On the other hand, in the *unnoticed* category, the *initiation distance* was not influenced by the *scenario*. In the *shop scenario*, the average of the *initiation distance* was 2.0 +/- 0.1 m and 1.5 +/- 0.3 m for “approach from frontal” and “approach from non-frontal” directions, respectively, and in the *meet scenario* the average of the *initiation distance* was 2.0 +/- 0.6 m and 1.6 +/- 0.4 m for “approach from frontal” and “approach from non-frontal” directions, respectively.

Since the *initiation distances* in “approach from frontal” and “approach from non-frontal” directions were obviously different, we measured the *initiation angle* to find the extent of these two *greeting patterns*. In the “approach from frontal” category, the maximum angle between the vector from the *visitor* to the *host* and the *visitor's* orientation was 55° on the left and 60° on the right side in the *shop scenario* and 65° on the left and 50° on the right side in the *meeting scenario*. On the other hand, in the "approach from non-frontal" category, the ranges of minimum to maximum angle between the vector from the *visitor* to the *host* and the *visitor's* orientation were 55~120° and 60~130° on the left and right sides in the *shop scenario* and 65~130° and 50~135° on the left and right sides in the *meeting scenario*. The minimum of this angle was the same as the maximum in the "approach from frontal" cases.

Table 3.2 Analysis of initiate position (distance and angle) and distance to talk

Scenario	Visibility	Greeting pattern	Initiate distance	Initiate angle (maximum)	Talk distance
Shop (100 cases)	Notice (34 cases)	Greet immediately (16/34)	2.2 +/- 0.2	-	0.7 +/- 0.1
		Greet after approaching (18/34)	2.5 +/- 0.3	-	0.8 +/- 0.4
	Not notice (66 cases)	Approach from frontal (18/66)	2.0 +/- 0.1	55~60	0.7 +/- 0.1
		Approach from non-frontal (48/66)	1.5 +/- 0.3	120~130	0.7 +/- 0.2
Meeting (100 cases)	Notice (53 cases)	Greet immediately (42/53)	3.3 +/- 1.5	-	0.7 +/- 0.2
		Greet after approaching (11/53)	6.2 +/- 1.0	-	1.2 +/- 0.5
	Not notice (47 cases)	Approach from frontal (17/47)	2.0 +/- 0.6	65~50	0.8 +/- 0.4
		Approach from non-frontal (30/47)	1.6 +/- 0.4	130~135	0.6 +/- 0.1



(a) From front direction **(b) From non-front direction**

Figure 3.6 Detailed analysis of *initiation position* in *unnoticed* category

We conclude that the *hosts* chose their positions not only considering the distance but also the direction, depending on the position relationships. As shown in Figure 3.6-a, when the *hosts* came from the *visitor's* frontal side, they always went straight toward the *visitor*. When the *hosts* came from behind the *visitors* (Figure 3.6-b), instead of going toward the *visitors*, the *hosts* went to their side to make sure that they were in the *visitors'* field of view before starting to talk. In addition, the distance at which they started to greet the *visitor* was influenced by whether the *host* came from the *visitor's* frontal side.

3.1.2.3. Talking Position

Next, we measured the position relationships between the *hosts* and the *visitors* when they started to talk (e.g., explaining products or leading movement) in each category. As a result, the *host* kept walking toward the *visitor* while greeting until the *host* was within a proper distance for talking to the *visitor*. We found that this distance, which averaged about 0.7 m, was common to both scenarios, except for the "greet after approaching" category in the meeting scenario.

3.1.2.4. Analyzing Utterances

Finally, we investigated how the *host* starts to talk with the *visitor*. We found that the utterances the *host* used to initiate the conversation were influenced by whether the *visitor* was considered to participate in the conversation or not. After the *visitor* noticed the *host's* arrival, the *host* greeted the *visitor* with an expression like "Welcome." It seemed to them as if they had already agreed to participate in a conversation. We called this mental agreement the *participation state*. When the *host* initiated the conversation from the side of the *visitor* without making eye contact, the *host* first needed to attract the *visitor*'s attention. This situation is called *visitor not participating* in the conversation. Consequently, when the *host* was noticed by the *visitor* or was coming from the frontal direction of the *visitor* within a certain distance, the *visitor* was considered to be participating in a conversation with the *host*, and thus the *host* needed to make an utterance immediately. When the *host* was coming from the non-frontal direction of the *visitor* within a certain distance ("Approach from non-frontal" case in Table 3.2, 48 trials in shop scenario and 30 trials in meeting scenario), only the *host* was considered to be participating in a conversation toward the *visitor* (but the *visitor* was not yet participating). It is not necessary for the *host* to utter something at once. However, to make the *visitor* participate in the conversation, the *host* first needs to either adjust the spatial formation with the *visitor* or say a phrase like "Excuse me" to attract the *visitor*'s attention (31/48 trials in shop scenario and 22/30 trials in meeting scenario).

We found that the above phenomena were shared by both scenarios, except for the threshold distance when they started a conversation. We concluded that the basic phenomena in initiating conversation were common among scenarios and environments.

3.1.2.5. Summary

In this data collection, we conducted our observation experiment in a simple lab situation. For meeting scenario, we consider that the environment is as the same as the real world and the situation is very common. While for the shop scenario, the decoration of our shop is simple and not all the participants had training or experience in how to behave as a shopkeeper in a shop. However, our purpose is to find common human behavior when initiating conversation instead of shopkeeper-specific behavior. We consider that it is appropriate to assume that the participants have the common sense needed to naturally initiate conversation with others.

We found four key points for initiating conversations: patterns of initiation behavior, initiation position, talk distance, and utterance. Moreover, we found several factors that influence them: scenario, plan, visibility, and greeting pattern. Patterns of initiation behavior are influenced by plan and visibility (situation dependent); initiation position and talk distance are influenced by scenario, visibility, and greeting pattern (situation and environment dependent). Utterances are influenced by greeting pattern (situation dependent).

3.2. A ROBOT THAT ADDRESSES INITIATION PROCESS

We implemented our model in a robot so that it appropriately addressed the *initiation of conversation*, i.e., choosing an appropriate position to start talking with appropriate timing.

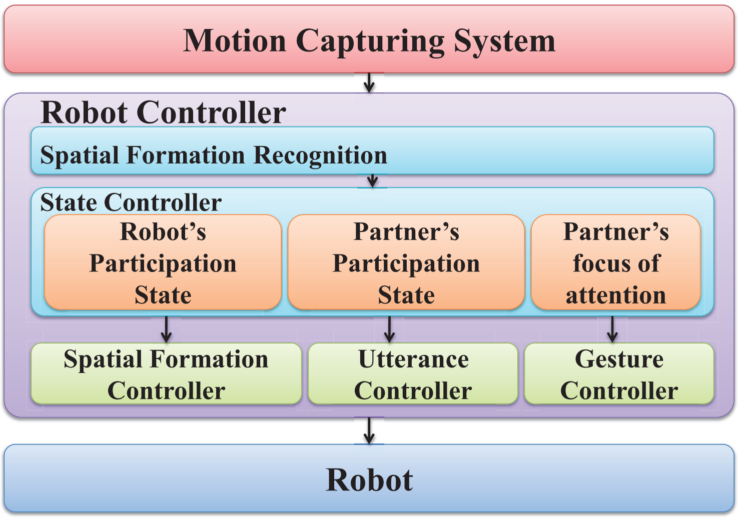


Figure 3.7 System configuration

3.2.1. GENERAL FRAMEWORK

We used a development framework that we had used successfully before to control the robot automatically [81]. Figure 3.7 shows an outline of our framework, which has three components: a humanoid robot, a motion capture system, and a robot controller (software). Control of the robot is carried out automatically without an operator. The *spatial formation recognition* function uses as input the position and orientation information of the robot, human and target from the motion capture system to recognize the spatial formation. The *state controller* receives the information from the *spatial formation recognition* and sends the state information to the *spatial formation*, *utterance,* and *gesture controllers*. The *spatial formation controller* calculates the target position for the robot every 100 ms and then generates and sends commands that consist of forward velocity and rotation velocity to the robot automatically to control its movement. The developer writes commands in advance with a markup language that can both control the robot's *gesture* and *utterance*, and the robot automatically uses them according to the information from the state controller [81].

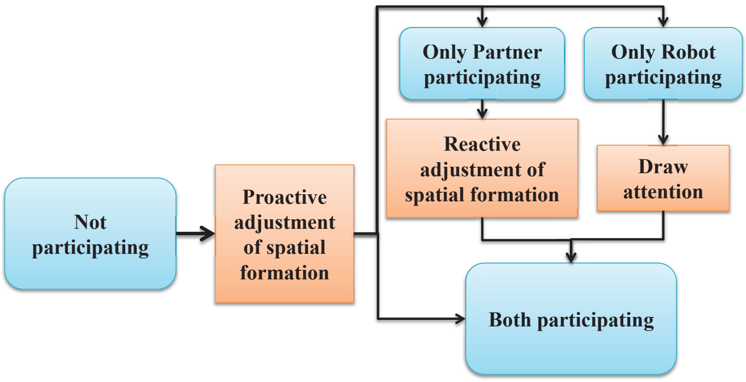


Figure 3.8 Flow of initiating conversation

Figure 3.8 shows the robot's flow for *initiation of conversation*. There are two paths that can be taken until the conversation starts. In one case, the robot initiates *participation.* It approaches, stops at an appropriate position (*proactive adjustment of spatial formation*), and attracts the visitor to participate in the conversation with a *drawing attention* action.

In the other case, the visitor initiates the conversation. While the robot is moving to a certain position (for *proactive adjustment of spatial formation*), the visitor prepares to initiate the conversation. Thus, the visitor's participation state changes to *participating* first, and then the robot adjusts its spatial formation to be appropriate for the *participation state*. In this case, it performs a *reactive adjustment of spatial formation*.

3.2.2. HARDWARE

We used Robovie-II, a 1.2-m-tall humanoid robot with a 0.3-m radius that is characterized by its human-like body expressions. It has a 3-DOF head and 4-DOF arms.

Its mobile base is equipped with wheels. Its maximum speed is about 0.7 m/s. And in our experiment we set the maximum speed of the robot as 0.5m/s for security reasons.

Since our research focus is to confirm our model's validity, we used a motion capture system as the sensor input. The motion capture system acquires body motions and outputs the position data of markers to the system. It outputs the data in real time with a 100-ms output cycle, and the error is less than 2 mm. Twenty-three markers were placed on the human and robot bodies, and four markers were attached to each product that was used for a *subsequent plan*.

3.2.3. SPATIAL FORMATION RECOGNITION

3.2.3.1. PARTICIPATION STATE

We define the *visitor's* and *robot's participation states* to indicate whether the human and the robot are participating in a conversation. We define the *participation states* of the robot and the human as PS_R and PS_H. When the robot is participating in a conversation, PS_R =1; otherwise, PS_R =0. When the human is participating in a conversation, PS_H =1; otherwise, PS_H =0.

We also define a *joint participation state* to show the relationship between the robot and the visitor in the conversation as PS_J (i.e., PS_R, PS_H). There are four state variables of the *joint participation state* in the implementation.

- **No one participating**

This state variable, which indicates a situation where neither the robot nor the visitor is participating in the conversation, is defined as PS_J= (0, 0).

- **Only robot participating**

This state variable indicates a situation where only the robot is participating in a conversation with the visitor, i.e., PS_J= (1, 0). Although the robot is considered to be participating in a conversation with the human, the human does not realize that the robot is approaching. In this case, the robot is allowed to greet the human, but it can also adjust its position to a better place instead of talking immediately. In addition, in this state, the robot should say something like "Excuse me" to draw the human's attention and initiate conversation. As the human starts to participate in the conversation, the robot begins to greet the human.

- **Only visitor participating**

This state variable indicates a situation where only the visitor is participating in a conversation with the robot, i.e., $PS_J= (0, 1)$. This means that only the human is considered to be participating in a conversation with the robot. It is possible that the *visitor* recognizes the robot and wants to say something to the robot before the robot greets him/her. However, as we found in the observation experiment, implicit behaviors always come before the explicit ones. Meanwhile, before the explicit contact (like saying a word), implicit behaviors such as standing position, body orientation and gaze would be established first. Since in our model the *participation state* could be detected by analyzing the spatial formation, the robot would always realize the *visitor's* intention and participate in the conversation at once. In this case, the robot must adjust the spatial formation to participate in the conversation and greet the human.

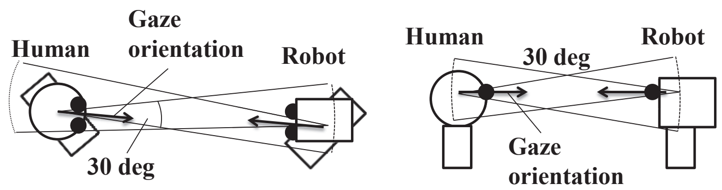
- **Both participating**

This state variable indicates a situation where both the robot and the visitor recognize the conversation possibility and are paying attention to each other. We record it as $PS_J= (1, 1)$. This means that since both the robot and the human are participating in the conversation with each other, the robot should immediately greet the human.

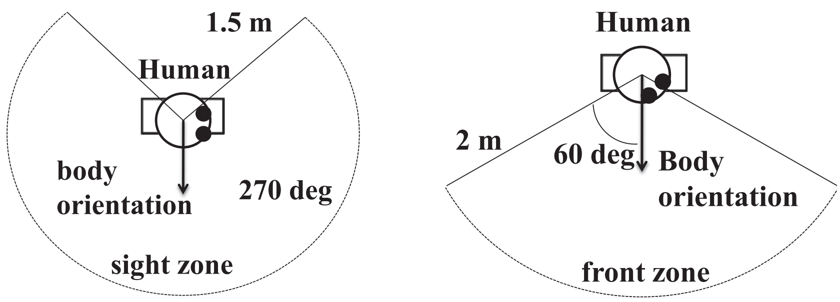
Dialogue act tags have been annotated for each phrase in a database of dialogue between several pair of speakers, according to the following set, based on the tags proposed in [40], taking into account dialogue acts such as affirmative or negative reaction, expression of emotions like surprise or unexpectedness, and turn-taking functions.

3.2.3.2. Participation Zone

Estimation of the *participation state* is a key component of this study. From our observations of human interaction, we found that people initiated conversation (a) when their gaze met within a certain distance and (b) inside the visitor's field of view within a certain distance when the visitor didn't notice the other's arrival. From these observations, we hypothetically developed a *participation zone* that consists of three parts: *gaze*, *sight,* and *front zones*. The *gaze zone* is the space established by one's gaze; if two people are in each other's *gaze zone* (their gazes meet), they perceive an obligation to participate in a conversation. The *sight zone* is a cone-shaped space established in front of a person to represent one's sight; if one person wants to initiate participation with another, he must enter the visitor's *sight zone* first (when their gaze does not meet). The *front zone* is an obtuse fan-shaped space established in front of a person to represent one's frontal side; if a person enters the visitor's *sight zone* and keeps the visitor in his own *front zone*, he perceives an obligation to participate in a conversation. When both people enter each other's *front zones*, they both perceive an obligation to participate in a conversation.



(a) gaze zone



(b) sight zone **(c) front zone**

Figure 3.9 Participation zone

With the three *participation zones* defined above, it is possible to estimate whether a person is participating in a conversation with another, and thus to determine the proper initiation pattern, initiation position and utterance.

Next, we report the method of estimating *participation zones*. In addition, estimation of the *visitor's* focus of attention is also needed when the *host* has a *subsequent plan*. The parameters we use below are derived from our observation experiment or models that were used successfully in previous research efforts. As reported in Section 3.1.2.5, parameters for *gaze zone* are situation- and environment-dependent. Even though the *front zone* and the *sight zone* are independent of the situation and the environment, it may also be necessary to adjust their parameters to position the robot.

- **Estimation of participation zone**

Since it is not easy to detect a person's gaze accurately, we used a simple technique that analyzes the person's head orientation instead. Figure 3.9-a illustrates the *gaze zone,* which is set as a 30° cone-shaped area (parameter was adjusted according to the accuracy of our motion capture sensor) in front of a person's (or robot's) head within a changeable distance. When the robot is in the human's *gaze zone*, we assume that the human is looking at the robot and realizes the robot is approaching.

We use Eq. 1 to calculate whether the robot is in the human's *gaze zone*:

$$IsInGazeZone(P_H, P_R, \theta_G) = \begin{cases} 1 & Dist(P_H, P_R) < InitiateDistance_{Gaze} \text{ and } |Angle(\theta_{P_H,P_R}, \theta_G)| < 30\deg \\ 0 & (otherwise) \end{cases} \tag{1}$$

where P_R is the position of the robot in the environment near the person and P_H is the position of the person. $Angle(\theta_{P_H,P_R}, \theta_G)$ is a function that indicates the constraint of the human's gaze orientation. We used $InitiateDistance_{gaze}$, which we analyzed in Section 3.1.2.2, as the length of the *gaze zone* and set it to 2.5 m in the evaluation experiment based on our observations (initiation distance of *Greet after approaching* in shop scenario in Table 3.1). θ_G is the human's gaze direction. Parameter $Dist(P_H, P_R)$ is in the x-y coordinate, and $Angle(\theta_{P_H,P_R}, \theta_{G,})$ is in the x-y-z coordinate. If the value of the position of robot P_R is not 0, the robot is in the human's *gaze zone*.

We set up precise parameters to define the *sight zone* from our observation results (initiation distance and initiation angle of Approach from non-frontal direction in Table

3.1), and thus the zone was set to a 270° fan-shaped area in front of the body of a person (or robot) within a 1.5-m distance (Figure 3.9-b).

We defined Eq. 2 to calculate whether the robot is in the human's *sight zone*:

$$IsInSightZone(P_H, P_R, \theta_H) = \begin{cases} 1 & Dist(P_H, P_R) < InitiateDistance_{Sight} \text{ and } |Angle(\theta_{P_H,P_R}, \theta_H)| < (InitiateAngle_{Sight}/2) \\ 0 & (otherwise) \end{cases} \quad (2)$$

where we use $InitiateDistance_{Sight}$ (1.5 m) and $InitiateAngle_{Sight}$ (270°), which we analyzed in Section 3.1.2.2, as the length and angular region of the *sight zone*. All of the parameters here are in the x-y coordinate.

We set-up precise parameters to define the *front zone* from the social distance [27], observations reported in Section 3.1.2 (initiate distance and initiate angle of Approach from frontal in Table 3.1), and the preliminary tests. Accordingly, the zone was set to a 120° fan-shaped area in front of the body of a person (or robot) within a 2.0-m distance (Figure 3.9-c).

We defined Eq. 3 to calculate whether the robot is in the human's *front zone*:

$$IsInFrontZone(P_H, P_R, \theta_H) = \begin{cases} 1 & Dist(P_H, P_R) < InitiateDistance_{Front} \text{ and } |Angle(\theta_{P_H,P_R}, \theta_H)| < (InitiateAngle_{Front}/2) \\ 0 & (otherwise) \end{cases} \quad (3)$$

where we use $InitiateDistance_{Front}$ (2.0 m) and $InitiateAngle_{Front}$ (120°), which were analyzed in Section 3.1.2.2, as the length and angular region of the *front zone*. All of the parameters here are in the x-y coordinate.

When these conditions are satisfied, the *participation state* changes from *not participating* to *participating*. However, the opposite is not true; since the transition of the *participation state* from *participating* to *not participating* needs verbal interaction, it is not controlled in this estimation module.

3.2.3.3. Visitor's focus of attention

As reported in Section 3.1.1, whether the visitor is paying attention to the target product, which the robot would explain as a *subsequent plan*, influences the robot's standing position. Therefore, we need to recognize the visitor's focus of attention.

We used a previously reported method [37] that identifies an object in transactional segments as the focus of implicit attention. A person's transactional segment is defined as the space in front of him/her when there is no obstacle between the person and the object. When the angle between the forward direction of the person's body and the vector from his/her body center to an object is less than 90° and the distance between him/her and the object is less than 2 m, the object is identified as the person's implicit attentional target (Figure 3.10).

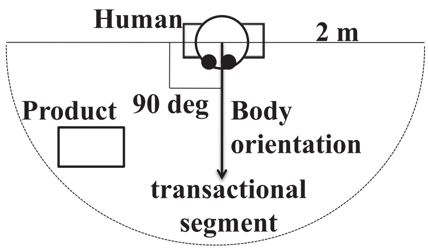


Figure 3.10 Transactional segment

If an object is in a person's *transactional segment*, we assume that the person is paying attention to it. We used Eq. 1 to calculate whether an object is in the person's *transactional segment*:

$$IsInTransactionalSegment(P_H, P_O, \theta_H) = \begin{cases} 1 & Dist(P_H, P_O) < 2000mm \text{ and } |Angle(\theta_{P_H,P_O}, \theta_H)| < 90\deg \\ 0 & (otherwise) \end{cases} \tag{4}$$

Here, P_O is the position of an object in the environment, P_H is the person's position, θ_{P_H,P_O} is the vector from the person's body center to P_O, and θ_H is the person's body orientation. $Dist(P_H, P_O)$ is the distance between the object and the person. $Angle(\theta_{P_H,P_O}, \theta_H)$ is the angle between the vector from P_H to P_O and the person's body orientation. All of the parameters here are in the x-y coordinate. If the value of the position of object P_O is not 0, the object is in the human's *transactional segment* and the human is paying attention to it. Here, we only used this simple method to estimate the person's focus of attention due to our sensor and experimental setting. This model gives the robot the basic ability to provide services according to the *visitor's* focus of attention.

