Measuring Communication Participation to Initiate Conversation in Human-Robot Interaction

Chao Shi • Masahiro Shiomi • Takayuki Kanda • Hiroshi Ishiguro • Norihiro Hagita

C. Shi (✉) • M. Shiomi • T. Kanda • H. Ishiguro • N. Hagita
Advanced Telecommunications Research Institute International IRC/HIL
2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan
e-mail: seki85@atr.jp

M. Shiomi
e-mail: m-shiomi@atr.jp

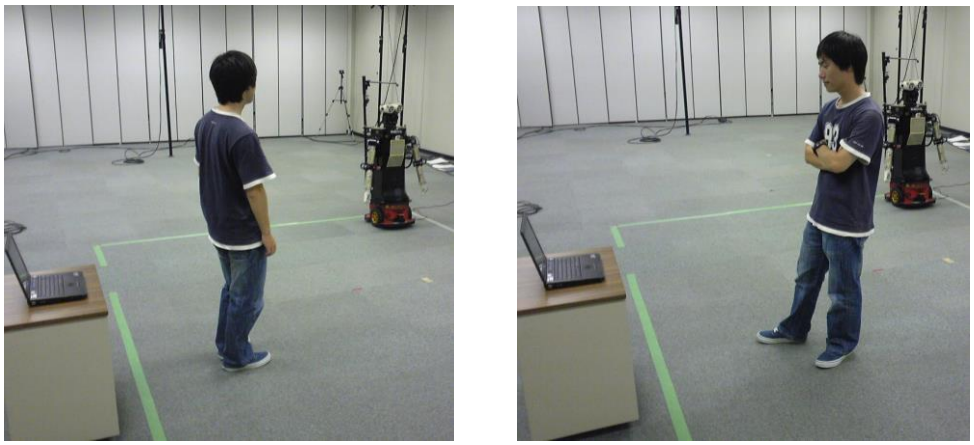
T. Kanda
e-mail: kanda@atr.jp

H. Ishiguro
e-mail: ishiguro@atr.jp

N. Hagita
e-mail: hagita@atr.jp

C. Shi • H. Ishiguro
Department of Systems Innovation, Osaka University, 1-3 Machikaneyama Toyonaka Osaka 560-8531 Japan

Abstract Consider a situation where a robot initiates a conversation with a person. What is the appropriate timing for such an action? Where is a good position from which to make the initial greeting? In this study, we analyze human interactions and establish a model for a natural way of initiating conversation. Our 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. Theories have addressed human communication related to these concepts, but they have only covered situations after people start to talk. In this research, we created 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



(a) Looking at robot (b) Looking at a product

Figure 1 Situations in a shop

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 compared with other two alternative conditions.

Keywords *Behavior modeling • Initiation of interaction • Natural-HRI*

1 Introduction

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 (Fig. 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 (Fig. 1a). Or he/she might look at another product displayed in the shop (Fig. 1b). 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?

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 Fig. 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.

In our previous work [7], 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 then conducted an evaluation experiment to compare our model with two baseline models, and our proposed model was evaluated as the best.

An earlier conference paper first described our approach to initiating a conversation in RSS 2011 [7]. The current article chronicles the whole observation, implementation and evaluation process from start to finish in one place, providing additional details, and offers new evaluation results needed to support our finding that the proposed solution is effective toward initiation of conversation.

In this paper, we build on this previous work in providing more detailed analysis, discussions and additional evaluations. Firstly, we added a detailed explanation of what exactly the participants performed, in order to explain our model more clearly and allow readers to extend the knowledge that we found from our observations. Secondly, we added a more detailed explanation, which will enable other researchers to reproduce/extend our proposed model. Thirdly, we added a system evaluation and an objective evaluation based on our formal evaluation experiment to further evaluate our model in an objective way. The system evaluation clearly showed how well our model works. And the objective evaluation provides more detailed information to tell what the robot exactly performed and what is different in each condition, and therefore shows the effectiveness of our model more persuasively. Fourthly, we also added further discussion to explain why lab situation settings are used for this research work, instead of using realistic scenarios such as a real field observation.

