

Spatial Formation Model for Initiating Conversation

Chao Shi^{1,2} Michihiro Shimada,¹ Takayuki Kanda,¹ Hiroshi Ishiguro,^{1,2} and Norihiro Hagita¹

¹ATR IRC Laboratories

2-2-2 Hikaridai Keihanna Science City
Kyoto 610-0228 Japan

²Osaka University

1-3 Machikaneyama Toyonaka
Osaka 565-8531 Japan

shi.chao@irl.sys.es.osaka-u.ac.jp, {m.shimada, kanda, ishiguro, hagita}@atr.jp

Abstract— In a situation where a robot initiates conversation with a person, when is the appropriate timing and where is an appropriate position from which to say the first greeting word? In this study, we analyze human interaction and establish a model for a natural way of initiating conversation. The model mainly concerns the participation state [1] and spatial formation [2]. When a person is going to participate in a conversation, at a moment when a particular spatial formation occurs, she would feel that she is participating in the conversation; once she perceived her participation she would try to maintain particular spatial formations. There are theories in human communication for these concepts [1, 2], but they only cover the situation after people have started to talk. We build a model that precisely describes the constraints and expected behaviors for the phase of initiating conversation. The proposed model is implemented in a humanoid robot, and it is confirmed as effective in an evaluation experiment based on a shopkeeper scenario.

Keywords—Behavior modeling; initiation of interaction; natural-HRI

I. INTRODUCTION

How do you meet someone and start a conversation? This might be a trivial problem for a person; however, it is not at all trivial for a robot. In a typical situation, one would stop at a certain position toward the target person, speak a greeting word, and find that they are engaging in a conversation. We can easily do this almost unconsciously. What we, as humans, would consciously think about would be the contents of the conversation once it had started.

In contrast, for a robot it is difficult to replicate what humans unconsciously perform. It needs to know every detail of behavior, such as where and when it should stop, and what it should say; but since we do this unconsciously, it is not easy to describe the details of what we are doing. For instance, let us consider a situation in a shop (Fig. 1), where a visiting customer is visiting has an appointment with a salesperson robot to get an explanation of a product. The customer might wait for the robot at the entrance while paying attention to the direction from which the robot is coming (Fig. 1a), or perhaps he might look at a product displayed in the shop (Fig. 1b). The expected behavior for the robot is apparently 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 participation* in natural human-robot interaction. Clark modeled human communication based on the notion that people in a conversation share the view for each other's *participation state* [1], such as a speaker, hearer, and side participants. Kendon's analysis on spatial formation, known as F-formation, overlaps

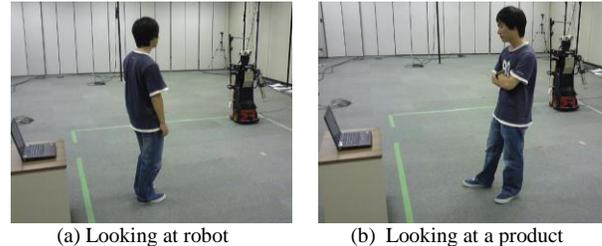


Figure 1 Situations in a shop

this view so that the participants in a conversation form a particular shape [2]. Researchers in HRI well recognize the importance of the participation state and spatial formation [3-6]; however, no study has revealed how a robot should behave to different kinds of initiation interaction depending on the situation (which we refer to as *initiation of participation*). In short, the above examples of the problem in Fig. 1 have not yet been solved.

To cope with this problem, we analyzed human behavior during the *initiation of participation*. It informs us of the importance of two functions in our model:

- recognition of the interlocutor's spatial formation
- constraints on the robot's spatial formation to maintain the participation state

We implemented a model with these two functions and demonstrated the effectiveness with an evaluation experiment.

II. RELATED WORKS

A. Natural HRI and Engagement

It is assumed that social robots may engage in “natural” interaction with humans, i.e., interaction in the same way as humans do with other humans. The use of human-like body properties for robots has been studied for providing greater naturalness in the interaction. Often, studies have focused on the interaction after robots meet people. For instance, studies have been conducted on the use of pointing gestures [7, 8] and gaze [9-12].

Similar to the concept of *initiation of participation*, 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 robots' gaze behavior [6] as well as people's gaze behavior for recognizing the engagement state [13, 14].