In some environments where objects are placed tightly, the recognition precision is one limitation of the model. However, one could easily use other methods for the task, since many researchers have already addressed this issue.

3.2.3.4. RECOGNITION OF THE PARTICIPATION STATE

We recorded the situation where the robot is in the human's *gaze*, *front,* and *sight zones* as H_Gaze, H_Front, and H_Sight, and the situation where the human is in the robot's *gaze* as R_Gaze, R_Front, and R_Sight. Table 3.3 shows the relationship among *joint participation state* PS_J and the three *participation zones*.

Table 3.3 Definitions of Joint Participation State

	H_Gaze	H_Front	H_Sight	Else
R_Gaze	(1,1)	(1,1)	(1,0)	(0,0)
R_Front	(1,1)	(1,1)	(1,0)	(0,0)
R_Sight	(0,1)	(0,1)	(0,0)	(0,0)
Else	(0,0)	(0,0)	(0,0)	(0,0)

3.2.4. SPATIAL FORMATION CONTROL

A conversation is always carried out when both people perceive themselves to be participating in it. When a robot attempts to initiate a conversation with a *visitor*, the most important thing is to ensure that both the *visitor's* and its own *participation state* are set to *participating.* We created a *spatial formation controller* to control the robot's position and orientation to achieve this.

This unit controls the robot's standing position with a motion capture system. The system seeks the optimal standing position for the robot in a search area. A cell establishing a 20 x 20-cm standing position divides the search area (Figure 3.11). This module estimates the values of all cells in the search area and selects the one with the highest value as the optimal standing position. Then the robot goes directly toward the position, stops and adjusts its orientation. The position is updated every 100 ms.

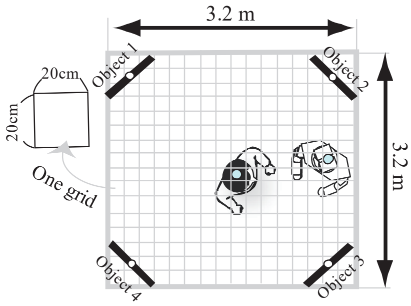


Figure 3.11 Searching grid

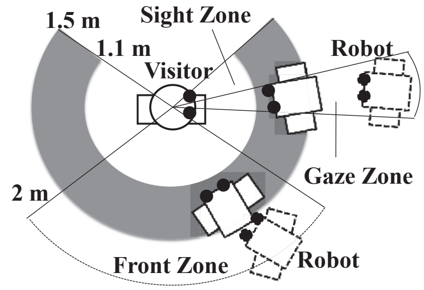


Figure 3.12 Reactive adjustment of spatial formation

From our observations of human-human interaction, we found the following: (a) The *host* kept facing the *visitor* and gazing at him/her within a certain distance when the *visitor* was participating; (b) when the *visitor* was not participating in the conversation, people always went to the position from where they could easily explain the target product or direction to the *visitor* if necessary. Thus, we created two models to control the spatial formation.

- **Reactive adjustment of spatial formation**

When the visitor is participating in the conversation, the robot needs to not only immediately participate in it but also get closer to the *visitor* and turn to him/her. We define this adjusting of position and orientation as the *reactive adjustment of spatial formation.* When the visitor is participating in the conversation, the robot should immediately start this adjustment, even if it has a previously made plan. We identified three rules for the *reactive adjustment of spatial formation* (Figure 3.12):

1) The robot should be at a position that allows itself to remain in the *sight zone* of the *visitor*.

2) Our observation on human-human interaction in Section 3.1.2.2 showed that the proper talking position is about 0.7 m, which ranges from 0.5 to 1.2 m. However, it is risky to place the robot too close to the visitor. Thus, in our implementation, we set the robot at a position that maintains a distance of about 1.1 to 1.5 m to the *visitor* (used successfully earlier [37]).

3) The robot should not turn to other orientations. It must keep facing the visitor to keep *participating*.

We calculated the distance between the robot position and each cell so that the robot could choose the nearest cell as its target position.

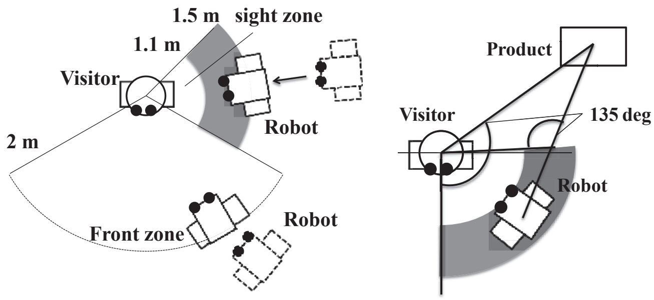


Figure 3.13 Proactive adjustment of spatial formation

We calculated the values of each cell for reactive adjustment using Eq. (5):

$$Value_{ReactiveAdjustment}(P_H, P_R, P_T, \theta_H) = \begin{cases} 1/Dist(P_R, P_T) & Dist(P_R, P_H) < 1500 \text{ and } |Angle(\theta_{P_H,P_R}, \theta_H)| < 60\,\text{deg} \\ 0 & (otherwise) \end{cases} \tag{5}$$

where P_T is the position of each cell, as shown in Figure 3.9. P_R is the temporal position of the robot. All of the parameters are in the x-y coordinate.

Position P_T of the cell with a maximum value must be chosen as the approaching target position to which the robot directly moves.

- **Proactive adjustment of spatial formation**

When neither the *visitor* nor the *robot* is participating in the conversation, the *robot* should approach the *visitor* first. Through our observations we found that the *host* tended to approach the *visitor* while considering whether he had a *subsequent plan* (29/32 trials in shop scenario, 18/26 trials in meeting scenario, as shown in Table 3.1). Since at this time the *robot* has the freedom to choose the location, we define this approach as the *proactive adjustment of spatial formation,* which has two rules (Figure 3.13):

1) When the *robot* only needs to greet the human without referencing an object or a place (*without plan*), it can simply go to the *visitor's front zone* when approaching from the front. Otherwise, it needs to enter the *visitor's sight zone* and keep a certain distance (1.1-1.5 m).

We defined *proactive adjustment* in the *without plan* cases by Eq. 6:

$$Value_{ProactiveAdjustment,Withoutplan}(P_H, P_R, P_T, \theta_H) = \begin{cases} 1/Dist(P_R, P_T) \quad (Dist(P_R, P_T) < 2000 \text{ and } |Angle(\theta_{P_H,P_R}, \theta_H)| < 60\deg) \\ \text{or } (1100 < Dist(P_R, P_H) < 1500 \text{ and } |Angle(\theta_{P_H,P_R}, \theta_H)| > 60\deg) \\ 0 \quad (otherwise) \end{cases} \tag{6}$$

where all of the parameters are in the x-y coordinate.

Position P_T with maximum value must be chosen as the approaching-target position.

2) When the robot needs to introduce some objects or places (*with plan*), it should choose the greet position that will keep the target object (or direction) visible to both the *visitor* and itself after the conversation has started. In this paper, we set this target in the field of view (270° from our observations) of both the *visitor* and the robot.

We defined *Proactive adjustment* in the *with plan* cases by Eq. 7:

$$Value_{ProactiveAdjustment,Withplan}(P_H, P_R, P_T, \theta_H, \theta_R) = \begin{cases} Value_{ProactiveAdjustment,Withoutplan}(P_H, P_R, P_T, \theta_H) \\ |Angle(\theta_H, \theta_{P_H,P_O})| < 135\deg \text{ and } |Angle(\theta_R, \theta_{P_R,P_O})| < 130\deg \\ 0 \quad (otherwise) \end{cases} \tag{7}$$

where P_O is the position of the target object. θ_{RB} is the robot's body orientation. All of the parameters here are in the x-y coordinate.

Position P_T with maximum value must be chosen as the approaching-target position.

3.2.5. Utterance and Gesture Control

We controlled the robot's utterances with a simple utterance controller that manages four functions: greeting, drawing attention, guiding, and explaining. A human developer pre-wrote the sentences, and the robot automatically uses them based on information from the state controller. The robot greets visitors when both of their *participation states* are *participating* and draws attention when only the *visitor* is not *participating*. When both are participating in the conversation, if the *visitor* is paying attention to the target product, the robot first explains it or guides the visitor to it.

The gesture controller accepts two types of input. One is from the state. When the state is *participating*, this controller makes the robot maintain eye contact or joint attention with the visitor. As the other type, it also receives input from the *utterance controller* to synchronize pointing gestures with utterances.

3.3. EXPERIMENT

We conducted an experiment that included both *system* and *user evaluations* to verify that our proposed model is useful for a robot to initiate conversation. From the viewpoint of our model, the two scenarios share the same patterns for initiating conversation, and thus either of them would be sufficient for this evaluation. In the shop scenario, the environment and the situation were more complex than that in the meeting scenario, making it possible to test the model with more varied situations. Accordingly, we decided to use the shop scenario as our evaluation experiment. The experiment was conducted in a lab room, under the assumption that it was a small computer shop with three products (Figure 1.1). A visitor visits this shop by appointment with a sales-robot to receive an explanation of one of the products. When he visits the shop, he waits for the sales-robot. When the robot arrives, they meet and initiate conversation. Finally, the robot explains the product. This setting places the focus of the evaluation on the interaction for initiating conversation. As we explained in Section 3.1.2.5 and Section 3.2.3, the parameters of the model we used in this experiment may need to be adjusted when using it in some other situations and environments. However, the knowledge of *participation zone* and *initiation of conversation* remains the same. The aim of the experiments is to investigate the validity of an initiation model that considers such regular patterns rather than the specific situation-dependent parameters. In this regard, we believe that using this simplified typical shop scenario is sufficient to show the effectiveness of the model.

3.3.1. HYPOTHESIS AND PREDICTION

From our observations, we found that people's behaviors during the *initiation of conversation* are influenced by such factors as the interlocutor's *participation state*. Therefore, we developed this hypothesis:

Hypothesis: Robot implemented with the *participation state* models would provide a better impression of interaction behaviors and make the participants prefer it better than robots that not implemented with the *participation state* models.

When using the proposed model, we assume that the robot can maintain its *participation state* effectively by adjusting its positions and timings as it greets and explains things to participants. We use *appropriateness of the standing position when the robot greets the visitor*, and *appropriateness of the standing position when the robot explains the target product* to evaluate the robot's behavior in the conversation. On the contrary, a robot using alternative methods that fail to consider the *participation state* might fail to adjust these positions and timings. Therefore, our hypothesis argues that if a robot considers the constraints for maintaining the participation state, as our proposal does, it can provide better impressions than alternative methods.

For comparison, we prepared two alternative methods: *guide* and *best-location*. The former method makes the robot initiate the conversation as quickly as possible by approaching a target within a certain distance. The latter method makes the robot stand at an appropriate location for explaining a product as quickly as possible before initiating the conversation. The details of the alternative methods are described in Section 3.3.2. Based on the above idea, we made this prediction:

Prediction 1: The proposed model for initiating conversation will outperform the alternative methods in the following areas: feeling of *appropriateness of the standing position when the robot greets the visitor*, *appropriateness of the standing position when the robot explains the target product*, and *overall evaluation*.

In the data collection, the timing of the first utterances by people to initiate conversations depended on situations such as visibility; for example, they start to greet when the target notices them even if the distance between them seems far ("Greet immediately" case in Table 3.2, 16/34 trials in shop scenario, 42/53 trials in meeting scenario), although they approached before greeting when the target did not notice them. The proposed model considers such visibility to control the robot behaviors. If we successfully implement our ideas, our proposed method will make the robot behave as people do. On the contrary, the alternative methods that fail to consider such visibility will require more time to prepare the robot to speak first because they only consider the positions of the robot and the target, not visibility, in initiating conversation. This means that the robot would not greet the participant even the participant had already paid attention to it until it gets to a position closer to the participant. This may make the participant wait for the robot, which can obviously be seen as a waste of time. This may influence the participant's impression on the robot's appropriateness of greet position. Accordingly, we predict that:

Prediction 2: Our proposed model of initiating conversation will decrease the *time from the beginning to the first utterance* compared to the alternative methods.

In the data collection, the timing of explaining or guiding also depends on the situation; they started explaining or guiding after approaching the target if they were far away. The proposed model considers such spatial settings to control the robot behaviors. If we successfully implement our ideas, our proposed method will make the robot behave as people do. On the contrary, the alternative methods will create different spatial settings, so the explaining or guiding timing will be different. In the *best-location* method, since the robot speaks first after reaching the proper position for explaining the product, we predict that such timing will closely follow the timing of the greeting. On the other hand, in the *guide* method, since the robot speaks first after reaching the target, sometimes the greeting position is far from the proper position for explaining the product. Such timing will be far from acceptable greeting timing. Thus, this time can partially and indirectly indicate the appropriateness of the choice of the explaining position and may influence the participant's impression of the robot. We predict that:

Prediction 3: The proposed model of initiating conversation will decrease the *time from the end of greetings to explanations* compared to the alternative methods.

If predictions 2 and 3 hold, the total interaction time with the robot that uses the proposed model will be less than the total interaction time with the robots that use the alternative methods. Based on these two predictions, we further predict that:

Prediction 4: The proposed model for initiating conversation will decrease the *total time* compared to the alternative methods.

3.3.2. CONDITIONS

The proposed model is compared with two alternative methods, which do not use the knowledge proposed in the paper but exploit other existing knowledge to provide the best interaction in the scenario.

a) Proposed method (***proposed***): The robot behaves based on our proposed model. It first approaches the visitor while considering the *subsequent plan*, and initiate conversation with the visitor at the proper timing according to the participation state of both of itself and the visitor. The robot would then judge if it is necessary to guide the visitor to pay attention to the target product by analyzing the visitor's focus of attention and then behave accordingly. At last, it explains the target product to the visitor from a proper position and orientation.

b) Always greet and guide (***guide***): In this strategy, although the robot does not have a complicated model for conversation initiation, it behaves as politely as possible and initiates the conversation as quickly as possible. It first goes directly toward the visitor. When the distance between them is reduced to 2 m, the robot stops, greets the visitor, and asks the visitor to look at the product. As the visitor approaches the product and looks at it, the robot goes to the best location for explaining the product, i.e., the location based on O-space, and explains it.

c) Always start the interaction at the best location for explaining (***best-location***): In this strategy, the interaction is designed to be as simple and quick as possible. When the robot finds a visitor, it immediately stands at an appropriate location for explaining the product, i.e., the location based on O-space, and starts to talk.

In the *guide* and *best-location* conditions, we used a previous model [37] in which the robot chooses a position near the human and the product, while keeping the product visible to both the robot and the human. We use the following model for the robot to appropriately control its position:

$$Value_{Bestlocation}(P_H, P_R, P_O, P_T, \theta_H) = \begin{cases} 1/Dist(P_R, P_T) & (1100 \le Dist(P_R, P_H) < 1300 \text{ or } 1100 \le Dist(P_R, P_O) \le 1200 \\ & \text{and } |Angle(\theta_H, \theta_{P_H, P_O})| < 90 \deg \text{ and } |Angle(\theta_R, \theta_{P_R, P_O})| < 150 \deg \\ 0 & (otherwise) \end{cases} \tag{8}$$

In advance, the experimenter wrote the text for the robot's utterances in five categories: (1) drawing attention, (2) greeting, (3) guiding, (4) explaining, and (5) epilog. In the *guide* and *best-location* conditions, the robot says the texts from the *greeting*, *guiding*, *explaining,* and *epilog* categories. In our *proposed* method, the robot always says the texts in the *greeting, explaining,* and *epilog* categories because the decision to say the texts in *drawing attention* and *guiding* are dependent on the *visitor's participating state* and *focus of attention*. If the *visitor* is participating in a conversation with the robot (focusing attention on the target product), the robot doesn't say the texts in the *drawing attention* (*guide*) category. Otherwise, it says those texts.

The exact utterances the robot spoke are as follows:

Drawing attention: Excuse me.

Greeting: Welcome, my name is Robovie and I'm in charge of PC sales. (Welcome would be omitted when the robot perform drawing attention first)

Guiding: We have got a new laptop PC over there, please just take a look.

Explaining: Let me show you this laptop PC. We just got it last week, and it is very popular now. The memory of this PC is 4GB, and its battery life is about 6 hours. In addition, the price is 100,000 yen normally but it is now on a campaign and only cost 80,000 yen.

Epilog: The introduction of this PC is over. Please just look around in our store at pleasure.

3.3.3. PROCEDURE

Fifteen native Japanese-speaking people (seven men, eight women, average age: 27 +/- 11, range from 18 to 56) were paid for their participation in our experiment that was conducted in a 6 x 10-m room. Due to the visibility limitations of the motion capture system, the experiment area was restricted to a 3 x 4.5-m area. We used the robot and motion capture system described in Section 3.2.2.

First, the participants put on the markers of the motion capture system, which was then calibrated by the experimenter. Then, the scenario and instructions were provided to the participants, instructing them to evaluate the interaction of the robot from the standpoint of a shop owner who needed to choose one robot from the candidates. They played a visitor in various ways so that they could completely judge the appropriateness of the behavior of each robot. They evaluated three types of robots from the shop owner's perspective to let them judge various spatial formations for initiating conversations, since each method has strengths and weaknesses.

They simulated the behaviors of five types of visitors that decided all by themselves (as a result, the five types of visitors played by each participants are not all the same), such as someone waiting in front of the product or someone at the store entrance, and interacted five times under each condition. In each condition, after interacting with the robot five times and pretending to be a different visitor in each interaction, they filled out a questionnaire to rate their impressions. The experiment used a within-subject design and the order of conditions was counterbalanced.

The experiments were recorded on video together with the motion capture system (recording the coordinates of the markers). In addition, the recognition results of the states and the detailed parameters such as positions, distances and angles of both the robot and the participant were also recorded by the robot system every 100 ms.

3.3.4. MEASUREMENT

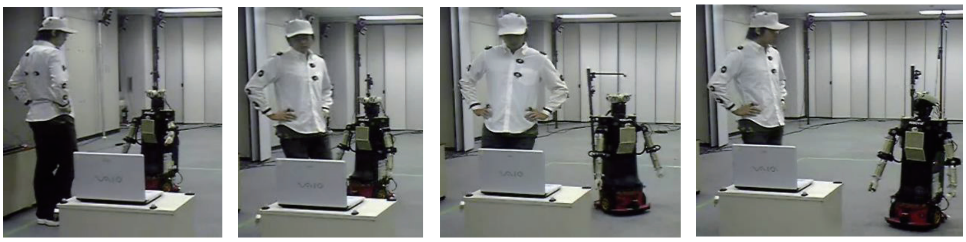
3.3.4.1. System Evaluation

First, we confirmed the *recognition accuracy of the participation state* of our system for both the robot and the visitor using the recorded experimental data. The system recorded all of the participation states of both the robot and the visitors in each trial. Thus, the *joint participation states* were also recorded. To confirm whether the recognition of the *joint participation state* was correct, two coders classified the *joint participation states* into the four state variables explained in Section 3.2.3 for all of the trials. The two coders that analyzed the data are two people who have no knowledge about robotics and HRI, but not the same people who coded the data collection (human observation experiment in Section 3.1) results. And they did not know about the purpose of the data and the model proposed in our research. We then compared the coding and system recognition results.

Second, we confirmed the *appropriateness of the robot's initiating behavior*. Based on the *joint participation state*, the robot moved and spoke first to the visitor in each trial. Since the robot spoke first, its visitor quickly realized that the robot wanted to talk to him/her and thus listened to the robot. Here, it is important to determine whether the robot spoke first at the proper position and timing.

We asked the two coders who classified all 75 trials whether the robot spoke first to the visitor at the proper position and timing. For each trial, they classified the position and timing at which the robot first spoke into two cases: *proper* and *improper*.

Third, we evaluated whether *maintaining of the participation state* was achieved. As discussed above, after starting the conversation, the robot should continue it until the end of its presentation. However, sometimes the visitor moved to another place, disrupting the conversation. For example, the robot showed the visitor the product (Figure 3.14-a in *joint participation state* PSJ = (1, 1)), but then the visitor moved toward the target product and disrupted the conversation (Figure 3.14-b, PSJ = (0, 0)). In this case, the robot must reposition itself to adjust the spatial formation (Figure 3.14-c) so that both are participating in the conversation again (Figure 3.14-d, PSJ = (1, 1)).



(a) (b) (c) (d)

Figure 3.14 Discontinuing and re-establishing the conversation

We used the coding results for the participation state to determine whether the conversation was discontinued in each trial. The coders again classified whether the conversation was disrupted by the robot or the visitor. We also calculated how long it took for the robot to re-establish the conversation with its visitor.

3.3.4.2. User Evaluation

The user evaluations included both subjective and objective assessments.

- **Subjective evaluation**

Participants completed a questionnaire for each condition after five interactions on a simple Likert scale of 1 to 7 that higher ratings are considered to be better. The questionnaire had the following items: *appropriateness of the standing position when the robot greeted the visitor*, *appropriateness of the standing position when the robot explained the target product*, and *overall evaluation*.

- **Objective evaluation**

In addition to the questionnaire, we focused on the following timings: (1) How much time does the robot take to initiate the conversation with the *visitor*? (2) After greeting, how much time does the robot take to prepare to explain the product? (3) How much time does the robot take to complete the entire scenario? The system recorded the *time from the beginning to the first utterance,* which is the time from the beginning of the experiment (the time of starting the robot system) to the time when the robot says the first word to the participant, the *time from the end of the greeting to the explanations,* which is the time from the end of the *greeting* utterance to the start of the *explanation* utterance, and the *total time,* which is the time cost in a whole trial.