The rest of this paper is organized as follows. Section 2 describes some related work, and Section 3 describes our approach to modeling people's behavior. Section 4 introduces our platforms and implementation of the model. In addition, we evaluated the model in both subjective and objective evaluations, which are explained in Section 5. Section 6 provides a discussion on the findings, and Section 7 summarizes our contributions.

2 Related Works

2.1 Natural HRI and Engagement

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 [8, 9] and gaze [10-13].

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 [14, 15].

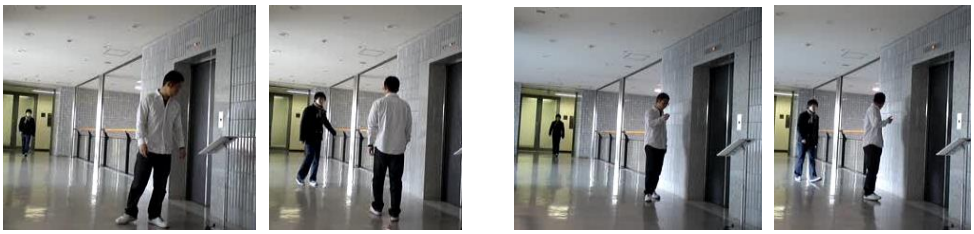
The main difference between the *initiation of conversation* and *engagement* is that the latter addresses a phenomenon that occurs after the people and the robots have established a common belief that they are sharing a conversation. In contrast, the phenomenon of *initiation of conversation*, which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

2.2 Initiating 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 [16], stop at a certain distance [17], start the conversation with a greeting [18, 19], 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 [20] and the finding that side participants stand close to the participants and often become the next participant [21]. 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 [22]. 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 [23, 24] and the path to approach and catch up with a walking person [25, 26], the standing position for presenting a product [27], the proper distances for passing behavior [28] and following behavior [29], and the selection criteria for choosing an

(a) Shop scenario



(b) Meeting scenario

Figure 2 Examples of initial positions in two scenarios

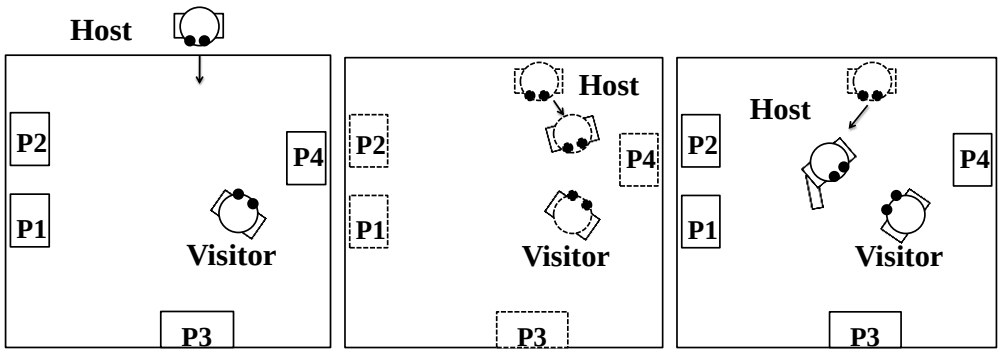
interaction partner [30]. A few studies have attempted to promote people's participation by encouraging behavior [5, 31] and detecting the requested behavior [32]. However, since these studies were aimed at encouraging people's participation, they only showed the one-sided behavior of the robot, not 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.

3 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 [33].

3.1 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*



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

Figure 3 Influence of *subsequent plan* in *initiate position*

of conversation. Based on this assumption, we divided each scenario into two situations.

Shop scenario: This interaction was conducted in a 5 x 5-m room in which four objects were placed (Fig. 2a). 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 (Fig. 2b). 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.

3.2 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 (Fig. 3b); 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* (Fig. 3c). 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*).

3.2.1 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

Table 1 Analysis of initiation behavior

Scenario	Plan	Visibility	Initiation behavior	
			Approaching visitor	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%)

measure the level of agreement between two sets of dichotomous ratings or scores [34]. 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.