The major difference between the *participation state* and *engagement* is that the latter addresses a phenomenon

occurring after people and robots have established a common belief that they share the conversation. In contrast, the phenomenon of *initiation of participation*, which this study addresses, concerns the situation before or just at the moment that they establish the common belief that they are sharing a conversation.

B. Initiating Interaction

In human communication studies, there are not many studies about how humans initiate conversation beyond the basic facts that they stop at a certain distance [15], start the interaction with a greeting [16, 17], share a recognition of each other’s *participation state* [1], and arrange themselves in a spatial formation [2]. Recent studies have started to reveal more detailed interaction, e.g., side participants stand close to the participants and often become the next participant [18], but this new knowledge has so far been limited.

In HRI, spatial formation has been studied in relation to initiating interaction. Michalowski et al. revealed the relation between the robot’s environment and the person’s participation state toward the interaction [4]. Hüttenrauch et al. found that people follow an F-formation in their interacting with robots [19]. Kuzuoka et al. studied the effect of body orientation and gaze in controlling the F-formation [3]. There have also been studies on generating more natural robot behavior, such as the approach direction [20] and path [21], the standing position [22], passing behavior [23], and following behavior [24]. A few studies have attempted to promote people’s participation by such means as encouraging behavior [5, 25] and detecting request behavior [26]. However, these studies were aimed at encouraging people’s participation, and thus they do not show how robots should behave in the *initiation of participation*.

III. MODELING INITIATION OF PARTICIPATION

To find regular patterns in people’s behavior at the moment of *initiation of participation*, we observed two persons’ interaction when they started conversation. We specifically focused on their spatial formation and gaze, which have been discussed as important factors in the literature on human communication [27].

A. Data collection

The data collection was conducted in two different settings, *shop* and *meeting* scenarios, to find consistency and difference across different purposes and environments. In each scenario, we set up a situation where one person initiated conversation with the other. Here, we assumed that whether a participant plans to explain some object or show the way after the initial greeting will somehow influence how that person behaves in the *initiation of participation*. Based on this assumption, we divided each scenario into two situations.

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



(a) Shop scenario



(b) Meeting scenario

Figure 2 Examples of initial positions in two scenarios

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

We set the initial position of the *host* to be out of sight from the *visitor*, and then the *host* entered to the environment to initiate conversation. With this setting, we were able to observe how they behaved both non-verbally and verbally to initiate a conversation.

Twenty paid undergraduate students (ten pairs, eleven men and nine women) participated in this data collection. They repeated each scenario ten times (after five trials, they switched roles, so each participant acted in one role five times in one scenario). We asked the *visitor* to position himself/herself differently every time so that we could collect diverse data. Beyond this instruction, the participants were instructed to behave freely.

B. Data analysis

Participants took diverse spatial formations and behaviors when they initiated conversation. For example, some *hosts* moved straight to the *visitor* and greeted the *visitor* saying, “Welcome, may I help you” in the central area; some moved to the side of the *visitor* and only spoke the first word when they reached a close position to the *visitor*. Aiming to retrieve a systematic pattern in such an *initiation of participation*, our observation focused on the position and timing of the *host*’s

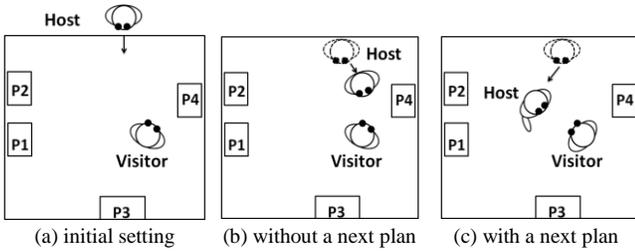


Figure 3 Influence of further plan in initiation position

performance, i.e., (1) where to initiate participation (*initiation position*), and (2) when to initiate participation (*initiation timing*).

In preliminary analysis of how the *host* behaved, we found that their choice of *initiation position* was influenced by whether they had a next plan, such as "guiding the *visitor* to a different location". Figure 3 shows an example. As shown in Fig.3 b, when a *host* only needed to greet the *visitor*, he typically went directly to the *visitor*; but when he needed to introduce a computer P3 to the *visitor*, he tended to go to a position at which the computer P3 would be visible to both of them. We defined the former case as *without plan* and the latter case as *with plan*. We also found that their choice of *initiation timing* was influenced by whether the *visitor* noticed the *host's* arrival. It is intuitive that when the *visitor* notices the *host*, the *host* makes greeting immediately.