3.3.5. RESULT OF SYSTEM EVALUATION

3.3.5.1. RECOGNITION ACCURACY OF PARTICIPATION STATE

Cohen's Kappa coefficient from the two coders' classification was 0.83, indicating highly consistent classification results. After the classification, to analyze the consistent trajectories for modeling, the two coders discussed and reached a consensus on their classification results for the entire coding process. Then we compared their coding results with the system recognition results. We compared the recognition result of the system and the coding result of the coders and recorded the time of the two result matches as T_{right}. Accordingly, we define the rate of system accuracy as

$$RecognitionAccuracy = T_{right} / T_{entire} \tag{9}$$

The system's recognition accuracy of the *joint participation state* was 90.2% of the coder's coding results, proving that with our system, the robot can accurately recognize its relationship with its visitor.

We analyzed the 10% difference and found that the system correctly recognized the changes in the participation state; the only difference was the timing of the changes (Figure 3.15). In the two results, the changing of the *joint participation state* was the same, e.g., from (0, 0) to (1, 1). As the *joint participation state* changes, the timings of the changes in the two results were sometimes different. We calculated the difference in the time from its occurrence to its end, and the average was 1.210 +/- 0.399 sec (range from 0.067 to1.747 sec).

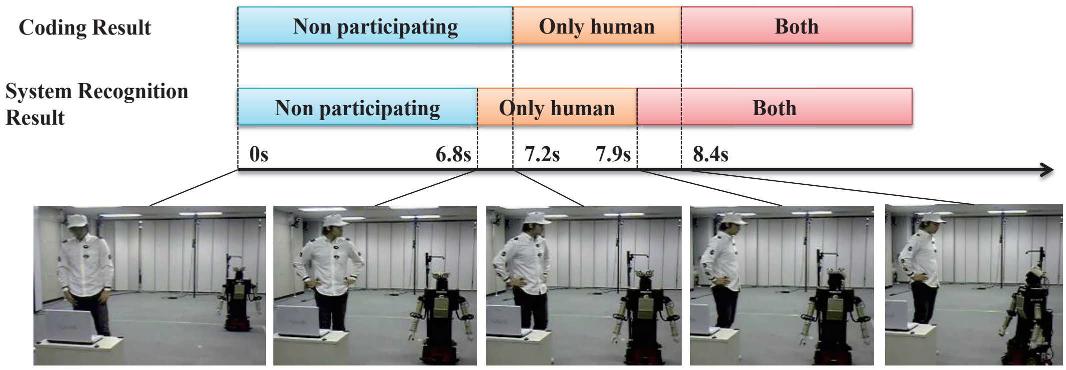
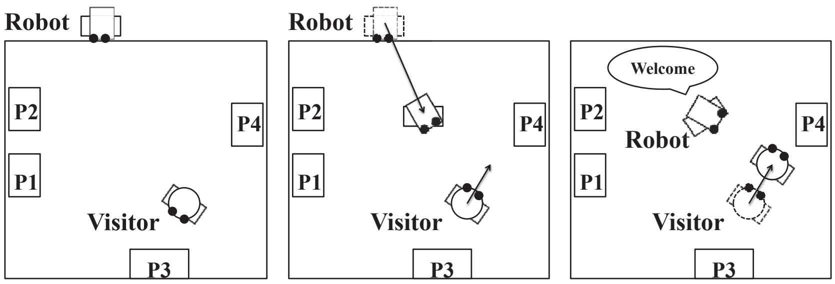


Figure 3.15 Difference between coding and system recognition result

3.3.5.2. *APPROPRIATENESS OF ROBOT'S INITIATING BEHAVIOR*

Cohen's Kappa coefficient from the two coders' classification was 0.91, indicating that their classification results were highly consistent. After the classification, the two coders discussed and reached a consensus on their classification results. Their coding result shows that in 69 trials (92.1%), the robot behaved appropriately.



(a) From non-front (b) moment of both participating (c) greeting the visitor

Figure 3.16 Inappropriate cases of robot's initiating behavior

In the six trials in which they thought the robot failed to behave appropriately, the robot first approached from the non-frontal direction (Figure 3.16-a). As the robot came nearer, the visitor suddenly turned around and passed and ignored it (*unnoticed*). There was a moment during which both the robot and the visitor were in each other's *frontal zone* (Figure 3.16-b). However, since the visitor moved very quickly, there was a system delay before the robot spoke. When the robot finally greeted the visitor, it was a little too late (Figure 3.16-c).

3.3.5.3. *Maintaining the participation state*

The classification results of the two coders for the participation state showed that in 62 of 75 trials the conversation was disrupted, i.e., the *joint participation state* PS_J changed from (1, 1) to (0, 0). The coders also classified whether the conversation was disrupted by the robot or by the visitor. The coding results of the two coders were identical, showing that in all 62 trials, the visitor moved and interrupted the conversation.

When its visitor moves, the robot should follow him/her to readjust the spatial formation and thus re-establish the conversation as soon as the visitor stops. We calculated the time from when the visitor stopped to when both the robot and the visitor began to participate in the conversation again. The average of this re-establishing time was 4.613 +/- 1.267 sec (range from 1.500 to 9.800 sec).

3.3.6. Result of User Evaluation

We used a Shapiro-Wilk test to preliminary analyze the experiment data, and confirmed that each set of data is normally distributed ($p > .05$ in all the data sets) before conducting further analysis.

3.3.6.1. *Verification of prediction 1*

Our first prediction was that the proposed model for initiating conversation will outperform the alternative methods in the following areas: feeling of *appropriateness of the standing position when the robot greets the visitor*, *appropriateness of the standing position when the robot explains the target product*, and *overall evaluation*.

For the "overall evaluation" score (Figure 3.17), we conducted a repeated measures ANOVA and found a significant main effect ($F(2,28)=9.125$, $p=.001$, partial $\eta 2$

= .395). A multiple-comparison by the Bonferroni method revealed that the score for the *proposed* condition was significantly higher than that for both the *guide* (p=.021) and *best-location* (p=.002) conditions. No significant difference was found between the *guide* and *best-location* conditions (p=.5). Therefore, our first prediction was supported.

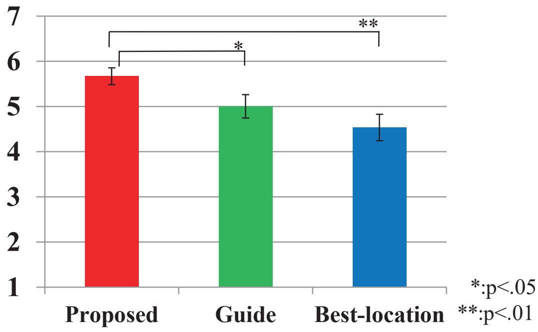


Figure 3.17 Overall evaluation

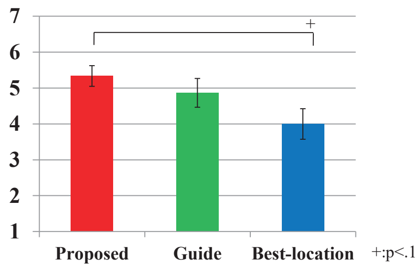


Figure 3.18 Appropriateness of standing position when it greeted

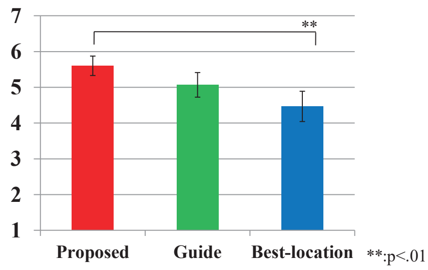


Figure 3.19 Appropriateness of standing position when it explained

For "appropriateness of standing position when it greeted" (Figure 3.18), a repeated measures analysis of variance revealed a significant main effect (F(2,28)=4.697, p=.017, partial η2=.251), but a multiple-comparison by the Bonferroni method showed only non-significant differences (*proposed* vs. *guide*: p=.706*, proposed* vs. *best-location*: p=.058, and *guide* vs. *best-location*: p=.199).

For "appropriateness of standing position when it explains the target product" (Figure 3.19), a repeated measures analysis of variance revealed a significant main effect (F(2,28)=9.126, p=.001, partial η2=.395). The Bonferroni method showed a significant difference between the *proposed* and *best-location* methods (p=.003), but other comparisons were not significant (*proposed* vs. *guide*: p=.209 and *guide* vs. *best-location*: p=.111).

3.3.6.2. Verification of Prediction 2

Our second prediction was that our proposed model of initiating conversation will decrease the *time from the beginning to the first utterance* ($T_{initiate}$) compared to the alternative methods.

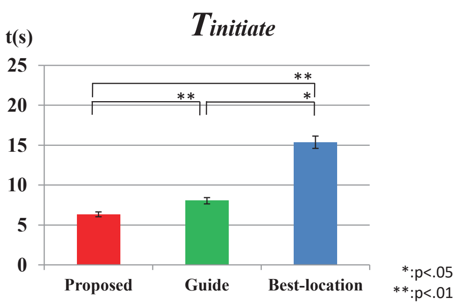


Figure 3.20 Average of $T_{initiate}$

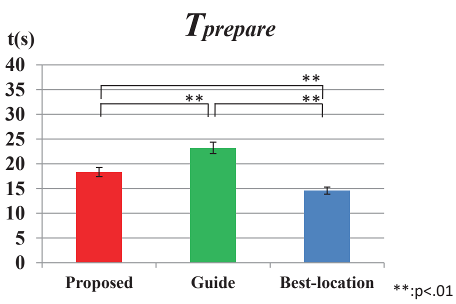


Figure 3.21 Average of $T_{prepare}$

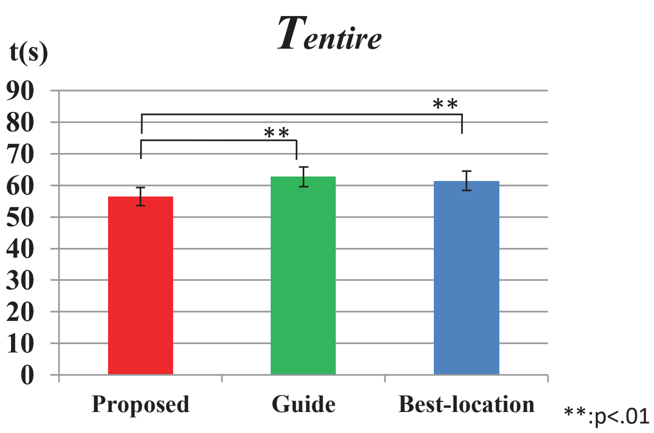


Figure 3.22 Average of T_{entire}

For our second prediction (Figure 3.20), $T_{initiate}$ averaged 6.333 sec in the proposed condition, 8.037 sec in the *guide* condition, and 15.363 sec in the *best-location* condition. We conducted a repeated measures ANOVA and found a significant main effect ($F(2,148)=108.252$, $p<0.001$, partial $\eta2$ = .594). A multiple-comparison by the

Bonferroni method revealed that the $T_{initiate}$ of the *proposed* condition was significantly less than that of both the *guide* ($p<.001$) and *best-location* ($p<.001$) conditions and that it was significantly less for the *guide* condition than for the *best-location* condition ($p=.021$). Thus, our second prediction was supported.

3.3.6.3. VERIFICATION OF PREDICTION 3

Our third prediction was that the proposed model of initiating conversation will decrease the *time from the end of greetings to explanations* ($T_{prepare}$) compared to the alternative methods.

For our third prediction (Figure 3.21), $T_{prepare}$ averaged 18.345 sec in the proposed condition, 23.209 sec in the *guide* condition, and 14.568 sec in the *best-location* condition. We conducted a repeated measures ANOVA and found a significant main effect ($F(2,148)=38.160$, $p<0.001$, partial $\eta2 = .340$). A multiple-comparison by the Bonferroni method revealed that the $T_{initiate}$ levels of both the *proposed* and *best-location* conditions were significantly less than that of the *guide* ($p<.001$) condition and that it was significantly less for the *best-location* condition than for the *proposed* condition ($p=.001$). Thus, our third prediction was partially supported.

3.3.6.4. VERIFICATION OF PREDICTION 4

Our fourth prediction was that the proposed model for initiating conversation will decrease the *total time* (T_{entire}) compared to the alternative methods.

For our fourth prediction (Figure 3.22), T_{entire} averaged 56.459 sec in the proposed condition, 62.747 sec in the *guide* condition, and 61.431 sec in the *best-location* condition. We conducted a repeated measures ANOVA and found a significant main effect ($F(2,148)=22.464$, $p<0.001$, partial $\eta2 = .233$). A multiple-comparison by the Bonferroni method revealed that the T_{entire} of the *proposed* condition was significantly less than both the *guide* ($p<.001$) and *best-location* ($p<.001$) conditions. But the comparison between *guide* and *best-location* was not significant (p=.708). Thus, our fourth prediction was supported.

3.3.6.5. SUMMARY

In summary, our *proposed* system was evaluated as the best method overall among those compared. Its effect in the overall evaluation can partially be explained by the difference between the *proposed* and *best-location* conditions in the appropriateness of the standing position when the robot explained the target product. However, this does not account for the difference between the *proposed* and *guide* conditions. And for the appropriateness of the standing position when the robot greeted the participant, only an almost significant result (p=.058) between the *proposed* and *best-location* conditions could be found. The results for *Tinitiate*, *Tprepare*, and *Tentire* show that with the *proposed* system, a robot can initiate conversation much more quickly than in the other two conditions and that, moreover, the *guide* condition outperforms the *best-location* condition. In addition, using the *proposed* system the robot completed the interaction with the *visitor* much more quickly than with the other two methods. This may also partially explain the results of the overall evaluation. We consider that prompt reaction behaviors from the robot, depending on the participation state, have a strong positive impact on an interaction.

Thus, our proposed model was evaluated as the best approach.

3.4. Discussion

3.4.1. When will this capability be used?

We believe that the capability of a robot to naturally initiate conversation is a major function to be implemented in future social robots. Although many other research projects have assumed that people and robots have already met and started interaction, this is generally not the case in the real world. Perhaps at an early deployment phase robots might not need to initiate interaction by themselves, since people would be interested in their novelty and approach them. In such cases, robots do not need to deal with the constraints of spatial configuration in order to initiate interaction.

However, when robots actually do start to work in the real world without attracting so much attention, people will often not initiate interaction by themselves. In such cases, robots will often fail to initiate interaction 35]. This problem will be more serious when the robot has a concrete role, e.g., shopkeeper. The shopkeeper scenario used in this study is one future situation where a robot is expected to play such a role. There are many other situations that involve a first meeting, such as a tour guide in a museum, a shopping assistant, and nursing care in a hospital, all of which have been considered applications of social robots in past research.

In our observations, we have found that the front and sight zones were stable in different environments and situations. That means it is possible to use these models for social robots working in many situations, as mentioned above. As for the gaze zone, although its parameters are dependent on the environment, we can also easily use it by first identifying the proper parameters for each situation.

3.4.2. Limitations

First, in our experiment, there was only one visitor in the shop, while in a real shop there might be multiple customers at any given time. A greater number of people in the environment would create several difficulties, such as obstacles for a moving robot, determining the target visitor among several people, and interruptions by other visitor when the robot is approaching the target visitor. In this paper, we did not provide models to solve this problem. This is certainly a limitation of our model. However, it would be possible to extend our model by adding several functions provided by other researchers. For example, when a person becomes an obstacle for the robot that is approaching a target visitor, the robot would be able to avoid the person by simply using a path-planning or collision-avoidance mechanism. In such a situation, the robot might need to keep a distance from other persons as it talks to the target visitor. Overcoming such limitations would be necessary before adapting our system to more crowded situations. As for decision making, we need to create a high-layer controller to find the appropriate target among people. This is out of this research's scope, but some past research works such as estimating visitors' state would be useful for this kind of mechanism. How to deal with an interruption by other visitors would depend on the robot's applications; if the robot is working as a shop employee, it would be better to change the target to the person and immediately start conversation. If the robot is working in a special service such as welcoming a VIP, the robot should not change the target. Actually, in the meeting scenario, when the host started approaching the visitor, staff members of the research institute sometimes walked through the lobby and passed by. As future work, by further analyzing these data or conducting additional experiments, we could create such a high-layer controller to help the robot make decisions when multiple people are in the situation.

Second, the decoration of our shop is very simple and there were only three products arranged separately. Such situations are commonly found in the real world. For example, when a robot works as a staff member in a gallery to explain individual artworks hanging on the wall to the visitors, our model can help the robot to recognize the focus of the visitor's attention correctly. On the other hand, there are certainly many environments in which several objects exist within the view of a single visitor simultaneously. For example, in a real shop, the goods might be placed more compactly, e.g., four laptop PCs on the same desk or dozens of displays hanging on the wall close to each other. In this case, there might be multiple products in a person's transactional segment simultaneously, making the detection of the person's focus of attention more complicated. We believe it is not always necessary for the robot to recognize the one specific object the *visitor* is looking at; recognizing the aggregation that the *visitor* is paying attention to is enough for the robot to provide basic service. Actually, in our daily life, in many cases it is not necessary for the clerk to know the customer's focus of attention at such a precise level. Based on our model, the use of gaze detection would help the robot to further improve the recognition accuracy of the visitor's focus of attention. Even if stable gaze detection is still difficult, such a function enables the robot to limit the candidates of objects to which the person pays attention. For example, with such a function the robot might be able to recognize whether the visitor is paying attention to the apples or oranges, and this could help the robot to provide services more appropriately.

Third, since our proposed model was tested in a specific scenario, its generalizability is limited. Perhaps the context affects the preferences for a robot's behavior. For example, in a busy business scenario, the *always starting interaction at the best location to explain* condition might work better than the proposed model. We believe that our shopkeeper scenario is rather neutral, so it probably reflects interaction in many daily scenarios, but this needs verification.

As we mentioned in the paper, the parameters in our model dealt with Japanese people and our own robots. But when they are adapted, adaptation parameters must be considered. For instance, factors such as cultures, type of robots and environment would influence parameters.

One may need to adjust the parameters when using robot for people from other cultures. For example, when the model is to be used in the countries such as The Netherlands and Denmark, the average height of people is much taller than Japan. John et al., suggested that height is a significant determinant of personal space [82], thus we consider that distance parameters retrieved from our study might need to be adjusted when using to interact with people of significant different height to make sure the interlocutors feel comfortable.

We only evaluated the model with our own humanoid robot, while others may use other type robots to interact with people. Different appearance could influence people' feeling and attitudes towards the robots [83, 84]. It is proper to imagine that a robot which has a lovely appearance of a famous cartoon character such as Mickey Mouse might easily attract many people to interact with it with joy. While a robot with a horrible appearance might sometimes frighten some people or make them feel uncomfortable. We suppose that these different feelings and attitudes caused by the different appearance of robots might also influence some parameters of the model. For example, we expect that it might be better to set the talking distance parameter for a horrible robot bigger than that for a lovely robot, but more evidences are required when one consider adjusting parameters.

The environment might also have influence on the parameters that used in the models. For example, when using the models in environments that everyone need to keep quiet, such as in a museum, library or a gallery, even there are not many people around, apparently it is not proper for the robot to greet a person from a long distance. It would also be better to reduce the distances in the models so that the robot could greet and then talk to other people in a low voice.

4. INITIATION OF MOTION INTERACTION: DISTRIBUTIONAL HANDING

4.1. MODELING DISTRIBUTION BEHAVIOR

In many countries, clerks from shops or companies distribute flyers, coupons, or pamphlets in shopping malls, museums, or on outdoor streets. The giver initiate interaction with passersby with his/her handing motion. We modeled the distributing behavior based on our observations of the handing behavior of people who distribute flyers in real environments.

People distribute things in various ways. For instance, some passively wait for pedestrians to take a flyer, and others actively approach pedestrians and offer them flyers. These different behaviors might have different effects on persuading pedestrians to accept the objects. In fact, we found that the success rate of the givers' handing performance widely diverged between 12.5% and 77.5%. Our main goal is to identify effective distributing behavior.

4.1.1. Data Collection

We collected scenes of flyer-distributing behaviors (a person offers flyers to pedestrians) by observations in a shopping mall in Osaka, Japan. The area includes a 3-6 m wide corridor that is approximately 70 m long with four shops nearby and one big hall about 300 m^2 that connects to a corridor that links the hall to a busy train station.

We collected video and position data of all the people in the area (video data for our observation and position data for calculating the detailed parameters). Both the corridor and the hall were covered with our people-tracking infrastructure using 49 3D-range sensors attached to the ceiling (a combination of Panasonic D-Imager, ASUS Xtion, and Velodyne HDL-32E) to estimate pedestrian locations every 33 ms [85].

Among the data collected over one year [86], we analyzed the pedestrian data from 10 am to 8 pm on six Sundays. We manually searched for scenes where givers distributed flyers to pedestrians. We identified ten givers who distributed more than 40 times and analyzed the first 40 distribution behaviors of each giver.

4.1.2. Analysis of Distributional behavior

4.1.2.1. How did these givers approach their targets and hand over their items

To determine an effective *distributional handing* method, we analyzed and evaluated the different methods exhibited by the givers. We categorized their behaviors with a focus on three behavioral elements: *gaze*, *approach,* and *arm* motions. All the givers' *gaze* behaviors were similar. When a pedestrian was chosen as the handing target, the givers kept *gazing* at the pedestrian to maintain eye contact until the handing was finished.