The coding results are shown in Table 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 [27]. 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

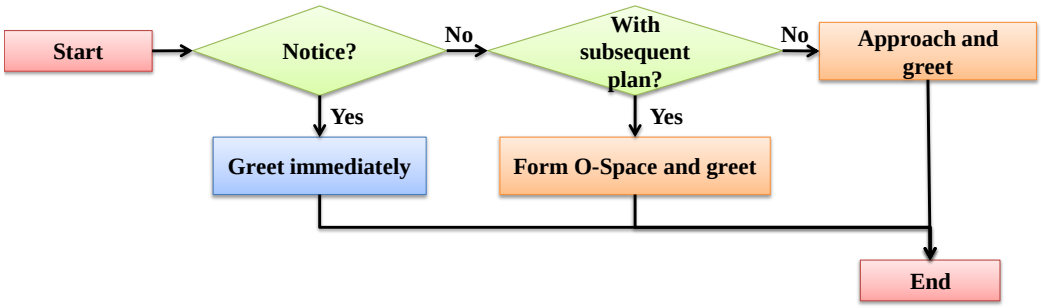


Figure 4 Choice of *initiate timing* and *position*

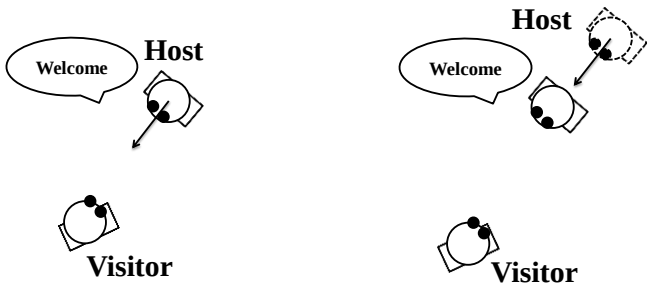
visitors. As shown in Fig. 4, in summary, we found that the choice of *initiation behavior* was influenced by whether the *hosts* had a further plan to explain something to the *visitor*. However, this choice is also influenced by *visibility*. If the *visitor* noticed the *host* within a certain distance, the *host* moved to the *visitor* to initiate the conversation.

3.2.2 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.

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 (Fig. 5a), the *host* greets *visitors* after approaching them (Fig. 5b); 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.



(a) Greet immediately (b) Greet after apporaching

Figure 5 Detailed analysis of *initiation position* in *notice* category

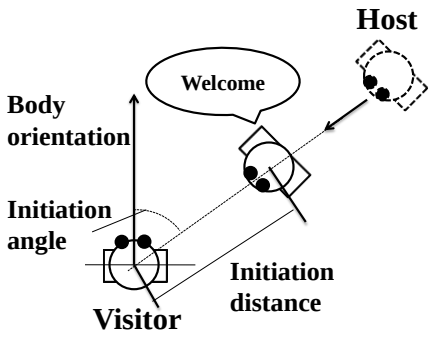


Figure 6 Initiation distance and initiation angle

We further analyzed the position relationships between the *host* and *visitor*. First, we measured the distance (*initiation distance*) and angle (*initiation angle*) (Fig. 6) 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*

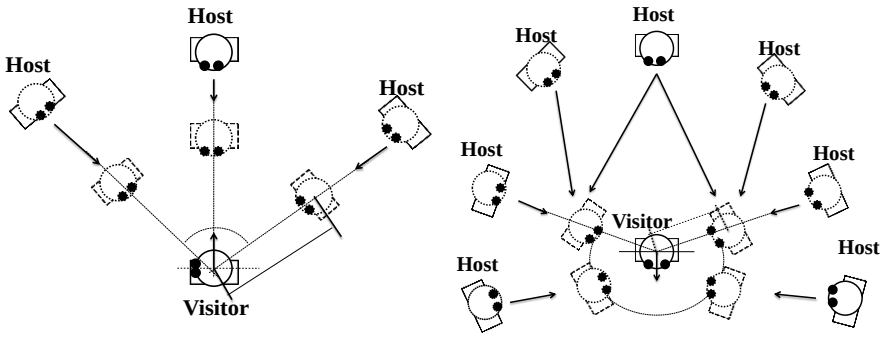
Table 2 Analysis of initiate position (distance and angle) and distance to talk

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

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.

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 Fig. 7a, 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* (Fig. 7b), 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.