To confirm the relationships among three key factors, we coded the situation. We defined the following categories and analyzed the moment when the *host* initiated the conversation:

1. Next Plan

- **With plan:** The *host* had a next plan to explain a product to the *visitor* (in *shop* scenario) or to take the *visitor* to another location (in *meeting* scenario).
- **Without plan:** The *host* did not have such a further plan but only aimed to talk with the *visitor*.

2. Visibility

- **Notice:** Before the *host* began to make an utterance to initiate participation, the *visitor* realized the *host's* arrival first and looked at him/her.
- **Not notice:** The *visitor* did not notice the *host's* arrival before the *host* initiated the participation.

3. Initiation Position (analyzed for *with plan* cases)

- **Closest position:** The *host* went directly to the *visitor*.
- **Position where next target is visible:** The *host* went to the position where the next target (e.g., product or a route for the next location) would be visible to them.

Whether the *host* had a *next plan* was provided by the participants at the data collection. Regarding *visibility* and *initiation position*, two coders conducted coding process, classifying all the data into the above defined categories. Cohen's Kappa coefficient from the two coders' classification was 0.82 for *visibility* and 0.87 for *initiation position*, indicating that their classification were highly consistent.

TABLE 1 DATA ANALYSIS RESULT

Next Plan	Visibility	Initiation position	
		Closest position	Location where next target is visible
With plan	Not notice	11	47
	Notice	39	3
Without plan	Not notice	55	-
	Notice	45	-

The coding result is shown in Table 1, which confirms our observation. We found that when the *visitor* hadn't noticed the *host's* arrival while the *host* had a next plan, the *host* tended to choose the position by considering the next plan (47 cases out of 58). The *host* always moved to the *closest position* to greet to the *visitor* when he don't have next plan, as in the coding coders did not find any exception. Moreover, the *host* tended to greet the *visitor* immediately when the *visitor* noticed to him/her, and when *not noticed*, the *host* typically approached close to the *visitor*. Even if he has the next plan the *host* greeted to the *visitor* when *notice*, thus he/she could hardly have a chance to adjust his/her position for the next plan (39 out of 42). In summary, we found that the choice of *initiation position* was influenced by whether the *host* had a further plan to explain something to the *visitor*. However, the choice of timing takes precedence: If the *visitor* notices the *host*, the *host* initiates the conversation immediately.

We further analyzed the detailed parameters of the spatial formation. In the *notice* category, we found that the distances at which they started to greet were different between the two scenarios. In the *shop scenario*, the largest distance was about 2.5 m (even if they noticed each other at a distance farther than 2.5 m, they did not make an utterance until they reached this distance threshold), while in the *meet scenario* the farthest distance was about 8 m. Thus, we considered this parameter situation-dependent (or environment-dependent).

Regarding the *initiation position* and the words they utter when they start, as shown in Fig. 4a, when the *host* came from the front direction to the *visitor*, he/she typically greeted the *visitor* from a distance of about 2 m. The *host* kept walking toward the *visitor* while greeting until he/she got nearer to the *visitor* within about 1.5 m. This type of interaction happened across an angle within 120 degree from the front, and thus we considered 120 degree the threshold of the front of the *visitor*.

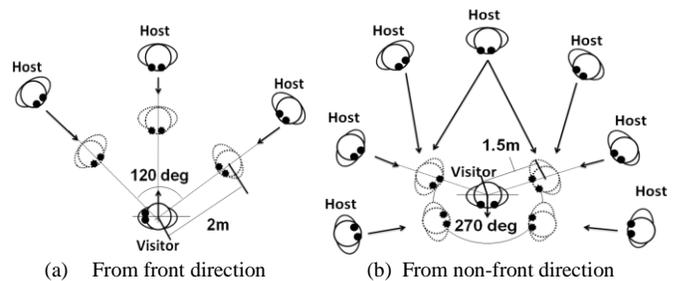


Figure 4 Detailed analysis of initiation position in *without plan*

As shown in Fig. 4b, when the *host* came from a non-frontal direction, he tended to stop near the *visitor* within a distance of about 1.5 m in the *visitor*'s field of view and say an utterance like "Excuse me" to first draw the *visitor*'s attention, and then he greeted the *visitor*. Here, by our observation, a *visitor*'s "field of view" is about 270 deg. in front of the *visitor*'s body.