On the other hand, we found differences in the *approach* and *arm* motions. We categorized the distributional handing of the givers into the following four patterns:

(a) (b) (c)

(1) Extend arm first and wait for pedestrian

(a) (b) (c)

(2) Wait for pedestrian and extend arm nearby

(a) (b) (c)

(3) Extend arm first and approach pedestrian

(a) (b) (c)

(4) Approach pedestrian and extend arm nearby

Figure 4.1 Four types of *distributional handing behaviors*

- **Extend arm first and wait for pedestrians**: The giver stayed at a certain place to wait for the arrival of pedestrians. She noticed a pedestrian, gazed at him (Figure 4.1-a), and then began to extend her arm to hand him a flyer (Figure 4.1-1-b). When she fully extended her arm and completed the handing motion, the pedestrian remained slightly away from her (Figure 4.1-1-c).

- **Wait for pedestrian and extend arm nearby**: The giver waited for the pedestrians to arrive while continuing to hold the flyer at waist-height (Figure 4.1-2-a). She did not start to extend her arm until the pedestrian came close (Figure 4.1-2-b). When she completed her arm-extending motion, the distance between the giver and the pedestrian was acceptable for the pedestrian to take the flyer (Figure 4.1-2-c).
- **Extend arm first and approach pedestrian**: The giver waited for pedestrians and noticed a pair of them (Figure 4.1-3-a). After choosing the pedestrian in black as her handing target, she extended her hand that held the flyer and approached the pedestrian (Figure 4.1-3-b). Instead of just moving to the side of the pedestrian and waiting, she kept approaching the pedestrian. The giver did not stop walking until she reached a place at which the pedestrian could easily accept the flyer.
- **Approach pedestrian and extend arm nearby**: The giver noticed a pedestrian approaching from the right side and moved toward him (Figure 4.1-4-a). As the distance between the pedestrian and the giver shrunk, the giver started to extend her arm to distribute the flyer while simultaneously approaching him (Figure 4.1-4-b). Finally, the giver simultaneously stopped near the pedestrian and completed her arm-extending motion.

Based on our observations, two coders who were not informed about our research hypothesis analyzed the collected data and separately classified all 400 handing trials from the ten givers. Cohen’s kappa coefficient from the two coder’s classifications was 0.873, indicating that their evaluations were highly consistent. They discussed disagreements to reach a consensus about their classification results.

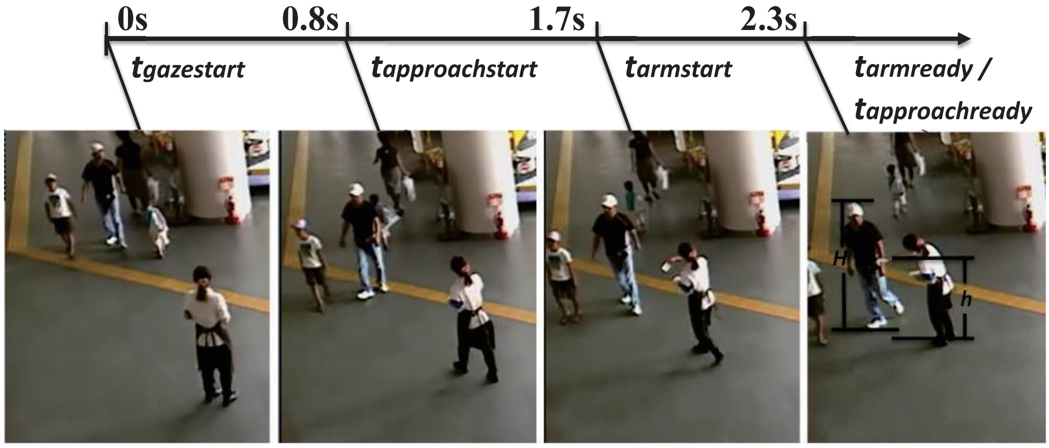
Table 4.1 Successful ratios based on behavior type

Behavior type	Successful ratio
Extend arm first and wait for pedestrian	21.2% (21/99)
Wait for pedestrian and extend arm nearby	33.0% (58/176)
Extend arm first and approach pedestrian	25.0% (5/20)
Approach pedestrian and extend arm nearby	72.4% (76/105)
Total	40.0% (160/400)

The coding result is shown in Table 4.1. The distribution style largely influences the *successful ratios*. The givers achieved the highest *successful ratio* when they distributed in the *approach pedestrian and extend arm nearby* type (72.4%). Therefore, we modeled close *approach and handing* behaviors for *distributional handing*.

4.1.2.2. DETAILED MODELING OF DISTRIBUTIONAL HANDING

To create *approach pedestrian and extend arm nearby* behavior, we further analyzed the details of each behavior and the timing.



(a) (b) (c) (d)

Figure 4.2 Details of behaviors and timing

Timing: Figure 4.2 shows a typical example of the timing of the *gaze*, *approaching*, and *arm* behaviors. As the giver looked around her environment, she started *gazing* when she chose a distribution target, then she started *approaching*, and finally she started an *arm* behavior when she got closer to the pedestrian. The most important constraint on the timing is that the giver completed her *arm* behavior (hereinafter $t_{armReady}$) just as she completed the *approaching* behavior ($t_{approachReady}$) (max error = 0.7s):

$$t_{armReady} = t_{approachReady}. \qquad (1)$$

Thus, we can compute the timing of *arm* behavior $t_{armStart}$ to satisfy this constraint after establishing an approach plan.

Approaching: We analyzed the detailed trajectories of the *approaching* behaviors of the givers. When offering objects, givers typically keep approaching pedestrians until they are close to the giver's front right/left side. In 12 out of 105 trials where the pedestrian directly approached the giver's initial position, the giver avoided the pedestrian's route and approached the pedestrian from the side (Figure 4.3). Overall, we can model a situation where a giver approaches pedestrians from the front left/right side but not directly from the front (Figure 4.1-4). As shown in Figure 4.4, we denote the giver position and the pedestrian when the giver stopped approaching (i.e., $t=t_{approachReady}$) as G_{ready} and P_{ready}. We decomposed the distance between P_{ready} and G_{ready} into *frontal* (element of distance in the direction of the pedestrian's motion) and *horizontal* (direction orthogonal to the pedestrian's motion). The location of G_{ready} can be computed using two parameters, $D_{frontal}$ and $D_{horizontal}$. To identify precise parameters, we conducted further analysis. Three givers performed *distributional handing* in our lab. They performed the *approach pedestrian and extend arm nearby* method. We collected the data with a motion capture system that tracks the position data in 100 Hz with error less than 2 mm. From the data, we computed $D_{horizontal}$ to be 0.7 m and $D_{frontal}$ to be 1.3 m and used these values for our system.

Figure 4.3 Avoid and handing

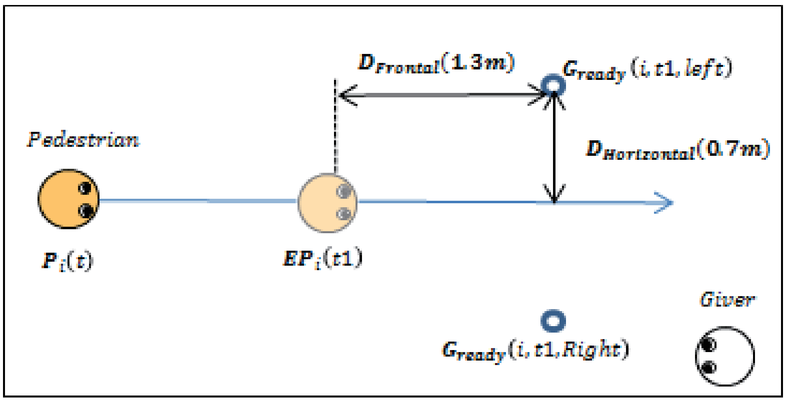


Figure 4.4 Candidates of G_{ready} at future moment t1

Arm: We further analyzed the data collected with a motion capture system. In the collected data, the givers typically held the flyer at their waist-height from the beginning to $t_{armStart}$. As she got closer to the pedestrian, she started to extend her arm at about the same height as the pedestrian's waist-height (defined as h, illustrated on the right in Figure 4.2-d). We found that the height of the giver's hand was adjusted based on the pedestrian's height (defined as H) so that a pedestrian can easily take the flyer. We set k as the ratio of the height of the giver's hand to the pedestrian's height and formulated the following constraint for the hand's height:

$$h = k \cdot H. \quad (2)$$

For our system, we used the average of k, which was 0.632 and ranged from 0.617 to 0.676.

4.1.3. INFLUENCE FROM PREVIOUS PEDESTRIANS

(a) (b) (c)

Figure 4.5 Continuous handing scene: after *pedestrian A* accepted a flyer, subsequent *pedestrians B* and *C* also took flyers.

We noticed that whether a pedestrian takes a flyer is influenced by the behavior of the previous pedestrians. That is, pedestrians tended to accept flyers if those who preceded them also took flyers. Figure 4.5 shows one such scene where pedestrians continuously accepted flyers from the giver. We scrutinized this phenomenon and categorized the giver's handing into two categories based on whether the handing trial is related to the other handing trials:

- **Individual**: Before the giver handed the flyer to the pedestrian, he had not noticed that the giver was handing out flyers to other pedestrians.

- **Former one not accepted**: Before the giver handed the flyer to the pedestrian, he had already noticed that the former pedestrian refused it from the giver.
- **Continuous** (former one accepted): Before the giver handed the flyer to the pedestrian, he had already noticed that the previous pedestrian accepted it from the giver.

Two coders analyzed the collected data and classified all the 400 handing trials analyzed in Section 3.1.2. Cohen's kappa coefficient from their classifications was 0.914, indicating that their evaluations were highly consistent. They discussed disagreements to reach a consensus in their classification results.

Table 4.2 Influence from previous pedestrians

Previous influence	Successful ratio
Individual	35.5% (55/155)
Former one not accepted	31.6% (42/133)
Continuous (former one accepted)	56.3% (63/112)
All	40.0% (160/400)

The result is shown in Table 4.2. The successful handing ratio was 35.5% (55/155) for *individual*, 31.6% (42/133) for *former one not accepted* and 56.3% (63/112) for the *continuous* type. A Chi-square test revealed that the *successful ratio* of *continuous* handing was significantly higher than that for both *individual* ($\chi2$ (1) = 10.542, p<.01, φ=0.198) and *former one not accepted* ($\chi2$ (1) = 14.120, p<.01, φ=0.239). This means that if a previous pedestrian took a flyer, the next pedestrian is also more likely to accept a flyer.

Since there were no significant differences between *individual* and *former one not accepted* ($\chi2$ (1) = 0.329, p=.566, φ=0.060), for simplicity, we treated these two as a single category: *independent distributing*. This means that regardless whether a target pedestrian saw that a previous pedestrian declined or just did not see anyone, a pedestrian would behave roughly the same.

In summary, this analysis informs our planning framework based on the following:

- When handing is successful to previous pedestrians, we anticipate a higher success rate among subsequent pedestrians.

- The robot should plan to continuously distribute to a series of pedestrians.

4.2. SYSTEM

4.2.1. ARCHITECTURE

Based on our analysis in Section 4.1, we learned the following: 1) our robot must try to continuously distribute to a series of pedestrians, and 2) it must approach them from the front and only extend its arm near the target pedestrian. Our system is designed to satisfy these requirements.

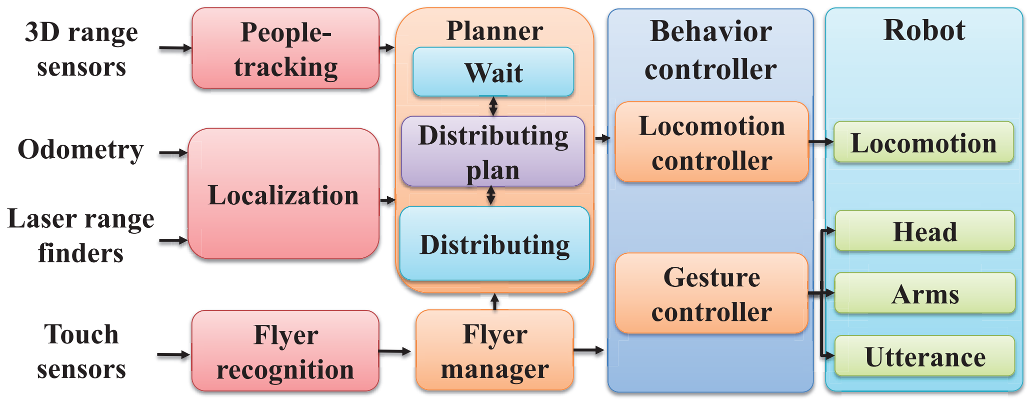


Figure 4.6 System overview

Figure 4.6 illustrates our software architecture. The main module is the *planner* in which the system plans paths for approaching pedestrians based on the above requirements. This planning is enabled by the information provided from the *people-tracking* and *localization* modules. The *flyer manager* controls the robot and prints flyers to be held in its hands.

Once a target is selected, precise timings are controlled by the *behavior controller* module that manages the locomotion and motion of the arms and the head direction of the robot. We explain these modules below.

4.2.2. HARDWARE AND BASIC INFRASTRUCTURES

4.2.2.1. ROBOT

We used a human-like robot, Robovie [87] (Figure 4.7), which is 1.2 m tall with a 0.3 m radius and is characterized by its human-like body expressions. It has a 3-DOF head and 4-DOF arms with 2-DOF hands. Its locomotion platform is a wheeled Pioneer3 DX. Two 30-m range laser sensors (Hokuyo UTM-30LX) were attached and used for localization and safety stop. It moved at a velocity of 500 mm/sec (1.8 km/h) forward and 45 degree/sec for rotations. Its forward and rotation accelerations are 400 mm/sec and 30 degree/sec, respectively.

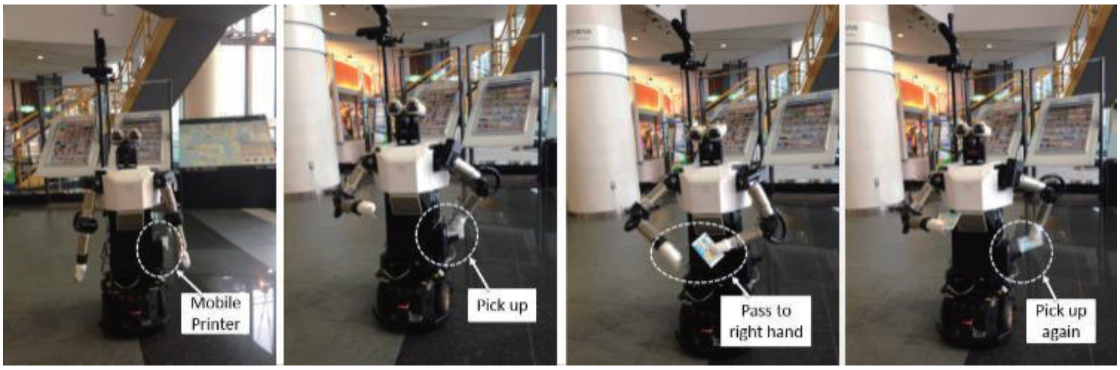


Figure 4.7 Robot picked up flyer from mobile printer

4.2.2.2. HAND AND FLYER MANAGER

The robot can hold a flyer in each of its hands. The *flyer-manager* module executes *print commands* and the robot *picks up* the flyer with its left hand and *passes* it to its right hand for distribution.

For use in later explanations, we define the state of the flyer at the printer (FS_P), the left hand (FS_L), and the right hand (FS_R). Each variable is 1 if a flyer is there, and 0 if not. We set the robot so that it is always distributing flyers with its right hand, and thus the robot is ready for distributing when $FS_R = 1$; otherwise, it waits for the flyer to be prepared.

Figure 4.8 shows the framework of the flyer manager. The flyer is sent from the printer to the left hand and then to the right hand. Figure 4.7 shows an example of the robot's motion. When the printer has finished printing a flyer and the robot's left hand is empty (FS_P = 1, FS_L = 0), the flyer manager controls the robot to *take* the flyer from the printer (FS_L=1 and FS_P=0). Then the printer prints another flyer, which the robot *passes* its right hand (FS_R = 1, FS_L = 0) so that it is ready for handing. The robot *picks up* the flyer from the printer when the printer has finished printing it (FS_P = 1).

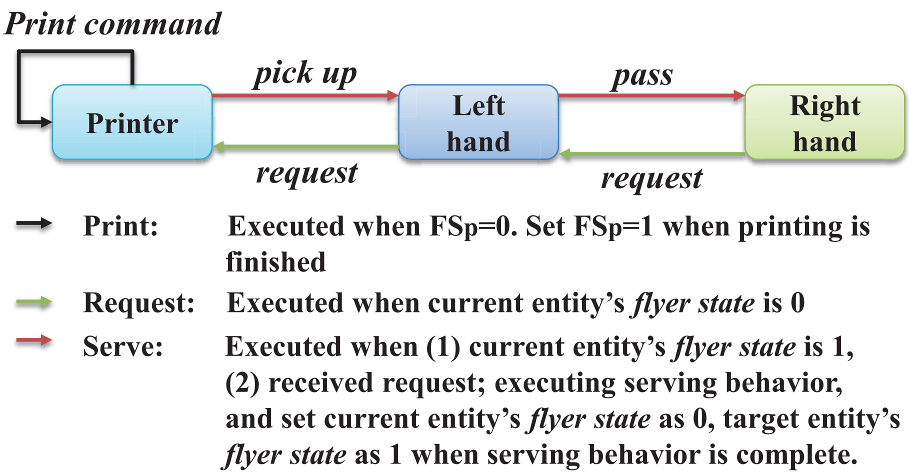


Figure 4.8 Flyer manager

4.2.2.3. LOCALIZATION AND PEOPLE-TRACKING

For robot localization, we used a particle filter with a ray-tracing approach on a grid map, which was built from odometry and laser scanner data. This module is called every 30 msec and updates the robot's position within 10-cm accuracy. People-tracking was done with its on-board laser range sensors, but to cover a large area, we used the *people-tracking* infrastructure explained in Section 4.1.1.

4.2.3. PLANNER

4.2.3.1. BASIC FRAMEWORK

Figure 4.9 illustrates the framework of our planner. When a flyer is ready (FS_R=1), the robot plans an approach to distribute flyers to each pedestrian in the area by calculating whether pedestrian i (p_i) is accessible from its current position c. Hereinafter, we refer to a plan to approach p_i from location c as $c \rightarrow p_i$. Then for each plan, the system evaluated utility $U(c \rightarrow p_i)$.

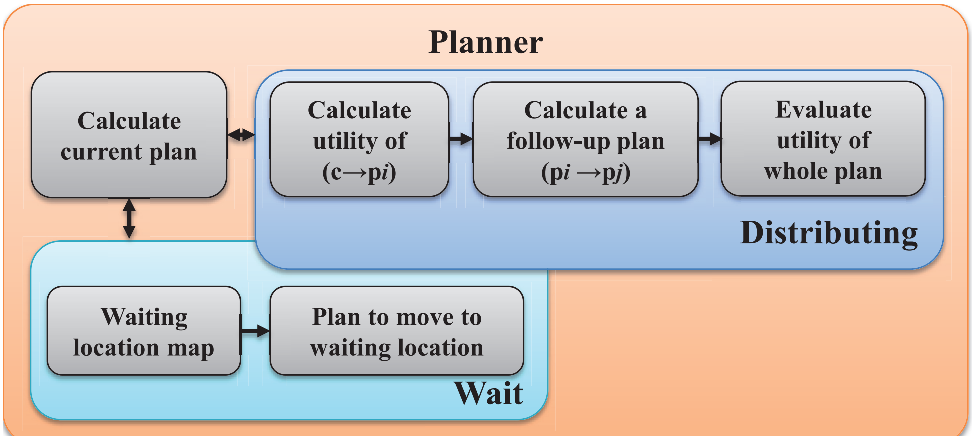


Figure 4.9 Processing in planner

The system also plans for the near future by estimating whether it is possible to continuously access the next pedestrian after the first pedestrian. It calculates whether pedestrian j is accessible from the location at which it gave a flyer to pedestrian i. This plan is referred to as $p_i \rightarrow p_j$. Finally, for each plan, the system evaluates utility $U(c \rightarrow p_i \rightarrow p_j)$, compares the utility with a threshold to eliminate unfeasible plans, and finally executes the plan with the highest utility.

If not all the pedestrians are accessible to the robot (or the utility of each plan fails to reach the threshold) or a flyer is not ready (FS_R=0), the robot transits to the wait mode and moves to a suitable location to wait, which is computed based on a waiting location map.

4.2.3.2. PLANNING DISTRIBUTION TO PEDESTRIANS

a) Calculating utility of each plan($c \rightarrow p_i$)

This module calculates the utility for each plan $c \rightarrow p_i$ by considering two factors: (1) the expected gain from distribution ($E(c \rightarrow p_i)$), and (2) the disturbance of other pedestrians ($D(c \rightarrow p_i)$). We define the utility of plan $c \rightarrow p_i$ as:

$$U(c \to p_i) = \max_{t_{pred} \in [0, t_{max}]} \left(E\left(c \to p_i, t_{pred}\right) - k_1\, D(c \to p_i, t_{pred}) \right), \qquad (3)$$

where $E(c \rightarrow p_i, t_{pred})$ represents the expected utility of the distribution when the robot distributes a flyer after t_{pred} seconds, $D(c \rightarrow p_i, t_{pred})$ represents the disturbance of other pedestrians around robot r after t_{pred} seconds, and k_1 is a coefficient parameter between two utilities, which was empirically set to 10.5. We set t_{max} as 15.0 sec for computational economy.

Below are the three computation steps for these utilities. The expected utility of the distribution is computed by considering two factors: time margin and continuous handing.