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

Figure 7 Detailed analysis of *initiation position* in *unnoticed* category

3.2.3 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.

3.2.4 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 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.

3.2.5 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).

4 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.

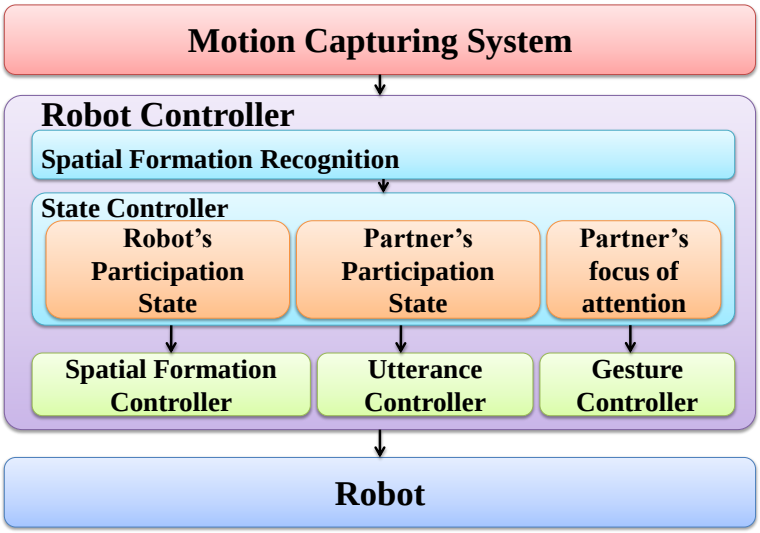


Figure 8 System configuration

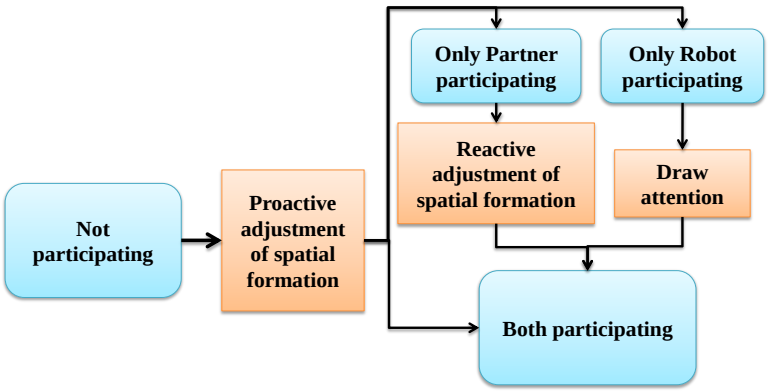


Figure 9 Flow of initiating conversation

4.1 General Framework

We used a development framework that we had used successfully before to control the robot automatically [35]. Figure 8 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 *utteranc*e, and the robot automatically uses them according to the information from the state controller [35].

Figure 9 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*.

4.2 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*.

4.3 Spatial Formation Recognition

4.3.1 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**

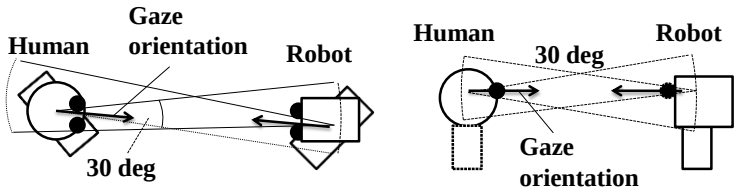
This state variable indicates a situation where only the robot is participating in a conversation with the visitor, i.e., PS_J= (1, 0). Although the robot is considered to be participating in a conversation with the human, the human does not realize that the robot is approaching. In this case, the robot 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**