We found that the utterances the *host* used to initiate the conversation were influenced by the *participation state* of the *visitor*. When the *visitor* noticed the *host*'s arrival, the *host* greeted the *visitor* with an expression like "Welcome." It seemed as if they had already reached an agreement to participate in the conversation. We named this mental agreement as the *participation state*. When the *host* initiated the conversation from a position to the side of the *visitor* without meeting gaze each other, the *host* needed to first draw the *visitor*'s attention. This situation can be considered the *visitor not participating* in the conversation.

We found that the above phenomena were common to both of the scenarios, except for the threshold distance when they start conversation. We concluded that the basic phenomena in initiating conversation were common among scenarios and environments.

IV. A ROBOT THAT ADDRESSES INITIATION PROCESS

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

A. Hardware

We used a humanoid robot that is 1.2 m tall with a 0.3 m radius and characterized by its human-like body expressions. It has a 3-DOF head and 4-DOF arms. Its mobile base is wheel-based.

Since our research focus is to confirm the validity of the model, we used a motion capturing system as the sensor input. The motion capturing 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.

There were 23 markers on the human's body and on the robot's body, and 4 markers were attached to each product.

B. General Framework

Figure 5 shows an outline of our framework. There are three components: a humanoid robot, a motion capturing system, and a robot controller (software).

The *spatial formation recognition* uses input from the motion capturing system and recognizes the spatial formation. The *state controller* receives the information from *spatial formation recognition*, and sends the state information to the *spatial formation controller*, *utterance controller* and *gesture controller*.

Figure 6 shows a robot's flow for *initiation of participation*. There are two paths until the start of

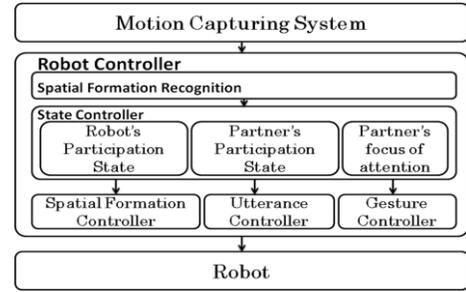


Figure 5 System configuration

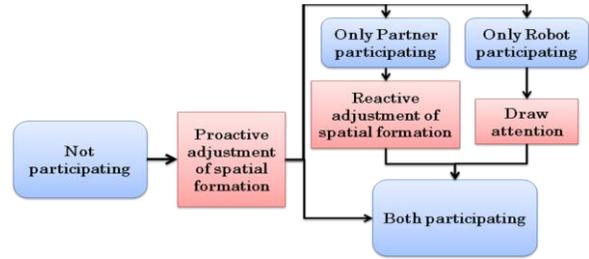


Figure 6 Flow of initiating conversation

conversation. In one case, it is the robot that initiates *participation*. It approaches, stops at appropriate position (*proactive adjustment of spatial formation*), and with a *draw attention* action it elicits the customer to participate to the conversation.

In the other case, it is the customer who initiates the conversation. That is, while the robot is moving to a certain position (for *proactive adjustment of spatial formation*), the customer takes an action to initiate the conversation. Thus, the customer's participation state transitions 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*.

C. Status

There are three state variables involved in the implementation.

- **Partner's participation state**

This state variable indicates an estimation of whether the partner perceives that he/she is participating in the conversation.

- **Robot's participation state**

This state variable indicates whether the robot is considered to be participating. This state is expected to match the partner's *participation state*. Thus, when these states mismatch, the robot should immediately try to resolve the conflict. The opposite case could happen where the robot is already participating (e.g., greeting the partner) but the partner is not yet participating (e.g., not recognizing the robot's presence or ignoring its greeting).

- **Partner's focus of attention**

This state variable represents an estimation of whether the partner is paying attention to an object.

D. Recognition

There are mainly two types of estimation conducted by interpreting spatial formation (including a gaze).

- **Estimation of the partner’s focus of attention**

We used a method reported in [22], which 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 a person when there is no obstacle between him/her and the object. When the angle between the forward direction of the people’s body and the vector from the people’s body center to an object is less than 90 degrees, and the distance between the people and the object is less than 2 m, the object is identified as the people’s implicit attentional target (Fig. 7).