Step 1: Estimation of time margin ($U_{timeMargin}$)

For each pedestrian p_i, it tests whether future moment t_{pred}, at which pedestrian p_i will be at expected position $EP_i(t_{pred})$ is a good position for distributing. In other words, it tests whether there is a distributing position of robot G_{ready} that is suitable for pedestrian's position P_{ready} (= $EP_i(t_{pred})$), given the robot's current position $G_{current}$ (=c). It computes the time required to move from c to G_{ready} with the path shown in Figure 4.10.

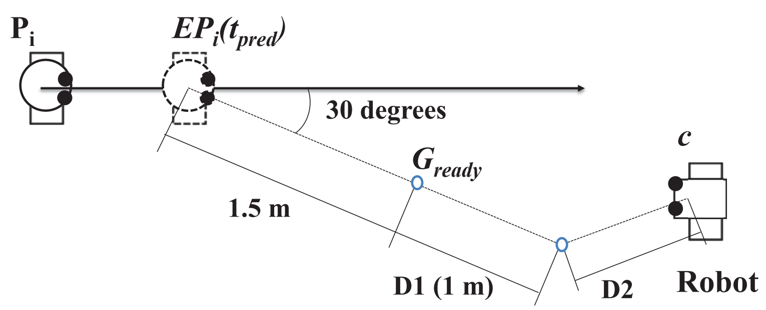


Figure 4.10 Planning a path to distribute to target pedestrian p_i at $EP_i(t_{pred})$ at future moment of t_{pred}. To do so, robot plans to move to G_{ready} by *WayPoint*.

As a simple implementation, $EP_i(t_{pred})$ is estimated by a linear interpolation with the past speed information of pedestrian i (Figure 4.4):

$$EP_i(t_{pred}) = P_i(t) + v^p \cdot t. \qquad (4)$$

G_{ready}, which corresponds to $EP_i(t_{pred})$, is calculated accordingly, and next we compute t_{arrive} as the time the robot takes from current location c to reach G_{ready} by *WayPoint*, which is a position on the line from G_{ready} to $EP_i(t_{pred})$ 1 m from G_{ready} (Figure 11). *WayPoint* is prepared to let the robot approach from the front of the target person. In this computation, to consider the robot's acceleration capability, we simulated its movements in small time steps of 0.1 sec and updated the velocity with the acceleration and angular acceleration capability.

As reported in Section 4.1.2, a productive giver does not stop to wait for the pedestrian, but keeps walking until she meets her pedestrian target. Thus, we evaluated a plan based on the timing when the robot meets a target pedestrian. Perhaps the best timing is when the giver reaches distributing position G_{ready} when the pedestrian reaches $EP_i(t_{pred})$; i.e., t_{arrive} equals t_{pred}.

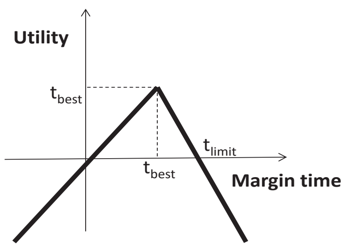


Figure 4.11 Calculation of U_{timing}

However, predictions are not always accurate. For instance, since people do not necessarily walk as predicted, it is better to choose a plan with a small margin time so that a robot can catch up from unexpected aspects in the prediction. Thus, we used the function shown in Figure 4.11 and designed a utility function to let the robot reach G_{ready} at t_{best} seconds earlier than the targeted person reached $EP_i(t_{pred})$, if possible:

$$U_{timeMargin}(c \to p_i, t_{pred}) = \begin{cases} timeMargin & (timeMargin \leq t_{best}) \\ \frac{t_{best}}{t_{best}-t_{limit}} timeMargin + \frac{t_{best} \cdot t_{limit}}{t_{limit}-t_{best}} & (timeMargin > t_{best}) \end{cases}, \qquad (5)$$

where

$$timeMargin = f_{tolerance} \cdot t_{arrive} - t_{pred}. \qquad (6)$$

We empirically set t_{best} to 1.6 sec and t_{limit} to 5.0.

We also tested whether the robot has enough time to extend its arm before arriving at G_{ready}, based on the constraint in Section 4.1.2:

$$ArmReady(t_{Pred}) = \begin{cases} true & (ArmTime \le t_{Pred}) \\ false & (otherwise) \end{cases}, \qquad (7)$$

where *ArmTime* estimates the moving time for each arm joint of the robot to reach the final position. The handing posture is calculated by the gesture controller, which is described below. If $ArmReady(t_{Pred})$ returns false, $U_{timeMargin}(t_{arrive}, t_{pred})$ returns negative infinity.

Step 2: Expected utility of distribution ($E(c \rightarrow p_i, t_{pred})$)

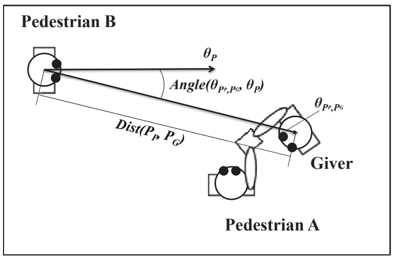


Figure 4.12 Estimation of continuous distribution

In addition, as we reported in Section 4.1.3, the *successful ratio* of distributing in *continuous* distributing is *significantly* higher than in *independent* distribution. Thus, we expect higher gain from *continuous* distribution. As shown in Figure 4.12, when a giver successfully distributed a flyer to pedestrian *A*, the system then evaluated the spatial formation between the giver and all the other pedestrians in the area to estimate whether they noticed the successful distribution. If pedestrian *B* seemed to notice that former pedestrian *A* accepted a flyer, distributing a flyer to pedestrian *B* is then considered a *continuous* distribution and is calculated below:

$$\begin{cases} 1 & Dist(P_P, P_C) < NoticeDist \text{ and } \left|Ang\left(\theta_{P_P,P_G}, \theta_P\right)\right| < 90 \\ 0 & otherwise \end{cases} \tag{8}$$

where P_G is the position of the giver at which she successfully completed a distribution to pedestrian *A* and P_P is the position of another pedestrian *B*. $Ang(\theta_{P_P,P_G}, \theta_P)$ is a function that indicates the angle (in degrees) between the pedestrian B's direction and the giver. θ_P is the pedestrian *B*'s walking direction. We used *NoticeDist* as the distance where the pedestrian noticed the former distribution and empirically set it to 5.0 m, based on our experiment environment.

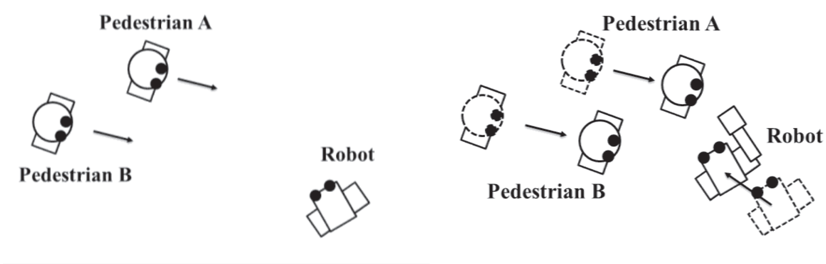
Overall, we defined the expected gain of a plan ($c \rightarrow p_i$) as:

$$E(c \rightarrow p_i, t_{pred}) = \begin{cases} U_{timeMargin}(c \rightarrow p_i, t_{pred}) \cdot Certain(t) & (independent) \\ U_{timeMargin}(c \rightarrow p_i, t_{pred}) \cdot Certain(t) \cdot (1+\delta) & (continuous) \end{cases}, \tag{9}$$

$$Certain(t) = \begin{cases} 1 - t/T_{effect} & if\ t < T_{effect} \\ 0 & other \end{cases}, \tag{10}$$

where δ is a parameter that represents the expected gain from the continuous distributing and was set to 0.6 based on the result in Table 4.2. *Certain(t)* is a function that represents the decay over time. This represents the effect where the future estimates of disturbances are uncertain because the prediction accuracy of the future behavior of pedestrians decreases with time. As a result, distributing plans which the robot could access pedestrian faster would have higher priority. We empirically set T_{effect} to 40.0 sec.

Step 3: Disturbance of other pedestrians ($D(c \rightarrow p_i, t_{pred})$)



(a) **(b)**

Figure 4.13 Disturbing other pedestrian when handing

When there are multiple pedestrians in the area, a robot approaching one pedestrian might disturb other pedestrians. Figure 4.12 shows an example where pedestrians *A* and *B* are walking in the area, and both are accessible to the robot. Since approaching pedestrian *A* costs less *estimatedTime* than pedestrian *B*, if no disturbance is considered, the robot would choose pedestrian *A* as a distributional target (Figure 4.13-a). However, when the robot approaches, it disturbs pedestrian *B*'s walking (Figure 4.13-b). To solve this problem, we used a distance-based comfort model [88]. The idea behind this model is that it is more comfortable for pedestrians if the distances to nearby persons are larger. Based on this model, pedestrian *j*'s discomfort during $c \rightarrow p_i$ is defined as:

$$D_j(\mathrm{c} \rightarrow p_i, t_{pred}) = \max_{t \in [0, t_{pred}]} \left(1 + Certain(t) \left(\frac{a}{dist(r,j,t)} + b\right)\right) \quad (11)$$

where *dist(r,j,t)* is the distance between robot *r* and pedestrian *j*' s body center at time moment *t*. *a* and *b*, which are the parameters for the distance-based comfort model, were imported from a previous work [88] to be 1017.76 and 1.180. *Certain(t)* is as the same as which in Eq. 10.

Finally, the total discomfort of all the pedestrians except distributing target pedestrian *i* in the area is calculated as below:

$$D(c \rightarrow p_i, t_{pred}) = \sum_{j \in S} D_j(c \rightarrow \mathrm{p}_i, t_{pred}), \quad (12)$$

where S represents a set of pedestrians around robot *r* except target pedestrian *i*.

With the above three steps, Eqs. 9 and 12 are derived from which utility $U(c \rightarrow p_i)$ of *plan* $(c \rightarrow p_i)$ can be calculated as defined in Eq. 3.

b) Calculating utility of entire distributing plan

After calculating a plan to approach pedestrian *i* ($\mathrm{c} \rightarrow \mathrm{p}_i$), the planner then calculates whether the robot can conduct a follow-up distribution to another pedestrian *j* ($\mathrm{p}_i \rightarrow \mathrm{p}_j$). We can similarly calculate $U(p_i \rightarrow p_j)$, the utility of plan $\mathrm{p}_i \rightarrow \mathrm{p}_j$ from Eqs. 3-12 by replacing *c* by p_i, which represents the robot's location after approaching pedestrian *i*.

The utility of the entire distributing plan is calculated below:

$$U(c \rightarrow p_i \rightarrow p_j) = U(c \rightarrow p_i) + U(p_i \rightarrow p_j). \qquad (13)$$

If plan $p_i \rightarrow p_j$ does not exist, $U(p_i \rightarrow p_j)$ returns 0.

Finally, among all the candidates of distributing plans, the system chooses the candidate plan with the highest utility as the next distributing plan to be executed:

$$U_{decision} = \max\left(U(c \rightarrow p_i \rightarrow p_j)\right). \qquad (14)$$

To prevent excessive switching of the target among multiple pedestrians (like oscillation), the target in the previous computation round is prioritized in the choice of target in the current round (the system doubled the utility). The plan is updated every 200 msec.

4.2.3.3. Alternative: planning to wait

This module provided a waiting location map, which is used by the robot to obtain its waiting position and orientation when there is no *distributional target*. A good waiting location is the place from which the robot can frequently approach many pedestrians, not simply the location where many pedestrians pass.

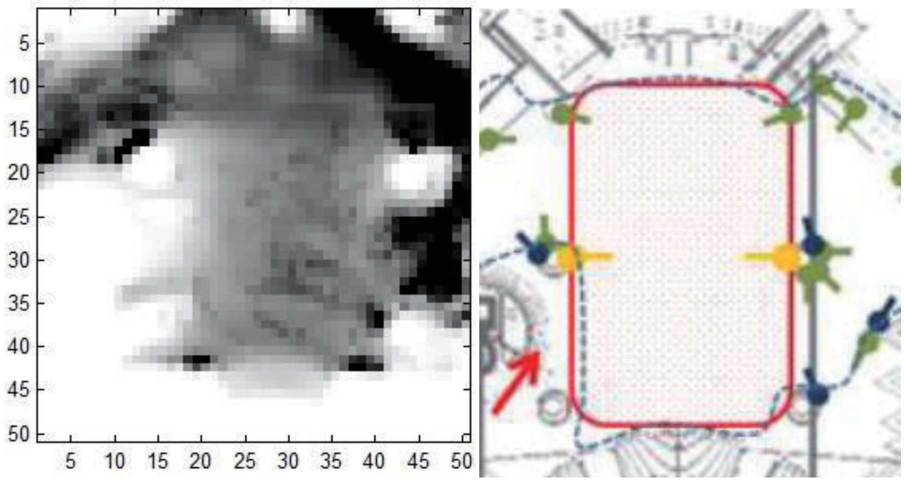


Figure 4.14 Example of grid map of average person density, values are in person/m^2

As we explained in Section 4.1.1, we collected pedestrian data in the area for a year. The data indicate that the density of the pedestrians changes over time; hence to reflect such temporal dynamics, we created grid maps of average person density for every 30 minutes. In this research, we used the hallway to conduct a robot distributing experiment and created a grid map with a 50 x 50 cm cell that divided the area. Figure 4.14 shows an example of the grid map of average person density, where the values are in person/m^2.

Next we created a map of candidates of waiting positions. The robot can move and approach pedestrians nearby from its waiting position. To incorporate this idea, we used the time needed by the robot to move from one grid to another. We also considered the robot's orientation. For each grid, 12 orientation candidates were set every 30 degrees. We calculated the value of each pair of position Pw and orientation θ_W candidates as follows:

$$Value = \sum_i^j \left[\left(\frac{Value_{G_i}}{\frac{Dist(P_W, P_{Gi})}{V_{Move}} + \frac{Angle(\theta_{P_W, P_{Gi}}, \theta_\omega)}{V_{Rot}}} \right) \right], \qquad (15)$$

where P_{Gi} is the position of grid i of the pedestrian density map, V_{move} and V_{rot} are the moving and rotating speeds of the robot, and $Value_{Gi}$ is the value of the pedestrian density in grid i. We chose the Pw and θ_W pair that yields the highest value as the waiting position and the orientation.

4.2.4. BEHAVIOR CONTROLLER

After a plan is selected by the *planner*, the *behavior controller* navigates the robot and manages its gaze and arm motions.

It controls the robot to reach the *WayPoint* first, and then it follows a line that connects $EP_i(t_{pred})$ and G_{ready} (Figure 4.8). Each position is updated every 100 msec. To make the robot reach the G_{ready} position at appropriate timing, we dynamically adjusted its velocity and controlled the robot to move at high speed first; to adjust the arrival timing, it starts to decrease its speed when it gets close to G_{ready}:

$$\begin{cases} v_{max} & \left(if\ \frac{RemainingDist}{RemainingTime} \cdot f_{tolerance} > v_{max}\right) \\ \frac{RemainingDist}{RemainingTime} \cdot f_{tolerance} & (otherwise) \end{cases}, \tag{16}$$

where v_{max} is the robot's maximum speed, *RemainingDist* is the summation of the distances from c to *WayPoint* and *WayPoint* to G_{ready}, i.e., D1 + D2 in Figure 11, and *RemainingTime* is the remaining time to t_{arrive} by considering the current time and the required time for rotating.

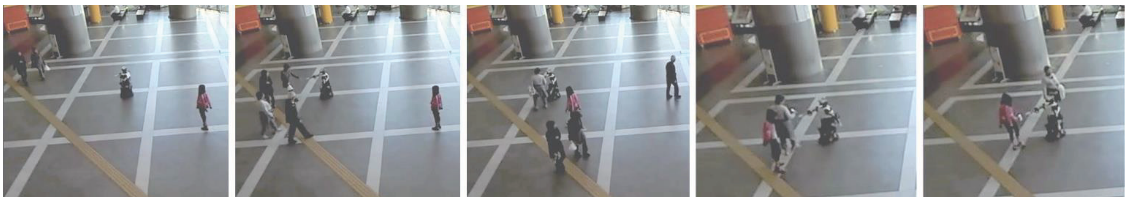
When the pedestrian accepts the flyer from the robot, the sensor attached to the robot's hand detects whether the flyer was taken and sets $FS_R = 0$. Otherwise, the robot waits for the target pedestrian to take the flyer; i.e., it continues to orient itself toward the target until the pedestrian passes the robot. We define δ as the angle between the pedestrian's moving direction and the vector from the position of the pedestrian to the robot. When $\delta >= 90$ degrees or the distance between robot and the target pedestrian exceeds 5 m, we consider the pedestrian to have passed the robot.

The gaze behavior is started from the very beginning, as explained in Section 4.1.2.1. The robot keeps directing its head direction toward the pedestrian until the *distributing* has finished.

For the *arm* motion, the gesture controller generates a handing motion. As illustrated in Figure 4.6, the height of the robot's hand was calculated using the target height information in Eq. 2. In addition, the robot said, "Please have a flyer" to the pedestrians at $t_{armstart}$ time.

4.2.5. EXAMPLE

Figure 4.15 shows a successful handing scene with our developed system. Two pedestrians came from the left, and another was standing on the right side (Figure 4.15-a), and two more were walking toward the hallway from the left (Figure 4.15-b). The robot calculated the utility of each candidate of the distributing plan and accessed the two pedestrians on the left because it anticipated a chance for a continuous handing. The robot chose one of the two nearby pedestrians to avoid disturbing the other pedestrians. It gave a flyer to the woman who took it, and at that moment, other pedestrians noticed that she had taken it (Figure 4.15-b). Then two more pedestrians (one in a white coat and another in red) approached the robot (Figure 4.15-c). Because the plan to hand a flyer to the pedestrian in white had the highest utility, the robot gave her a flyer, and she took it (Figure 4.15-d). The pedestrian in red came close to the robot, too. The robot gave a flyer to her, which she accepted (Figure 4.15-d).



(a) (b) (c) (d) (e)

Figure 4.15 Example of distributional handing robot

4.3. Evaluation of our Distributional Handing Method

We evaluated our distributional handing model in a field trial and compared our proposed method with a simple alternative method. Our evaluation criteria are the number of flyers successfully distributed by the robot. We also compared the developed robot with human givers to provide insight about the extent to which the robot's performance with these two models compares to the human performance.

4.3.1. Hypothesis and prediction

We compared our proposed model with an alternative method to evaluate its effectiveness. Since no commonly available method exists for distributional handing, we implemented a very simple wait-and-handing method. The robot stopped at its waiting position and handed flyers to pedestrians who passed nearby. This is the *wait for pedestrian and extend arm nearby* method explained in Section 4.1.2. Since we know that our *approach pedestrian and extend arm nearby* method outperformed the human givers, if it is implemented appropriately, it will probably also outperform this simple method.

If our proposed model is designed properly, we expect that the robot using it will perform more efficiently than the wait-and-handing method robot. First, since the robot with our proposed model can move around the environment and approach pedestrians, if target pedestrians are appropriately chosen, it should access more pedestrians than the robot that stopped and waited at a certain place. Accordingly, we made the following prediction:

Prediction 1 (access efficiency): With our proposed model, the robot will more efficiently access pedestrians than the wait-and-handing method robot.

Second, the robot in our proposed method plans for a distributing behavior that imitates the way productive human givers perform. According to the analysis in Section 4.1, it should yield more success than the other methods. Our proposed method's robot also plans to perform continuous handing, which should also yield more success. By combining both expected effects with prediction 1, if these calculations are appropriate, it will have a higher successful ratio of distributional handing than the wait-and-handing method robot:

Prediction 2 (flyer-distributing efficiency): With the proposed method, the robot will successfully give more flyers than the wait-and-handing method robot.

If the distributional handing models are proposed and implemented well, we also expect that the robot with our proposed model will be able to distribute flyers efficiently like a human giver. Thus, we also compare its performance with that of the humans.

4.3.2. METHOD

4.3.2.1. SETTINGS

The evaluation was conducted in the same shopping mall where we conducted the data collection and our first evaluation experiment. The robot was placed in a large 8 x 12 m hallway (Figure 17), which connects to an event hall and a train station, and there are restaurants and shops nearby.

4.3.2.2. COMPARISON

We compared the following two methods.

Proposed: The robot is controlled by the proposed model. When pedestrians arrive, the target decider (Section 4.2) chooses the handing target and the behavior controller manages the robot's local behavior (Section 4.2); otherwise the robot waits at the waiting location.

Wait-and-handing: The robot stops and waits at its waiting position, which was calculated with the same method used for the proposed condition. The robot does not move around but only extends its arm to pedestrians who pass nearby. As a pedestrian approaches within 3.0 m of the robot, the robot chooses her as the handing target and starts to look at her. It starts the handing motion when she is within 1.0 m so that she can comfortably accept the flyer. These distance parameters were empirically decided.

We prepared three afternoon time slots (2:00, 3:00, and 4:00) that more customers came to the shopping mall relatively, assigned each condition to the time slots with counter-balancing, and collected 200 minutes of data for each condition. We ensured that the time lengths of the collected data of each condition are identical in each time slot.

Human giver: For comparison, we sought human data from the previously collected data. As shown in Table 4.1, the average successful ratio of the ten givers we analyzed was 40%. To avoid outliers (too good or too bad givers), we used those human givers who provided average successful ratios that resembled the average level of 40%. We found three human givers who engaged in distributional handing around the time similar to our time slots and averaged their performances. We retrieved 200 minutes of data for analysis; the time lengths of the data in each time slot are identical to the robot experiment data.