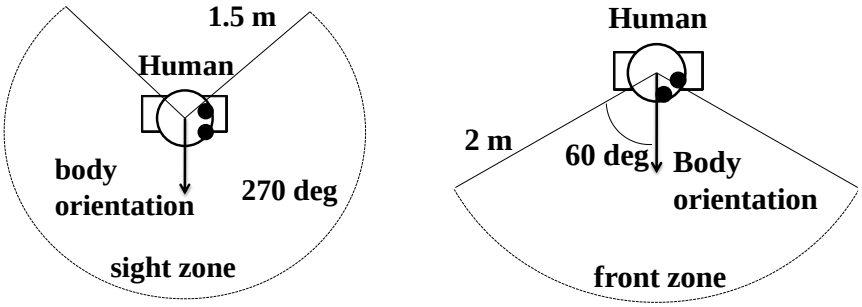
This state variable indicates a situation where only the visitor is participating in a conversation with the robot, i.e., PS_J= (0, 1). This means that only the human is considered to be participating in a conversation with the robot. It is possible that the *visitor* recognizes the robot and wants to 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



(a) gaze zone



(b) sight zone (c) front zone

Figure 10 Participation zone

participating in the conversation with each other, the robot should immediately greet the human.

4.3.2 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.

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 3.5, 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. Fig. 10a 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*:

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

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 3.2.2, 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 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 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 (Fig. 10b).

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 3.2.2, 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 [17], observations reported in Section 3.2 (initiate distance and initiate angle of Approach from frontal in Table 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 (Fig. 10c).

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 3.2.2, 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.

4.3.3 Visitor's focus of attention

As reported in Section 3, 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 [27] 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

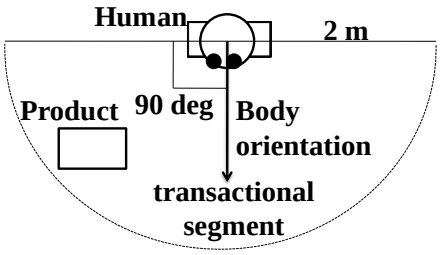


Figure 11 Transactional segment

Table 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)

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 (Fig. 11).

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*:

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

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.

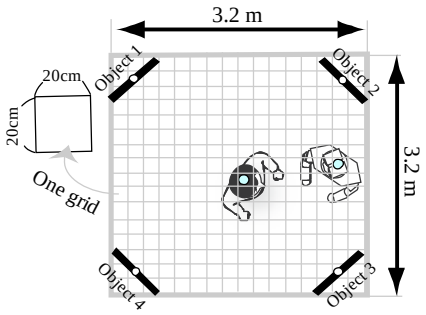


Figure 12 Searching grid

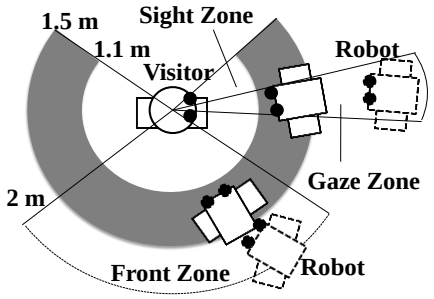


Figure 13 Reactive adjustment of spatial formation

4.3.4 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 shows the relationship among *joint participation state* PS_J and the three *participation zones*.

4.4 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 (Fig. 12). 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.

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**

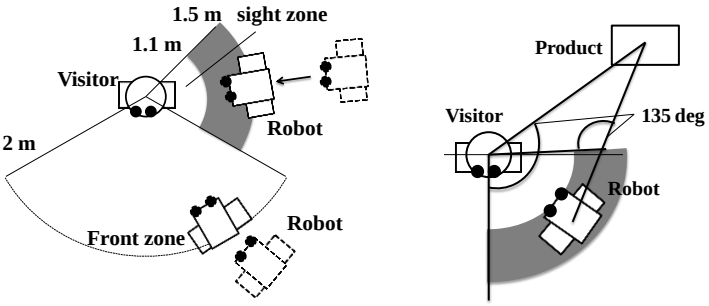


Figure 14 Proactive adjustment of spatial formation

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* (Fig. 13):

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 3.2.2 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 [27).

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.

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

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

where P_T is the position of each cell, as shown in Fig. 10. 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). 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 (Fig. 14):

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.

4.5 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.

5 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 (Fig. 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 3.2.5 and Section 4.3, 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.

5.1 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 5.2. 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 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.

5.2 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 [27] 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:

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

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.

5.3 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 4.2.

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.