4.3.2.3. Measurement

We counted the number of times the robot/person offered flyers (*#Handing*). However, since such a number is largely affected by the number of pedestrians, we normalized it by dividing by the number of pedestrians who passed through the experiment area (*#Pedestrians*). We defined the following evaluation criteria:

$$AccessEfficiency = \frac{\#Handing}{\#Pedestrians}. \qquad (17)$$

We also counted the number of times the robot/person successfully gave flyers (*#DistributedFlyers)* and defined the following criteria:

$$FlyerDistributingEfficiency = \frac{\#DistributedFlyers}{\#Pedestrians} \quad (18)$$

4.3.3. RESULTS

4.3.3.1. DATA ANALYSIS

There were 200 minutes of data for each condition. We measured the number of pedestrians using people-tracking infrastructure. We checked the average number of pedestrians in all time slots for each condition and confirmed that they are reasonably similar (ANOVA shows no significant difference ($F(2,117)=2.548$, $p=.184$)). Two people independently counted the number of pedestrians accessed by the robot/person, and the numbers of pedestrians to whom they successfully gave flyers. The results of the two coders were exactly the same, showing that the coding result is highly reliable.

4.3.3.2. VERIFICATION OF PREDICTION 1

Table 4.3 shows the results. The access and flyer-distributing efficiencies were counted for each five-minute time slot. The average access efficiency was 0.21 (s.d. 0.07) for the proposed method, 0.17 (s.d. 0.07) for the wait-and-handing method, and 0.25 (s.d. 0.11) for the three human givers.

An ANOVA revealed a significant main effect ($F(2,117)=8.399$, $p<.001$). A multiple-comparison with the Bonferroni method revealed that the access efficiency for the human givers was significantly higher than that for the wait-and-handing method ($p<.001$). No significant difference was found between the proposed and wait-and-handing methods ($p=.160$) or between the proposed method and the human givers ($p=.102$).

Thus, our first prediction was supported; the proposed method yielded higher access efficiency than the wait-and-handing method.

Table 4.3 Evaluation Results

	Access efficiency	Flyer-distributing efficiency	Person density (num./min.)
Proposed	**0.21** (s.d. 0.07)	**0.18** (s.d. 0.05)	5.7
Wait-and-handing	0.17 (s.d. 0.07)	0.12 (s.d. 0.05)	6.1
Human giver	**0.25** (s.d. 0.11)	0.10 (s.d. 0.04)	5.7

Bold face indicates values significantly higher than others.

4.3.3.3. Verification of prediction 2

The average flyer-distributing efficiency was 0.18 (s.d. 0.05) for the proposed method, 0.12 (s.d. 0.05) for the wait-and-handing method, and 0.10 (s.d. 0.04) for the three human givers.

An ANOVA revealed a significant main effect ($F(2,117)=31.093$, $p<.001$). A multiple-comparison with the Bonferroni method revealed that the distributing efficiency for the proposed condition was significantly higher than that for both the wait-and-handing method ($p<.001$) and the human givers ($p<.001$). No significant difference was found between the wait-and-handing method and the human givers ($p=.104$).

Our second prediction was supported; our proposed method yielded higher flyer-distributing efficiency than the wait-and-handing method.

4.3.3.4. Additional analysis

We further analyzed the details to determine why our proposed condition yielded higher flyer-distributing efficiency than the others. Since we expected that the robot in the proposed condition would appropriately access pedestrians to yield a higher successful handing ratio and would also plan to more frequently perform *continuous handing*, we analyzed whether all of these effects were visible. Two people classified each of the handings as either *individual* or *continuous,* which we described in Section 4.2. Cohen's kappa coefficients from their classifications were 0.832, 0.874, and 0.910

for the two conditions and the human giver, indicating that their evaluations were highly consistent.

Table 4.4 Detailed results

	Individual handing	**Continuous handing**	**Ratio of continuous handing**
Proposed	**77%** (86/112)	**99%** (111/112)	**0.500**
Wait-and-handing	51% (63/123)	**99%** (90/91)	0.425
Human giver	24% (46/189)	68% (67/98)	0.341

Bold face indicates values significantly higher than others.

Table 4.4 shows the analysis result. During *individual handing*, the successful ratios (whether pedestrian took the flyer) significantly differed across the conditions. The successful handing ratio was 77% for the proposed condition, 51% for the wait-and-handing condition, and 24% for the human giver condition. A Chi-square test revealed significant differences among them (χ^2 (2) = 79.786, $p<.01$). Residual analysis revealed that the successful ratio in the proposed condition is significantly higher than the others ($p<.05$), and the one in the human giver condition is significantly lower than the others ($p<.05$).

When the handing was *continuous*, the successful ratio was 99% for the proposed condition, 99% for the wait-and-handing condition, and 68% for the human giver condition. A Chi-square test revealed a significant difference between them (χ^2 (2) = 63.598, $p<.01$). Residual analysis revealed that both the proposed and wait-and-handing conditions are significantly higher than the human giver condition ($p<.05$).

To analyze whether more frequent *continuous handings* were performed, we calculated the *continuous handing* ratios from all handings. The *continuous handing* ratios for the proposed, wait-and-handing, and human giver conditions were 0.500, 0.425, and 0.341. A Chi-square test revealed significant differences among them (χ^2 (2) = 13.150, $p<.01$). Residual analysis revealed that the *continuous handing* ratio in the proposed condition is significantly higher than the others ($p<.05$), and in the human giver condition, it was significantly lower than the others ($p<.05$).

4.3.4. INTERVIEW

We were surprised that the proposed method outperformed the human givers. To better understand this result when the robot operated with our proposed model, we interviewed 17 pedestrians (16 who accepted flyers, and 1 who did not) and asked them why they took the flyers. The interviews were recorded and transcribed for analysis.

First, we systematically analyzed the answers from the 16 pedestrians who accepted flyers and classified their answers into four categories. The classification was confirmed by two human coders who did not know the research purpose. Their coding results were identical.

Table 4.5 shows the analysis result. The answers of six pedestrians were classified as *natural handing behavior.* We gathered the following comments from them:

Table 4.5 The Reason Why Pedestrians Took Flyers

	Ratio
Behavior-oriented reasons	
Handing behavior was natural	37.5% (6/16)
Influenced by precedent pedestrians	25.0% (4/16)
Robot-oriented reasons	
The robot was interesting	25.0% (4/16)
Handing behavior was impressive	12.5% (2/16)

"There was no particular reason. The robot slowly came to me and stuck out its arm. It behaved smoothly. So I just reached out and took it."

"It timely and quickly offered me a flyer. I thought, ok, I'll take it."

We classified the answers of four pedestrians as *influenced by other pedestrians*:

"I saw that the person before me took the flyer from the robot. So I thought, I'll take it too."

These two categories are *behavior-oriented reasons*, which might also occur if a person nicely distributes flyers.

On the other hand, the two other categories are considered *robot-oriented reasons*, which were caused by the fact that the robot distributed the flyers. The answers of four pedestrians were classified as the *robot was interesting*. One of them said:

"I took it basically because I was interested. I wanted to see how it would react."

The answers of two pedestrians were classified as *impressive handing behavior*:

"I was very impressed by the robot. It clearly said 'please take a flyer.' It even made eye contact with me. I felt that it really wanted me to take it."

We categorized them separately from the *natural handing behavior* category, because they would not consider it impressive if a person distributed the flyer.

One pedestrian did not accept the flyer from the robot for the following reasons: "I looked at the flyer and realized that it was just a map of this shopping mall. Since I've worked here for more than ten years, I didn't need it."

4.4. DISCUSSION

4.4.1. INTERPRETATION OF FIELD EVALUATION RESULTS

As hypothesized, our proposed condition performed the best in successfully handing flyers. It was also the best in terms of a successful handing ratio in *individual handing*, which indicates that it appropriately imitates a good giver's behavior. Because it was the best in terms of *continuous handing* ratio, we also believe that our proposed method successfully planned appropriate handing behaviors so that it performed *continuous handing* more frequently. Overall, we successfully implemented key factors identified from the analysis of human behavior.

We observed some differences between the proposed method and the human givers. The human givers yielded higher *accessing efficiency* than the robot with the proposed model. This is not surprising since the robot's motion is not as swift as the humans. However, overall, the proposed method yielded higher *flyer-distributing efficiency*. This was because it yielded a much higher successful handing ratio both in *individual* and *continuous handing*, because it conducted *continuous handing* more frequently.

Our interview results explain part of the reason why the robot yielded a higher successful handing ratio. Six out of 16 pedestrians gave robot-oriented explanations why they took flyers. They were either interested in the robot or impressed by its behavior. Although recently robots are often seen on television or the Internet, few people have really interacted with them. Thus, the robot's success was partly due to its novelty.

4.4.2. FOR FUTURE USE OF FINDINGS FROM THIS STUDY

Future robots that serve distributional handing services might be implemented with our method reported in this paper, although some of the situations when our study was conducted will change over time; this issue must be carefully addressed. First, although our study showed that our robot provided better *flyer-distributing efficiency* than the human givers, as more and more robots are introduced into actual environments, such novelty will fade. On the other hand, such robot capabilities as locomotion speed and flexibility will eventually improve. Currently, since the robot we used had poorer capability than the humans, it is not known whether robots will overall yield more or less success than humans.

4.4.3. LIMITATIONS AND FUTURE WORKS

In our field trial, we used sensors attached to the environment. But attaching sensors to environments might not be simple. Even though we believe that we can easily build a similar robot system that only uses sensors attached to the robot, we have not tested such a configuration yet. One requirement for such a system is that target detection needs to be done at a range from several to 10 m. One possible future work will confirm whether our developed method will work with on-board sensors. In addition, the parameters in our method were analyzed from or calibrated for Japanese people and our four robots. When our proposed method is used elsewhere, the parameters must be adapted.

5. Social Robots Attracting Passersby

5.1. Understanding Managerial Perspectives

We interviewed store managers to identify their expectations and requirements.

5.1.1. Procedure

5.1.1.1. Context

Our study was conducted in a suburban shopping mall that has 114 stores/restaurants. Most of the stores sell such common items as clothes, shoes, sporting equipment, or equipment for outdoor activities. The shopping mall is usually busy on weekends, and on weekdays it is used more by people from the nearby offices.

5.1.1.2. Participants

We contacted the store managers in one area of the mall through the mall administrators. We requested interviews with them to gather their thoughts on the use of social robots in their stores. 10 of 13 stores accepted our request.

5.1.1.3. Protocol

To explain and describe the current capabilities of autonomous robots, we showed videos in which robots engaged in three common store tasks: *announcement* (a robot periodically talks about advertisements without responding to individuals), *invitation* (whenever passersby approach, a robot looks at them and announces an advertisement), and *distribution* (a robot extends its arm to an approaching passerby to give her a flyer). We did not limit the options to these tasks; the managers were free to suggest other tasks they were interested in. Then we conducted semi-structured interviews about the following topics:

1) **Intention to use**: We asked whether they would like to use robots in their stores as well as their reasons for wanting to use a robot (e.g., what they expected the robot).

2) **Design requirements**: We asked them more specifically how they would like to use the robot by focusing on desired tasks and other design requirements (e.g., behaviors and appearances).

3) **Concerns**: We asked them what behaviors the robot must avoid if they are going to use them as well as the behaviors they are concerned about when others use a robot.

The interviews, which lasted an average of about 30 minutes, were recorded and transcribed. We classified their responses into categories based on their answers. Some answers were classified into multiple categories. Two independent coders classified them into categories. Their judgment matched reasonably well and yielded Cohen's kappa coefficient of 0.67 on average.

5.1.2. Interview results

5.1.2.1. Intention to use

Eight of the managers expressed that they wanted to use the robot, and two wanted to try and see whether it was effective. Table 5.1 shows the categorized results of the coding of their reasons for their intention to use.

Table 5.1 Reasons for intention to use

Inexpensive labor	**Total**	**7**
	Information value	7
	Inexpensive human-like labor	5
Uniqueness of robots	**Total**	**9**
	Efficiently attracts passersby	9
	Relieves stress	3

We identified two main ideas about their reasons for wanting to use a robot. Seven managers mentioned inexpensive labor:

> "Robots might be useful for sales promotions because they could tell people about our store." (*information value*)
>
> "Finding and hiring new employees is difficult. Too many people don't want to work in the service industry." (*inexpensive human-like labor*)
>
> "I don't afford enough employees to deal with a sudden large number of visitors. I don't want customers to wait too long. A robot could ease such busy situations." (*inexpensive human-like labor*)

Nine managers addressed the *uniqueness of robots*:

> "Not very many people have actually seen a real robot. If a passerby sees a real one, children will approach it, and adults might stop. If it were used for a sales promotion, I'd expect a large effect." (*efficiently attracts passersby*)
>
> "A robot is different from a high-pressure salesman. With a person, a customer is probably cautious. Once engaged, a person might not stop explaining until he gets a sale. For a robot, people wouldn't be so concerned. They might listen more to a robot's explanation than if I greet them." (*Relieves stress*)
>
> "Greeting passersby is stressful. Employees are often reluctant to do it. A robot would greatly reduce stress." (*Relieves stress*)

5.1.2.2. Design requirements

Table 5.2 shows the categorized result for the desired tasks. Since each store has different characteristics, each one requested a different type of task for the robot. Among the three tasks for which we showed examples, eight managers expressed interest in *invitation* because they want passersby to stop and pay attention to their stores. Seven expected robots to perform *distribution* tasks and provide discount coupons or flyers with store information that is too complicated for signboards. No one had any interest in *announcement* tasks. One manager specifically commented on this:

> "It's better for the robot to respond to people. In that case, people will perceive it as a robot."

Relevant to these tasks, four managers mentioned *enjoyment*. They expect a robot to provide entertainment, particularly for children, so that visitors will stay in their stores longer and create a positive atmosphere to increase sales.

Table 5.2 Expected tasks

Invitation	8
Distribution	7
Enjoyment	4
Greeting and chatting	4
Cashier	3
Cleaning and refilling	2
Translating	2

They also mentioned a couple of other tasks in addition to our examples. Four managers wanted robots to greet and briefly distract customers when clerks are too busy. Three managers wanted robots to serve as cashiers, and two wanted robots for such cleaning and replenishment tasks as wiping tables and refilling drinks in a restaurant because such tasks are time-consuming and onerous. Two wanted robots to serve as translators.

Table 5.3 shows the categorized results for the design requirements. Store managers made the following comments:

> "A bigger robot is better, because it will attract more attention." (*noticeable*)

> "I want it to wear a white coat. That way, people will recognize that it is associated with my store." (*indicate relationship with stores*)

“I prefer a round one that can walk around. It’s cute and moonfaced and looks charming.” (*familiarity*)

“On weekends, let it offer balloons to children, who are the primary targets of inviting services. During weekdays, let it attract adult passersby.” (*context dependency*)

“People will clearly realize that it is a robot, a machine. That is good for attracting passersby.” (*robot-likeness*)

Since each store manager has his/her own design preferences, no single design works for all.

Table 5.3 Design requirements

Noticeable	6
Indicates relationship with stores	5
Familiarity	4
Context dependency	3
Robot-likeness	3

5.1.2.3. Concerns

Table 5.4 shows the categorized results for behaviors to avoid when they use a robot. Four managers mentioned *bothering visitors*, and four mentioned *safety risks*:

“It should not obstruct people who want to enter the store or walk past. A robot can’t bother people.” (*bothering visitors*)

“Since it is a machine, it might fail and cause an injury. For example, its arm might jerk and hit a child. Adults would probably be able to avoid such risks. But small children might get too close to it.” (*safety risk*)

Table 5.4 Behaviros robots must avoid

Bothering visitors	4
Causing safety risks	4

Table 5.5 shows the categorized results for their concern when other stores use robots. Their opinions were split. Five managers were not concerned and actually

encouraged other stores to use robots, but five were concerned about robots being used by other stores:

> "If visitors stop around here, they might enter the store across from mine, or the store next door as well as mine. If the robot attracts visitors to this area, even if they go to another store, I'm still happy." (*no concern*)

> "If a store has a robot that sells similar items as my shop, I'd want them to operate it at a distance away from my store, so that I don't have to see or hear it." (*do not want competitors to use a robot around their stores*)

> "It's hard to predict how often others would use the robots. If everyone uses them, their effect will be diminished. A visitor can see robots in too many different places if every store uses them at the same time." (*do not want to be used by others at all*)

Table 5.5 Concerns about use by other stores

No concern (expecting ripple effect)	5
Do not want competitors to use around their stores	4
Do not want to be used by others at all	1

5.2. FIELD TRIAL

We conducted three case studies in actual stores to observe how the robots were employed.

5.2.1. PROCEDURE

We invited stores that responded to our interviews (Section 5.1) to deploy a social robot as a research trial. We offered two hours of use for four days to each store. All stores wanted to use it, and so far we served three stores in first-come first-served principle (we plan to serve to other stores too). Within the limitations of our autonomous robot's capabilities, we consulted with them about how they would like to use the robot. They could choose hardware from three robots (only one robot was capable of distributing flyers). We implemented services based on their requests and sought feedback from the managers about how to improve it. We documented these processes. The study was approved by our institutional review boards.

5.2.2. MEASUREMENTS AND BASELINE

For each store, we evaluated how frequently passersby stopped at and visited the store as follows:

- **Stop**: Determined whether a passerby *stopped* near the store. This includes cases where people stopped around the robot.
- **Visit**: This only includes cases where people *visited* the store and excludes cases where they only interacted with the robot. For instance, if a person bought something from a store or stopped and apparently observed its products (Figures 5.4-c and d), it was judged as a *visit*. If a person only stopped for the robot (Figure 5.3) or just glanced at the products, it was judged as *not a visit*.

Stopped and *visited* were coded from videos by two coders who did not know the research hypothesis. If they found the same person appeared again to *stop*/*visit*, only the first one was evaluated. The first coder coded all the data, and the second did confirmatory coding for 10% of the data.

We compared these ratios in two situations as follows:

- **Robot present**: ratio of people who *stopped/visited* the store among those who passed by while the robot was being operated during the eight-hour observation period.
- **Baseline**: ratio of people who *stopped/visited* the store among those who passed by during an eight-hour observation period. We selected the observation period with the same time and the same day of week as the *robot present* situation.

Finally, we interviewed the managers to determine whether they were satisfied and had an intention to use the robot again.

5.2.3. IMPLEMENTATION: ROBOT SYSTEM AND SERVICES

The managers wanted to use *invitation* and *distribution* services. We implemented a fully autonomous system for these services and consulted with each manager to adjust the services to suit each store. We limited the service to be non-mobile, where the robot rotates its body orientation without moving around. Based on current levels of autonomy, if a robot moves around, it is relatively difficult to prevent it from *bothering visitors* and causing *safety risks*. Our intention was to keep the implemented services rather simple so that the robots could robustly operate autonomously and be feasible for actual use.

Robot hardware and infrastructure

Managers so far choose only one robot. It is characterized by its human-like physical expressions. It is 120-cm high with a 40-cm diameter on a mobile platform. It has a 3-DOF head and 4-DOF arms (Figure 1.3, right). We used a technique for a people-tracking system [85] for detecting passersby.

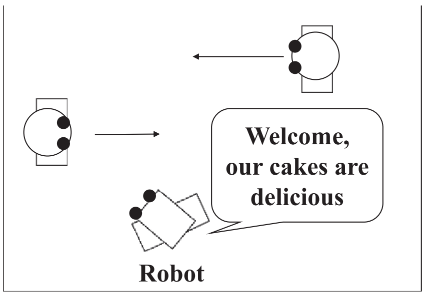


Figure 5.1 Invitation service

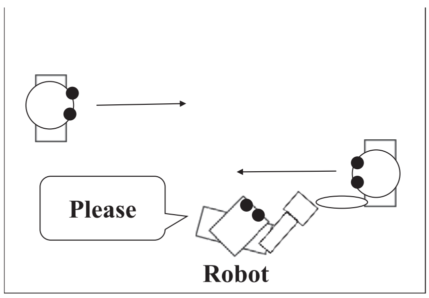


Figure 5.2 Distribution service

Invitation service

We modeled the shopkeeper behaviors when they greet passersby. They typically look at them and greet them with such advertisement utterances as "welcome, our shop sells. . . " The robot obtains the positions of potential visitors within 5 m from its people-tracking system and chooses as a target one who is approaching the closest to the robot (Figure 5.1). It orients its body and head direction to the target and periodically makes advertisement utterances.

Distribution service

We adapted a technique for distributing flyers reported in [89]. The system predicts the future locations of passersby and selects the person who will soon pass near the robot. Then it looks in her direction, and when she comes close, it extends its arm to give her a flyer, and says "please take it" (Figure 5.2). Its hand has a touch sensor, which detects whether the flyer was taken. When it detects that a flyer was taken, it says "thank you." The robot is equipped with a printer and prepares subsequent flyers by itself.

In both services, we followed the requests from managers and let the robot periodically give advertisement utterances if no target person was selected so that people far from the robot might hear such utterances and approach the robot.

5.2.4. CASE 1: INVITATION SERVICE AT A CAKE SHOP

5.2.4.1. STORE CHARACTERISTICS

The shop (Figure 1.3) faces a corridor of the mall. Inside its showcase, cakes and puddings are displayed. In the booth one clerk wears a baker's costume and tends to the store. On the weekends, they attract passersby by loudly greeting them.

5.2.4.2. MANAGER'S DESIGN DECISION

She chose a robot to perform an *invitation service*. She did not have a strong preference about the robot's appearance, and she chose one she had seen before. She wanted the robot to explain two types of information: 1) the unique features of her products, e.g., no additives, no artificial colors, and 2) advertisement of store items. She wanted it to wear a baker's costume that resembled that worn by the clerk. But since we couldn't find a costume that appropriately fit the robot, instead we put the shop's logo on its chest.

We implemented the *invitation* service, in which the robot randomly announces one advertisement utterance from three (no additives and no artificial colors, roll cakes, and pudding). After showing the initial implementation, she asked us to adjust the intervals between the utterances; we changed them to two seconds.

5.2.4.3. Results

Figure 5.3 People *stopped* around the robot, although this was not coded as a *visit*

Many passersby glanced at the robot or the store. Some stopped around the robot to interact with it (Figure 5.3, left). Families with children often interacted with it. We sometimes observed that while their children were interacting with the robot (Figure 5.4-a), parents visited the store (Figure 5.4-b and c) and bought something (Figure 5.4-d).

Two coders' judgments for *stop* and *visit* matched well and yielded a Cohen's kappa coefficient of .784. We applied a Chi-square test for the *stop* and *visit* ratios. It revealed that passersby *stopped* significantly more frequently around the store in the *robot present* situation (13.73%) than in the *baseline* (2.71%) ($\chi^2(1)= 915.023$, $p<.01$, $\varphi=.20$). Further, as illustrated in Figure 5.5, passersby *visited* the store significantly more frequently in the *robot present* situation (3.22%) than in the *baseline* (2.40%) ($\chi^2(1)=13.646$, $p<.01$, $\varphi=.025$).

(a) (b)

(c) (d)

Figure 5.4 Robot successfully enticed passersby to *visit*

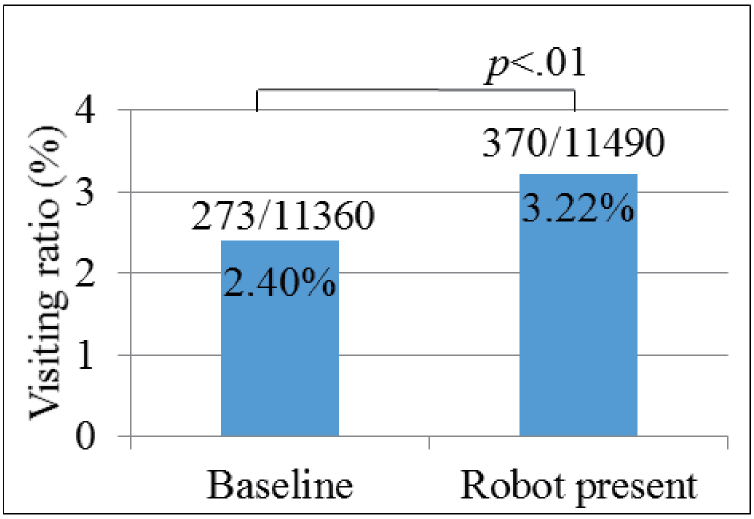


Figure 5.5 Ratio of *visited* passersby for cake shop

5.2.4.4. MANAGER'S FEEDBACK AFTER USE

She wanted to use it again. When we discussed the cost she was willing to pay, she said that she would use it if it cost less per hour than a human worker. She also expressed interest in buying the robot to avoid paying an hourly wage. She mentioned three benefits: 1) the robot nicely attracted passersby on weekends; 2) it increased sales; and 3) it reduced the workload of the shopkeepers by talking about promotions instead of the shopkeeper.

In contrast, she was less eager to use it again on weekdays. She felt that during weekdays, generally children approached the robot, and thus she didn't think it contributed to sales. She also felt embarrassment operating the store while surrounded by children (Figure 5.3, right).

5.2.5. CASE 2: INVITATION SERVICE AT A DRUGSTORE

5.2.5.1. STORE CHARACTERISTICS

The store (Figures 5.6 and 5.7), which is operated by the manager in a white coat, sells medicines and bottled drinks. No other clerk is employed. He does not actively attract passersby; when a customer stops, he explains medicines that are appropriate for customers.

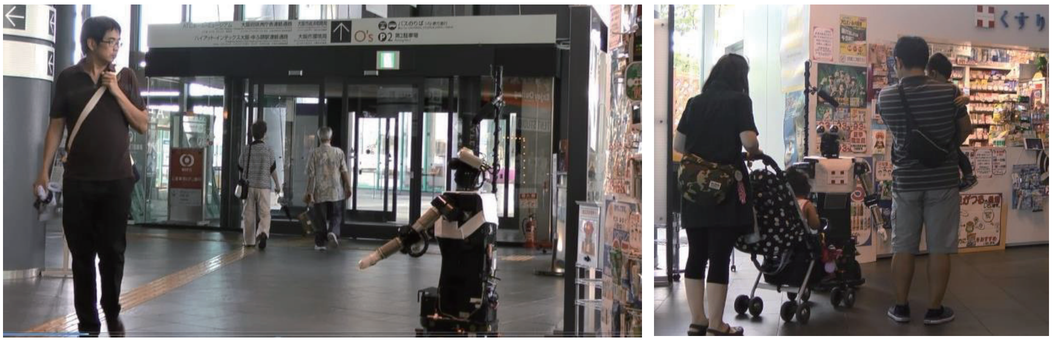


Figure 5.6 Passersby around drugstore

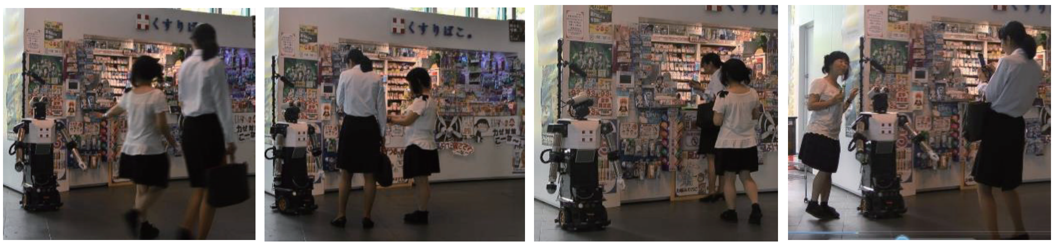
5.2.5.2. MANAGER'S DESIGN DECISION

He wanted the robot to move around and attract passersby and promote drinks instead of medicines. If visitors need medicine, he knew they would visit the store with or without special promotions. Moreover, a shopkeeper needs to determine very quickly which medicine is appropriate for each customer, which is not possible for a robot. Instead, he wanted the robot to attract children to buy juice and other drinks.

He chose an invitation service using the same robot (Figures 5.6 and 5.7) because it has a robot-like appearance and he had seen it before. He wanted it to wear a white coat, but we were unable to prepare a white coat for the robot, so he placed an emblem of the drugstore on its chest instead. We consulted with him and determined the interval between utterances to be two seconds. During the preparation, he wanted to increase the variation of the robot's advertisement utterances. Initially, it only mentioned juice for children, but we also added a comment for adults: "If you are tired, how about an energy drink?"

5.2.5.3. *RESULTS*

Many passersby glanced at the robot (Figure 5.6, left), and some interacted with it (Figure 5.6, right). For example, they waved their hands and talked to it. We observed cases where the robot successfully attracted passersby. Figure 5.7 shows one such scene. A mother and her daughter stopped in front of the robot (Figure 5.7-a) and heard that the store sells juice (Figure 5.7-b). She bought some juice while talking with the manager about the robot (Figure 5.7-c) and took a picture it with her daughter (Figure 5.7-d).



(a) (b) (c) (d)

Figure 5.7 Passersby who listened to robot and *visited* drugstore

The two coders' judgments for *stop* and *visit* matched well and yielded a Cohen's kappa coefficient of .848. We applied a Chi-square test for the *stop* and *visit* ratios. It revealed that passersby *stopped* more frequently around the store in the *robot present* situation (14.84%) than in the *baseline* (1.36%) ($\chi^2(1)$= 1504.145, p<.01, φ=.242). However, their frequency of visits in the *robot present* situation (1.13%) did not differ significantly with that in the *baseline* (1.03%) ($\chi^2(1)$=.447, p=.504, φ=.005).

5.2.5.4. *MANAGER'S FEEDBACK*

The manager wanted to use the robot again. He plans to move his store and asked whether the robot could come to his next location, too. Even though he admitted that the robot probably did not really contribute much to sales, yet he deemed it useful because it attracted browsers and advertised the store. When we asked how much he would pay, he said that he would pay as much as a human worker.

Although he expressed interest in using the same robot again, he wanted to improve its interactivity, e.g., saying "thank you" to customers, even if it were teleoperated by a shopkeeper. He believed that visitors wanted to interact with the robot, and such reactions from it would encourage visitors to make more purchases.

5.2.6. CASE 3: DISTRIBUTING DISCOUNT COUPONS FOR A DONUT SHOP

5.2.6.1. STORE CHARACTERISTICS

This store sells donuts. A single shopkeeper wears an orange-color T-shirt, hat, and an apron and runs the store. When passersby are around the store, she promotes her products by loudly making such utterances as, "how about some donuts?"

5.2.6.2. MANAGER'S DESIGN DECISION

The manager wanted to use the robot to distribute discount coupons (get a free donut with a purchase over 500 yen) that lasted until the end of the next month. Since he also wanted it to announce advertisements, he put ads in a coupon leaflet about relatively unknown donuts.

We implemented a distribution service. When it estimates that it can give a coupon to a passerby within three seconds, it starts the distribution service. Otherwise, it repeats its advertisement information from three candidates, such as "Hello, we have many kinds of delicious donuts." In communication with the manager, we adjusted the time intervals between utterances to three seconds.

He wanted it to wear the same uniform as the shopkeeper. We put the store's hat on the robot's head and attached a T-shirt to the front side of its body (Figure 5.8). Though he initially wanted a cute voice for the robot, after he observed its use in another store, he decided to use its default synthesized voice because passersby seemed to react positively to it. During the trial, he changed the robot's location to increase its visibility.

Figure 5.8 Scene at donut shop

5.2.6.3. Results

Many passersby noticed the robot and passed through the corridor while looking at it. Some took the coupon (Figure 5.8), and others visited the store. They typically stopped to accept a coupon (Figure 5.9, left), listened to the robot, looked at the coupon (Figure 5.9, center), and glanced at the store's shelf on which the donuts were displayed (Figure 5.9, right). Some then bought donuts with the coupons. It distributed 413 coupons during an eight-hour trial.

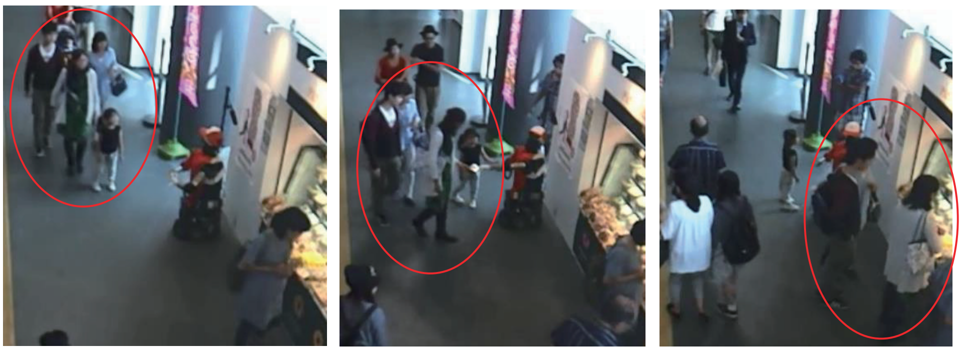


Figure 5.9 Passersby received coupons and *visited* donut shop

Two coders' judgments for *stop* and *visit* matched well and yielded a Cohen's kappa coefficient of .794. We applied a Chi-square test for the *stop* and *visit* ratios. It revealed a significance that passersby more frequently *stopped* around the store in the *robot present* situation (7.39%) than in the *baseline* (3.28%) ($\chi^2(1)$= 356.679, p<.01, φ=.091). Further, as illustrated in Figure 5.10, passersby *visited* the store more frequently in the *robot present* situation (3.35%) than in the *baseline* (2.94%) ($\chi^2(1)$= 6.006, p=.014, φ=.011).

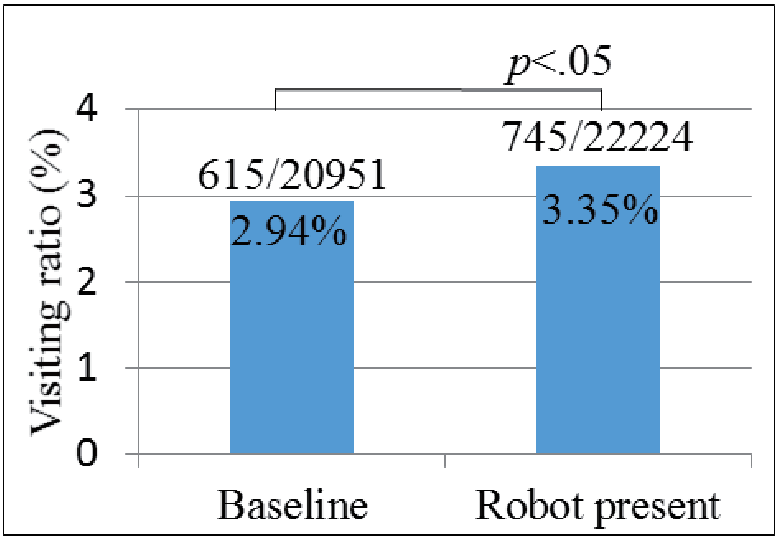


Figure 5.10 Ratio of *visited* passersby for donut shop

5.2.6.4. MANAGER'S FEEDBACK

The manager wanted to use the robot again. While he was not sure whether the robot contributed to the sales, he admitted that it attracted many passersby and advertised his shop. Since hundreds of coupons were distributed, he expected that some customers would return to buy donuts with them. When we asked how much he would pay, he said that he would pay as much as a human worker is paid.

He also mentioned a couple of possible improvements. He wanted the robot to have better interactivity to answer easy questions from visitors. According to him, recently many young people do not want to work in the service industry because they are required to communicate with visitors in a face-to-face manner. He wants a robot to be the store's main clerk and handle all communications; the human clerk will only provide such easy support as replenishing supplies. He also wants the robot to express itself more so that visitors will perceive it to be more than just a machine.

5.3. DISCUSSION

Regarding the problem that participants could not get proper distance cue during sound sources changed its range in previous experiment, a possible reason could be that HRTFs we used was measured in a fixed range (1.4m). When the sound sources moved, the amplitude of HRTFs is recalculated to fall off in inverse proportion to the distances, and ITD (interaural time difference) remains the same based on plane wave assumption. However, the fact is when the range of sound source changed, both ILD (interaural level difference) and ITD will change as well (ILD increased with decreasing distance because of the decreasing of head scattering effect, ITD decreased with decreasing distance), especially at close distance that curvature of the wave front become significant.

5.3.1. IMPLICATIONS

Our study revealed that store managers have serious interest in using robots in their stores. Even after seeing the limitations of the capability of today's autonomous robots, they expressed their desire for future use. In fact, for two stores, the robot increased passerby's frequency of visit. Frequency increased from 2.40% to 3.22% for the cake store, 2.94% to 3.35% for the donut store. This is rather impressive. Even though we expected that the robot would increase the number of people who stopped around the robot (in fact, people who stopped around stores largely increased for all the stores), persuading people to pay attention to something other than the robot itself (in our case, store products) is more difficult. One would concern that the amounts of increase are small; but, we consider this impactful to the stores, as the ratio of visit in the baseline is also not so large. We also believe that the store managers are already doing their best to attract passersby, e.g., signboards, music, and the labor of employees. Since robots undoubtedly add new value to such efforts on their current business, they found them useful.

We were also impressed to learn that the store managers expressed a willingness to pay a considerable amount for a robot's services, perhaps equal to that of a human worker. The services we prepared were all autonomous. The invitation service can be done with many commercially available robots. Distribution services need some improvements from such robots, but they could be done rather easily. Thus, as the price of robot hardware becomes more affordable, we predict that such a robot will widely spread in the near future.

We learned the following design implications from our case studies. First, the store managers were very interested in indicating the relationship between the robot and their stores. They wanted it to wear costumes. Unfortunately, our robot was not well prepared for this, but a future robot could be designed for easily changing its clothing or appearance. Second, the managers made a rather small number of requests, mostly about the contents and the timing of utterances. Perhaps we could prepare a robot that can be customized by the store managers themselves. Interactivity is one avenue for potential improvements.

5.3.2. Social acceptance from administrators

Regarding perceptions related to social acceptance from managers, we made some interesting observations. Their acceptance can be different from that by interlocutors especially when children are involved. The robot often attracted children to interact with it, and thus families often stopped near the stores. Some managers welcomed it; the drugstore manager was happy because it created an opportunity to attract attention to his store. But one manager worried that the presence of a crowd of children might change her store's atmosphere, and so she was discouraged from using the robot, although the children are willing to come.

We observed a kind of a ripple effect, i.e., decisions on acceptance were transmitted across society. Managers see other stores that use robots, and we also noticed that they communicate about the robot among themselves, which influenced their decisions about how to choose appearances and voices. In addition, we were contacted by another manager who wanted to use the robot for his store after seeing it used at another store. This ripple affect suggests that acceptance will be quickly shared, and once some stores start to use robots, others will quickly follow and employ them, too.

5.3.3. Novelty effect

Novelty remains one major reason why managers want to use robots. But novelty is ephemeral. Will they continue to use them for many weeks and months? This open question is beyond the focus of a single study. However, we speculate that strategies exist for keeping their robot attractive; a store that creates its own character like Mickey Mouse might attract visitors for a long time. Further, with robots that express themselves more, people might start to form a kind of relationship with the robots over a long-term use, which would also compensate the loss of novelty.

5.3.4. COMPARISONS WITH HUMAN

Since our robot successfully attracted passersby, comparing its effect with humans seems logical. Perhaps if we replaced it with a person, a similar effect would occur. However, note that although such an effect might be obtained with human workers, the store managers did not make this choice. This is probably relevant to their interview answers that argued for the uniqueness of the robot's value. For instance, they believed that robots attract the attention of customers well without causing/receiving stress. Some managers complained about the difficulty of finding people who are willing to do such services. Thus, while it is possible that humans may cause similar effect in attracting passerby, in reality humans were not alternative choice for the store managers.

5.3.5. CONTEXT DEPENDENCY

Although the frequency of *stop* increased in all three stores, the store visitors increased only for the cake and donut shops but not for the drugstore. We consider that this result is due to the difference of the nature of their stores. These two stores are designed for incidental visits, where a passerby just drops in without any previous intention to visit or buy. In contrast, as the manager himself mentioned, since drugstores are mainly for people who aim to visit, passersby might less incidentally visit even though the robot tried to attract them. Further investigation might improve our knowledge about designing robots for various contexts, e.g., a large store or a restaurant.

5.3.6. LIMITATIONS

Our observations are mostly from case studies that only involve a specific robot, stores, and people. Comparisons include factors beyond the presence of robots. For instance, in the case of the donut shop, although the store visitors increased, that was partly due to the fact that they discounted their product. We did not ask them not to do so because we wanted to observe how they would naturally use the robot. Regarding social acceptance, it would be also important to study how interlocutors perceived; however, we did not interview passerby and visitors, because we did not want to influence their behavior due to the presence of interviewers around the store. As far as we observe, people seemed happy and we did not hear any complaint. We relied on interviews with the managers. But the opinions in the interviews before the case studies (e.g., some were tolerant of use by others) would change after active usages. Their feedback after the trial use might be positively biased (e.g., how much they would pay) because perhaps they were being polite. Nevertheless, it was not our intention to accurately measure their attitudes; we wanted to roughly understand their views. Also, we don’t believe that they consciously distorted their opinions out of a misplaced sense of kindness. Although a majority responded, since the participating managers were self-selected, their positive views do not mean all of the remaining managers are positive. Perhaps the people who did not respond experienced hesitation, reluctance, or negative attitudes.

6. CONCLUSIONS

The present work explored the problem of natural initiation of interaction in human-robot interaction focused on initiating conversation and distributional handing behaviors.

For initiation of conversation, the contribution that makes this possible is a clear set of guidelines for how to structure a robot's behavior to start and maintain a conversation. This knowledge can be used by designers to create robots capable of engaging in a conversation with a person, possibly toward integrating robots into domestic and public environments.

More specifically, we first studied natural interaction at the moment of initiating conversation. In a shopkeeper scenario where a salesperson meets a customer, we then modeled natural human interaction. Our model was implemented in a humanoid robot and tested in an evaluation experiment. We compared our proposed model with two baseline models. The experimental results verified our proposed model as the best with respect to its more appropriate behaviors and the smallest time delay. The recognition accuracy of the participation state in the system evaluation was high, showing that the model can be used to recognize an individual's participation state in a conversation.

For distributional handing, we studied this behavior in which a *giver* distributes flyers to pedestrians in an actual shopping mall environment. Our approach developed a behavior model from the natural interaction of humans. We found that a person who distribute flyers well approaches pedestrians from their front side and only extends his/her arm when they are near. We also found that pedestrians more frequently accepted flyers when the handing was *continuous*, meaning that the distribution targeted a pedestrian who noticed that a previous pedestrian took the flyer. We modeled and implemented these two factors in our humanoid robot and conducted an evaluation experiment a real shopping mall, where the developed robot autonomously distributed flyers. This demonstrated that our developed robot successfully performed a flyer-distributing service. The flyer-distributing efficiency reached 0.18, meaning that it successfully gave flyers to 18% of the pedestrians, which was significantly better than a simple robot that waits for pedestrians to take flyers. We believe that this ratio is reasonably high. The pedestrians in this study were going through a shopping mall and are typically busy with other purposes. Flyer-distribution service is one possible future role in which a robot might serve. It is important that developed robots can successfully operate in real world environments autonomously and with real pedestrians.