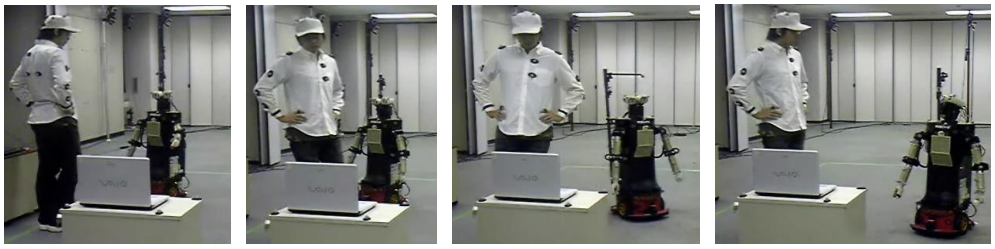
They simulated the behaviors of five types of visitors that decided all by themselves (as a result, the five types of visitors played by each participants are not all the same), such as someone waiting in front of the product or someone at the store entrance, and interacted five times under each condition. 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.

5.4 Measurement

5.4.1 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 4.3 for all of the trials. The two coders that analyzed the data



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

Figure 15 Discontinuing and re-establishing the conversation

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) 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 (Fig. 15a in *joint participation state* PSJ = (1, 1)), but then the visitor moved toward the target product and disrupted the conversation (Fig. 15b, PSJ = (0, 0)). In this case, the robot must reposition itself to adjust the spatial formation (Fig. 15c) so that both are participating in the conversation again (Fig. 15d, PSJ = (1, 1)).

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.

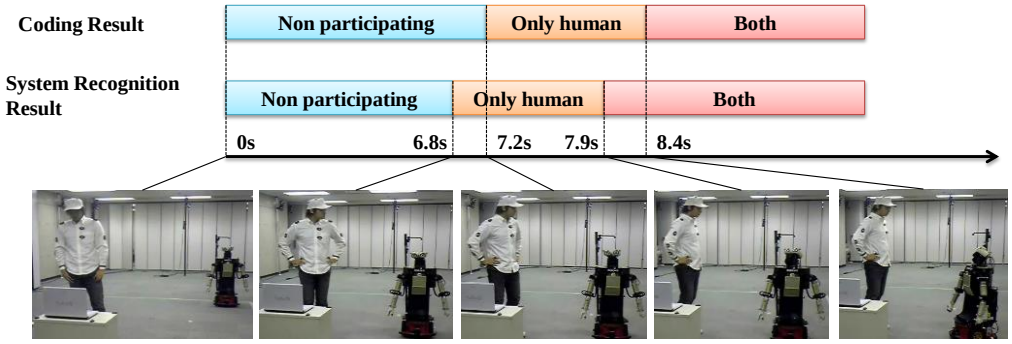


Figure 16 Difference between coding and system recognition result

5.4.2 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.

5.5 Result of System Evaluation

5.5.1 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

$$RecognitionAccuracy = T_{right} / T_{entire} \quad (9)$$

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 (Fig. 16). 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).

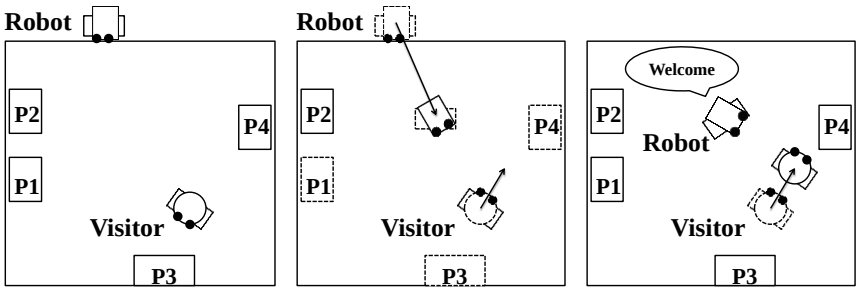
5.5.2 Appropriateness of robot's initiating behavior

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.

In the six trials in which they thought the robot failed to behave appropriately, the robot first approached from the non-frontal direction (Fig. 17a). 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* (Fig. 17b). 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 (Fig. 17c).

5.5.3 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*



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

Figure 17 Inappropriate cases of robot's initiating behavior

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).

5.6 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.