Furthermore, we conducted a field study to investigate the social acceptance of social robots by stores, particularly for attracting passersby, which today's robot can autonomously perform. From interviews with ten store managers, we identified two main reasons they want to employ such social robots in their stores:

1. Robots offer cheap labor and provide unique value that humans cannot.

2. They believe that robots are good at attracting the attention of visitors without causing or receiving stress.

We also conducted three case studies in which we observed how store managers employed social robots in their store and found:

1. Social acceptance: Each store manager requested different designs and services. But all of them want to show the connection between the robot and their shop such as dressing the robot with their shop's clothes.

2. Robot could autonomously perform as managers designed: In all the three shops, the managers were satisfied with the result that much more passersby stopped by their stores thanks to the robot.

3. For two out of three stores the robot successfully encouraged visitors to visit.

4. The store managers were satisfied with the results and expressed a desire to use the robots again. In addition, two store managers mentioned that they would like to employ such robots with same wage with human clerks. One mentioned that she would consider to employ such robots with cheaper wage than human clerks.

ACKNOWLEDGEMENT

At the end of this thesis, I am using this opportunity to thank everyone who made my Ph.D. project possible and an unforgettable experience.

First and foremost I want to thank my advisor Prof. Hiroshi Ishiguro, for giving me chance to explore such an exciting and challenging topic and the guidance necessary to make headway. He has taught me, both consciously and unconsciously, how good research is done. I appreciate all his contributions of time, suggestions, and funding to make my Ph.D. experience stimulating.

I would like to express my appreciation and thanks to my supervisor in ATR Dr. Takayuki Kand, for his brilliant ideas and suggestions. I am grateful for the excellent examples he has provided as a successful researcher and mentor.

I also wish to express sincere appreciation to my thesis committee members, Prof. Koh Hosoda and Prof. Kensuke Harada, for taking time to review my thesis and giving helpful comments.

I acknowledge my gratitude to my colleagues in ATR, Dr. Satoru Satake, Dr. Masahiro Shiomi and Deneth Kurunarathne, for their help in sensor setting and experimental data collection.

Finally, I wish to express my deepest thanks to my family who has always supported me.

BIBLIOGRAPHY

[1] Clark H. H., *Using Language*, Cambridge University Press, 1996.

[2] Kendon A., Spatial Organization in Social Encounters: the F-formation System, in *Conducting Interaction: Patterns of Behavior in Focused Encounters*, A. Kendon ed., Cambridge University Press, pp. 209-238, 1990.

[3] Kuzuoka H., Suzuki Y., Yamashita J. and Yamazaki K., Reconfiguring Spatial Formation Arrangement by Robot Body Orientation, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp. 285-292, 2010.

[4] Michalowski M. P., Sabanovic S. and Simmons R., A Spatial Model of Engagement for a Social Robot, *IEEE International Workshop on Advanced Motion Control*, pp. 762-767, 2006.

[5] Shiomi M., Kanda T., Ishiguro H. and Hagita N., A Larger Audience, Please! - Encouraging people to listen to a guide robot -, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp. 31-38, 2010.

[6] Sidner C. L., Kidd C. D., Lee C. and Lesh N., Where to Look: A Study of Human-Robot Engagement, *International Conference on Intelligent User Interfaces (IUI 2004)*, pp. 78-84, 2004.

[7] H.-M. Gross, et al., Shopbot: Progress in Developing an Interactive Mobile Shopping Assistant for Everyday Use, *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC2008)*, pp. 3471-3478, 2008.

[8] W. Burgard, et al., The Interactive Museum Tour-Guide Robot, *National Conf. on Artificial Intelligence (AAAI1998)*, pp. 11-18, 1998.

[9] S. Thrun, et al., Minerva: A Second-Generation Museum Tour-Guide Robot, *IEEE Int. Conf. on Robotics and Automation (ICRA1999)*, pp. 1999-2005, 1999.

[10] Y. Kuno, et al., Museum Guide Robot Based on Sociological Interaction Analysis, *ACM Conference on Human Factors in Computing Systems (CHI2007)*, pp. 1191-1194, 2007.

[11] G. Ferri, et al., Dustcart, an Autonomous Robot for Door-to-Door Garbage Collection:From Dustbot Project to the Experimentation in the Small Town of Peccioli, *IEEE Int. Conf. on Robotics and Automation (ICRA2011)*, pp. 655-660, 2011.

[12] A. Weiss, et al., Robots Asking for Directions: The Willingness of Passers-by to Support Robots, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp. 23-30, 2010.

[13] R. Kirby, J. Forlizzi and R. Simmons, Affective Social Robots, *Robotics and Autonomous Systems*, vol. 58, pp. 322-332, 2010.

[14] B. Mutlu and J. Forlizzi, Robots in Organizations: The Role of Workflow, Social, and Environmental Factors in Human-Robot Interaction, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2008)*, pp. 287-294, 2008.

[15] J. Pineau, M. Montemerlo, M. Pollack, N. Roy and S. Thrun, Towards Robotic Assistants in Nursing Homes: Challenges and Results, *Robotics and Autonomous Systems*, vol. 42, pp. 271-281, 2003.

[16] A. M. Sabelli, T. Kanda and N. Hagita, A Conversational Robot in an Elderly Care Center: An Ethnographic Study, *ACM/IEEE int. Conf. on Human-Robot Interaction (HRI2011)*, pp. 37-44, 2011.

[17] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro and N. Hagita, A Communication Robot in a Shopping Mall, *IEEE Transactions on Robotics*, vol. 26, pp. 897-913, 2010.

[18] Kuzuoka H., Oyama S., Yamazaki K., Suzuki K. and Mitsuishi M., GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions, *ACM Conference on Computer-supported cooperative work (CSCW2000)*, pp. 155-162, 2000.

[19] Scassellati B. M., Foundations for a Theory of Mind for a Humanoid Robot, ed: Massachusetts Institute of Technology, 2001.

[20] Breazeal C., Kidd C. D., Thomaz A. L., Hoffman G. and Berlin M., Effects of nonverbal communication on efficiency and robustness in human-robot teamwork, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2005)*, pp. 383-388, 2005.

[21] Kuno Y., Sadazuka K., Kawashima M., Yamazaki K., Yamazaki A. and Kuzuoka H., Museum Guide Robot Based on Sociological Interaction Analysis, *ACM Conference on Human Factors in Computing Systems (CHI2007)*, pp. 1191-1194, 2007.

[22] Mutlu B., Forlizzi J. and Hodgins J., A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior, *IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids'06)*, pp. 518-523, 2006.

[23] Mutlu B., Shiwa T., Kanda T., Ishiguro H. and Hagita N., Footing In Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2009)*, pp. 61-68, 2009.

[24] Nakano Y. I. and Ishii R., Estimating User's Engagement from Eye-gaze Behaviors in Human-Agent Conversations, *International Conference on Intelligent User Interfaces*, pp. 139-148, 2010.

[25] Rich C., Ponsler B., Holroyd A. and Sidner C. L., Recognizing Engagement in Human-Robot Interaction, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp. 375-382, 2010.

[26] Mondada L., Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space, Journal of Pragmatics, Vol. 41, pp 1977-1997, 2009.

[27] Hall E. T., The Hidden Dimension: Man's Use of Space in Public and Private, The Bodley Head Ltd., 1966.

[28] Goffman E., Behavior in public place: Notes on the Social Organization of Gatherings, The Free Press, 1963.

[29] Kendon A., Features of the structural analysis of human communicational behavior, in *Aspects of Nonverbal Communication*, W. v. R. Engel ed., 1980.

[30] Loth S., Huth K. and De Ruiter J.P., Automatic detection of service initiation signals used in bars, Front Psychol., 2013.

[31] Katagiri Y., Bono M. and Suzuki N., Conversational Inverse Information for Context-Based Retrieval of Personal Experiences, *Lecture Notes in Computer Science*, vol. 4012, pp. 365-376, 2006.

[32] Hüttenrauch H., Eklundh K. S., Green A. and Topp E. A., Investigating spatial relationships in human-robot interactions, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2006)*, pp. 5052-5059, 2006.

[33] Dautenhahn K., Walters M. L., Woods S., Koay K. L., Nehaniv C. L., Sisbot E. A., Alami R. and Siméon T., How May I Serve You? A Robot Companion Approaching a Seated Person in a Helping Context, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2006)*, pp. 172-179, 2006.

[34] Torta E., Cuijpers R.H., Juola J.F., and Pol D.V.D., Design of Robust Robotic Proxemic Behaviour, ICSR, pp.21-30, 2011.

[35] Satake S., Kanda T., Glas D. F., Imai M., Ishiguro H. and Hagita N., How to Approach Humans?: Strategies for Social Robots to Initiate Interaction, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2009)*, pp. 109-116, 2009.

[36] Carton D., Turnwald A., Wollherr D. and Buss M.., Proactively Approaching Pedestirans with an Autonomous Mobile Robot in Urban Environments, *The 13th International Symposium on Experimental Robotics,* pp 199-214, 2013.

[37] Ciolek T. M., and Kendon A., Environment and the Spatial Arrangement of Conversational Encounters, Sociological Inquiry, Vol. 50, pp. 237–271, 1980.

[38] Yamaoka F., Kanda T., Ishiguro H. and Hagita N., A Model of Proximity Control for Information-Presenting Robots, *IEEE Transactions on Robotics*, vol. 26, pp. 187-195, 2010.

[39] Pacchierotti E., Christensen H. I. and Jensfelt P., Evaluation of Passing Distance for Social Robots, *IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN2006)*, pp. 315-320, 2006.

[40] Gockley R., Forlizzi J. and Simmons R., Natural Person-Following Behavior for Social Robots, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2007)*, pp. 17-24, 2007.

[41] Weiss A., Mirnig N., Buchner R., Förster F. and Tscheligi M., Transferring Human-Human Interaction Studies to HRI Scenarios in Public Space, INTERACT (2), pp.230-247, 2011.

[42] Bergström N., Kanda T., Miyashita T., Ishiguro H. and Hagita N., Modeling of Natural Human-Robot Encounters, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2008)*, pp. 2623-2629, 2008.

[43] C. C. Kemp, A. Edsinger and E. Torres-Jara, Challenges for Robot Manipulation in Human Environments, *IEEE Robotics & Automation Magazine*, vol. 14, pp. 20-29, 2007.

[44] A. Sorokin, D. Berenson, S. S. Srinivasa and M. Hebert, People Helping Robots Helping People: Crowdsourcing for Grasping Novel Objects, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2010)*, pp. 2117-2122, 2010.

[45] M. Ralph and M. A. Moussa, An Integrated System for User-Adaptive Robotic Grasping, *IEEE Transactions on Robotics*, pp. 1-12, 2010.

[46] S. Kajikawa, T. Okino, K. Ohba and H. Inooka, Motion Planning for Hand-over between Human and Robot, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS1999)*, pp. 193-199, 1995.

[47] S. Shibata, K. Tanaka and A. Shimizu, Experimental Analysis of Handing Over, *IEEE Int. Workshop on Robot and Human Interactive Communication (RO-MAN1995)*, pp. 53-58, 1995.

[48] A. Agah and K. Tanie, Human Interaction with a Service Robot: Mobile-Manipulator Handing over an Object to a Human, *IEEE Int. Conf. on Robotics and Automation (ICRA1997)*, pp. 575-580, 1997.

[49] M. Huber, M. Rickert, A. Knoll, T. Brandt and S. Glasauer, Human-Robot Interaction in Handing-over Tasks, *IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN2008)*, pp. 107-112, 2008.

[50] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler and J. Forlizzi, Using Spatial and Temporal Contrast for Fluent Robot-Human Hand-Overs, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI 2011)*, pp. 489-496, 2011.

[51] M. Cakmak, S. S. Srinivasa, M. K. Lee, J. Forlizzi and S. Kiesler, Human Preferences for Robot-Human Hand-over Configurations, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2011)*, pp. 1986-1993, 2011.

[52] E. A. Sisbot, L. F. Marin-Urias, R. Alami and T. Simeon, A Human Aware Mobile Robot Motion Planner, *IEEE Transactions on Robotics*, vol. 23, pp. 874-883, 2007.

[53] E. A. Sisbot, A. Clodic, R. Alami and M. Ransan, Supervision and Motion Planning for a Mobile Manipulator Interacting with Humans, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2008)*, pp. 327-334, 2008.

[54] E. A. Sisbot, R. Ros and R. Alami, Situation Assessment for Human-Robot Interactive Object Manipulation, *IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN2011)*, pp. 15-20, 2011.

[55] M. P. Michalowski, S. Sabanovic and R. Simmons, A Spatial Model of Engagement for a Social Robot, *IEEE Int. Workshop on Advanced Motion Control*, pp. 762-767, 2006.

[56] N. Bergström, T. Kanda, T. Miyashita, H. Ishiguro and N. Hagita, Modeling of Natural Human-Robot Encounters, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2008)*, pp. 2623-2629, 2008.

[57] K. Dautenhahn, et al., How May I Serve You? A Robot Companion Approaching a Seated Person in a Helping Context, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2006)*, pp. 172-179, 2006.

[58] S. Satake, et al., How to Approach Humans?: Strategies for Social Robots to Initiate Interaction, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2009)*, pp. 109-116, 2009.

[59] C. Shi, M. Shimada, T. Kanda, H. Ishiguro and N. Hagita, Spatial Formation Model for Initiating Conversation, *Robotics: Science and Systems Conference (RSS2011)*, 2011.

[60] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox*, et al.*, "MINERVA: A Second-Generation Museum Tour-Guide Robot," *IEEE Int. Conf. on Robotics and Automation (ICRA1999)*, 1999.

[61] H. Kuzuoka, S. Oyama, K. Yamazaki, K. Suzuki, and M. Mitsuishi, "GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions," *ACM Conf. on Computer-supported cooperative work (CSCW2000)*, 2000.

[62] R. Siegwart, K. O. Arras, S. Bouabdallah, D. Burnier, G. Froidevaux, *et al.*, "Robox at Expo.02: A Large Scale Installation of Personal Robots," *Robotics and Autonomous Systems,* vol. 42, pp. 203-222, 2003.

[63] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki, and H. Kuzuoka, "Museum Guide Robot Based on Sociological Interaction Analysis," *ACM Conf. on Human Factors in Computing Systems (CHI2007)*, 2007.

[64] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, "Interactive Humanoid Robots for a Science Museum," *IEEE Intelligent Systems,* 22, pp. 25-32, 2007.

[65] R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Robotics and Autonomous Systems,* vol. 58, pp. 322-332, 2010.

[66] M. K. Lee, J. Forlizzi, S. Kiesler, P. Rybski, J. Antanitis, and S. Savetsila, "Personalization in HRI: A longitudinal field experiment," *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2012)*, 2012.

[67] T. Belpaeme, P. E. Baxter, R. Read, R. Wood, H. Cuayáhuitl, B. Kiefer*, et al.*, "Multimodal child-robot interaction: Building social bonds," *Journal of Human-Robot Interaction,* vol. 1, pp. 33-53, 2012.

[68] F. Tanaka and S. Matsuzoe, "Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning," *J. of Human-Robot Interaction,* vol. 1, pp. 78-95, 2012.

[69] A. Pereira, C. Martinho, I. Leite, and A. Paiva, "iCat, the chess player: the influence of embodiment in the enjoyment of a game," in *Proceedings of the 7th Int. joint Conf. on Autonomous agents and multiagent systems-Volume 3*, 2008, pp. 1253-1256.

[70] H.-M. Gross, H.-J. Boehme, C. Schroeter, S. Mueller, A. Koenig, C. Martin*, et al.*, "ShopBot: progress in developing an interactive mobile shopping assistant for everyday use," *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC2008)*, 2008.

[71] M. Svenstrup, T. Bak, O. Maler, H. J. Andersen, and O. B. Jensen, "Pilot Study of Person Robot Interaction in a Public Transit Space," *Research and Education in Robotics — EUROBOT 2008,* vol. 33, pp. 96-106, 2009.

[72] A. Weiss, J. Igelsböck, M. Tscheligi, A. Bauer, K. Kühnlenz, D. Wollherr*, et al.*, "Robots Asking for Directions: The Willingness of Passers-by to Support Robots," *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, 2010.

[73] M. Shiomi, K. Shinozawa, Y. Nakagawa, T. Miyashita, T. Sakamoto, T. Terakubo*, et al.*, "Recommendation effects of a social robot for advertisement-use context in a shopping mall," *Int. Journal of Social Robotics,* vol. 5, pp. 251-262, 2013.

[74] A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kühnlenz, and M. Buss, "A Methodological Variation for Acceptance Evaluation of Human-Robot Interaction in Public Places," *IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN2008)*, 2008.

[75] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model," *Int. Journal of Social Robotics,* vol. 2, pp. 361-375, 2010.

[76] E. Broadbent, R. Stafford, and B. MacDonald, "Acceptance of Healthcare Robots for the Older Population: Review and Future Directions," *Int. Journal of Social Robotics,* vol. 1, pp. 319-330, 2009.

[77] P. Salvini, C. Laschi, and P. Dario, "Design for Acceptability: Improving Robots' Coexistence in Human Society," *Int. Journal of Social Robotics,* vol. 2, pp. 451-460, 2010.

[78] B. Mutlu and J. Forlizzi, "Robots in Organizations: The Role of Workflow, Social, and Environmental Factors in Human-Robot Interaction," *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2008)*, 2008.

[79] Yamazaki K., Kawashima M., Kuno Y., Akiya N., Burdelski M., Yamazaki A. and Kuzuoka H., Prior-to-request and request behaviors within elderly day care: Implications for developing service robots for use in multiparty settings, *European Conference on Computer Supported Cooperative Work (ECSCW2007)*, pp. 61-78, 2007.

[80] Cohen J., A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46, 1960.

[81] Shi C., Kanda T., Shimada M, Yamaoka F., Ishiguro H. and Hagita N., Easy development of communicative behaviors in social robots, *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2010)*, pp. 5302-5309, 2010.

[82] Hartnett J.J., Bailey K.G. and Hartley C.S., Body Height, Position, and Sex as Determinants of Personal Space, Journal of Psychology, 1974.

[83] Woods S.N., Dautenhahn K., Schulz J., Child and adults' perspectives on robot appearance, Proceedings of the Symposium on Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction , pp. 126-132, 2005.

[84] Woods S.N., Dautenhahn K., Schulz J., Exploring the design space of robots: Children's perspectives, Interacting with Computers, Vol 18, No. 6, pp. 1390-1418, 2006.

[85] D. Brscic, T. Kanda, T. Ikeda and T. Miyashita, Person Tracking in Large Public Spaces Using 3d Range Sensors, *IEEE Transaction on Human-Machine Systems*, vol. 43, pp. 522 - 534, 2013.

[86] D. Brscic and T. Kanda, Changes in Usage of an Indoor Public Space: Analysis of One Year of Person Tracking, *IEEE Transactions on Human-Machine Systems*, 2015.

[87] T. Kanda, H. Ishiguro, M. Imai and T. Ono, Development and Evaluation of Interactive Humanoid Robots, *Proceedings of the IEEE*, vol. 92, pp. 1839-1850, 2004.

[88] Masahiro Shiomi, Francesco Zanlungo, Kotaro Hayashi, Takayuki Kanda, Towards a Socially Acceptable Collision Avoidance for a Mobile Robot Navigating Among Pedestrians Using a Pedestrian Model, *International Journal of Social Robotics(IJSR)*, Vol.6, Issue 3 ,pp.443-455, 2014.

[89] C. Shi, M. Shiomi, C. Smith, T. Kanda, and H. Ishiguro, A model of distributional handing interaction for a mobile robot, *Robotics: Science and Systems Conf. (RSS2013)*, 2013.

PUBLICATIONS

Journal Papers

Chao Shi, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, Norihiro Hagita, "Measuring Communication Participation to Initiate Conversation in Human-Robot Interaction", *International Journal of Social Robotics(IJSR)*,7 (5), pp. 889-910, 2015

石超, 佐竹聡, 神田崇行,石黒浩, "客引きロボット導入に向けた社会実験", *日本ロボット学会誌*,（Accepted）

Chao Shi, Satoru Satake, Takayuki Kanda, Hiroshi Ishiguro, "A robot that distributes flyers to pedestrians in a shopping mall", *International Journal of Social Robotics(IJSR)* (In Review)

Satoru Satake, Chao Shi, Takayuki Kanda, "Social Robots that Attract visitors to Stores: Case Studies about How They Are Employed", *International Journal of Social Robotics(IJSR)* (In Submission)

Conference Papers (peer reviewed only)

Chao Shi, Takayuki Kanda, Michihiro Shimada, Fumitaka Yamaoka, Hiroshi Ishiguro, Norihiro Hagita, "Easy development of communicative behaviors in social robots", *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS2010)*, pp. 5302-5309, 2010.

Chao Shi, Michihiro Shimada, Takayuki Kanda, Hiroshi Ishiguro, Norihiro Hagita, "Spatial Formation Model for Initiating Conversation", *The 2011 Robotics: Science and Systems Conference (RSS 2011)*, 2011.

Chao Shi, Masahiro Shiomi, Christian Smith, Takayuki Kanda, and Hiroshi Ishiguro, "A model of distributional handing interaction for a mobile robot", *The 2013 Robotics: Science and Systems Conference (RSS 2013)*, 2013.

Chao Shi, Satoru Satake, Takayuki Kanda, Hiroshi Ishiguro, "How would store managers employ social robots?", *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI2016)*, 2016.