5.6.1 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 (Fig. 18), 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

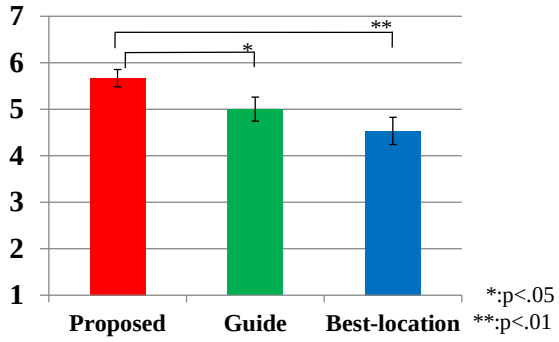


Figure 18 Overall evaluation

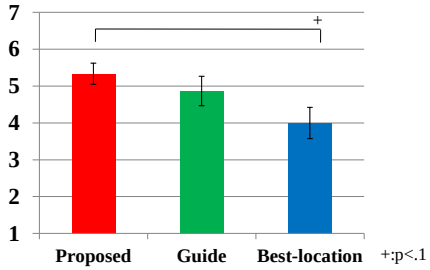


Figure 19 Appropriateness of standing position when it greeted

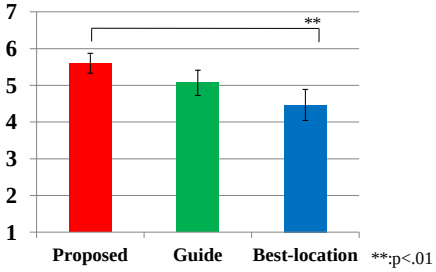


Figure 20 Appropriateness of standing position when it explained

found between the *guide* and *best-location* conditions (p=.5). Therefore, our first prediction was supported.

For "appropriateness of standing position when it greeted" (Fig. 19), a repeated measures analysis of variance revealed a significant main effect ($F(2,28)=4.697$, $p=.017$, partial $\eta 2=.251$), 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" (Fig. 20), a repeated measures analysis of variance revealed a significant main effect ($F(2,28)=9.126$, $p=.001$, partial $\eta 2=.395$). 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).

5.6.2 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.

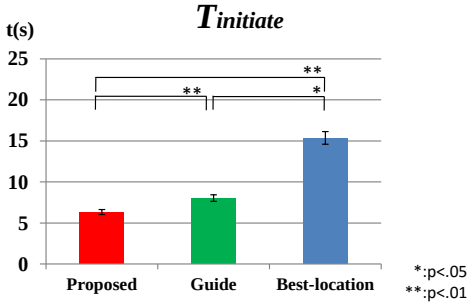


Figure 21 Average of $T_{initiate}$

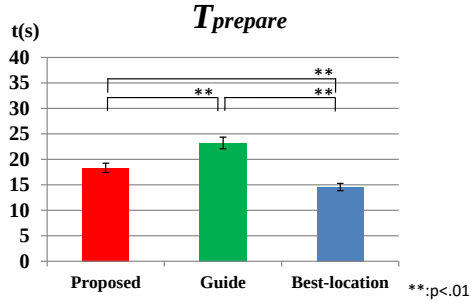


Figure 22 Average of $T_{prepare}$

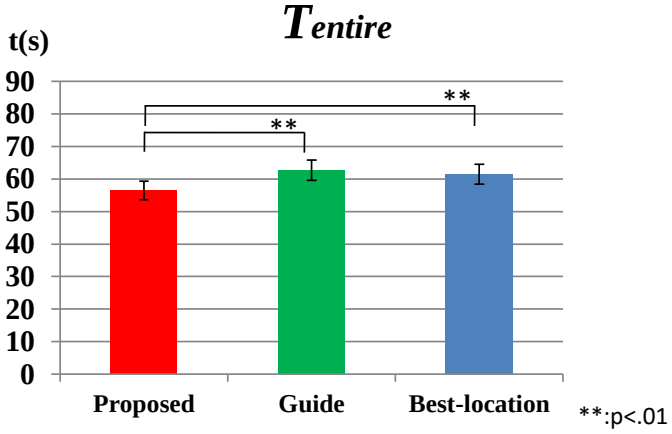


Figure 23 Average of T_{entire}

For our second prediction (Fig. 21), $T_{initiate}$ 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 $\eta 2 = .594$). 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.

5.6.3 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 (Fig. 22), $T_{prepare}$ 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 $\eta 2 = .340$). 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.

5.6.4 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 (Fig. 23), 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 $\eta 2 = .233$). 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.

5.6.5 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.

6 Discussion

6.1 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 [25]. 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.

6.2 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 there are multiple people 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 [36], 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 [37, 38]. 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.

7 Conclusion

In summary, this article reported on how a robot can initiate a conversation with people. 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.

Acknowledgements We'd like to thank everyone who helped with this project. This research was supported by KAKENHI 21118008.

References

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. Shi C., Shimada M., Kanda T., Ishiguro H. and Hagita N., Spatial formation model for initiating conversation, *Proc. Of Robotics: Science and Systems (RSS 2011)*, 2011.
8. 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.
9. Scassellati B. M., Foundations for a Theory of Mind for a Humanoid Robot, ed: Massachusetts Institute of Technology, 2001.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. Hall E. T., The Hidden Dimension: Man's Use of Space in Public and Private, The Bodley Head Ltd., 1966.
18. Goffman E., Behavior in public place: Notes on the Social Organization of Gatherings, The Free Press, 1963.
19. Kendon A., Features of the structural analysis of human communicational behavior, in *Aspects of Nonverbal Communication*, W. v. R. Engel ed., 1980.
20. Loth S., Huth K. and De Ruiter J.P., Automatic detection of service initiation signals used in bars, Front Psychol., 2013.

21. 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.
22. 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.
23. 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.
24. Torta E., Cuijpers R.H., Juola J.F., and Pol D.V.D., Design of Robust Robotic Proxemic Behaviour, ICSR, pp.21-30, 2011.
25. 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.
26. 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.
27. Ciolek T. M., and Kendon A., Environment and the Spatial Arrangement of Conversational Encounters, Sociological Inquiry, Vol. 50, pp. 237–271, 1980.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. Cohen J., A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46, 1960.
35. 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.
36. Hartnett J.J., Bailey K.G. and Hartley C.S., Body Height, Position, and Sex as Determinants of Personal Space, Journal of Psychology, 1974.
37. 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.
38. 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.

Chao Shi received the M. Eng. Degree in engineering science from Osaka University, Osaka, Japan, in 2011.

He is currently an Intern Researcher at Intelligent Robotics and Communication Laboratories (IRC) at the Advanced Telecommunications Research Institute International (ATR), Kyoto. His research interests include human-robot interaction, interactive humanoid robots and field trials.

Masahiro Shiomi received M. Eng and Ph.D. degrees in engineering from Osaka University, Osaka, Japan in 2004 and 2007, respectively.

From 2004 to 2007, he was an Intern Researcher at the Intelligent Robotics and Communication Laboratories (IRC) and is currently a Researcher at IRC at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan. His research interests include human-robot interaction, interactive humanoid robots, networked robots, and field trials.

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

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

Dr. Kanda is a Member of the Association for Computing Machinery, the Robotics Society of Japan, the Information Processing Society of Japan, and the Japanese Society for Artificial Intelligence.

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

In 1991, he was a Research Assistant with the Department of Electrical Engineering and Computer Science, Yamanashi University, Yamanashi, Japan. In 1992, he joined as a Research Assistant with the Department of Systems Engineering, Osaka University. In 1994, he became an Associate Professor with

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

Norihiro Hagita received his B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University in 1976, 1978, and 1986.

In 1978, he joined Nippon Telegraph and Telephone Public Corporation (Now NTT), where he developed handwritten character recognition. He was a visiting researcher in the Department of Psychology, University of California, and Berkeley in 1989-90. He is currently the Board Director of ATR and an ATR Fellow, director of ICT Environment Research Laboratory Group and Intelligent Robotics and communication laboratories. He is a member of IEEE, the Information Processing Society of Japan, and the Japanese Society for Artificial Intelligence, The Institute of Electronics, Information and Communication Engineers (IEICE) and The Robotics Society of Japan.