- **Method of estimating participation state of the partner and the robot**

Estimation of the *participation state* is the key component of this study. From our observations of human interaction, we found that (a) people initiated conversation when their gaze met within a certain distance, and (b) people initiated conversation inside the partner’s field of view within a certain distance when the partner didn’t notice the other’s arrival. From these observations, we hypothetically developed the idea of a *participation zone*. It consists of three parts: *gaze zone*, *sight zone* and *front zone*. *Gaze zone* is a space established by one’s gaze; if two people were both in each other’s *gaze zone* (gazes meet), they would perceive an obligation to participate in a conversation. *Sight zone* is a space established by one’s sight; if one person wanted to initiate the participation with someone, he must enter the partner’s *sight zone* first. *Front zone* is a space established in front of a person; if people enter the partner’s *sight zone* and keep the partner in his own *front zone*, he would perceive an obligation to participate in a conversation. When both people enter each other’s *front zone*, they would both perceive an obligation to participate in a conversation.

As reported in section 3, the length of the *gaze zone* is changeable according to the environment (in the evaluation experiment, we set it as 2.5 m according to our observations). Figure 8a illustrates the *gaze zone*. *Gaze zone* is set to a 30-degree cone-shaped area in front of a person’s (or robot’s) head within a changeable distance.

We set up the precise parameters to define the *sight zone* from our observation results, and thus the zone was set to a 270-degree fan-shaped area in front of the body of a person (or robot) within a 1.5-m distance (Fig. 8b).

We also set up the precise parameters to define the *stable zone* from the personal distance [15], observations reported in section 2, and preliminary tests. The zone was thus set to a 120-degree fan-shaped area in front of a person (or robot) within a 2-m distance (Fig. 8c).

When these conditions are satisfied, the *participation state* transitions from *not-participating* to *participating*. However, this is not true for the opposite; the transition of the *participation state* from *participating* to *not-participating*

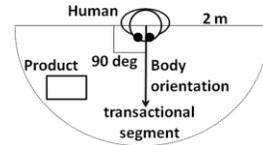


Figure 7 Transactional segment

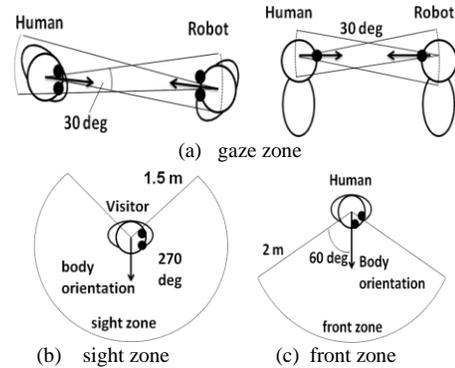


Figure 8 Participation zone

needs verbal interaction, and thus it is not controlled in this estimation module.

E. Spatial Formation Control

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

From our observations of human interaction, we found that: (a) The *host* kept facing the *visitor* and gazing at him 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 spatial formation.

- a) **Reactive adjustment of spatial formation**

When the customer is participating in the conversation, the robot not only needs to participate in it immediately but also needs to get closer to the customer and turn to her. We define this adjusting of position and orientation as *Reactive adjustment of spatial formation*. When the customer is participating in the conversation, the robot should start this adjusting at once despite its former plan. There are three rules for the *Reactive adjustment of spatial formation* (Fig. 9):

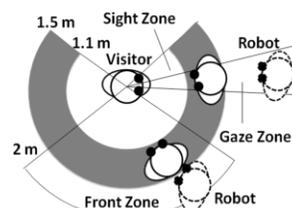


Figure 9 Reactive adjustment of spatial formation

- 1) The robot should be at a position that keeps itself in the *sight zone* of the *visitor*.
- 2) The robot should be at a position that keeps a distance of 1.1 m (used successfully in [23]) to 1.5 m (retrieved in our observation) with the *visitor*.

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

b) Proactive adjustment of spatial formation

When neither the *visitor* nor the robot is participating in the conversation, the robot should approach the customer first. Through our observations we found that *host* tended to approach the *visitor* while considering whether he had a *introduction plan*. Since at this time the robot has the freedom to choose the location, we define this approach as a *Proactive adjustment of spatial formation*. There are two rules for the *Proactive adjustment of spatial formation* (Fig. 10):

- 1) The robot needs to go into the *visitor's front zone* when coming from front. Otherwise, it only needs to go into the *visitor's sight zone* and keep a certain distance (1.1-1.5 m).
- 2) When the robot had a navigation plan, besides the first rule, it should also keep the target object (or direction) visible to both the *visitor* and itself in order to explain the target comfortably. In this paper, we set this as follow: the target should be in the field of view (set as 270 deg. by our observation) of both the *visitor* and the robot.

F. Utterance and Gesture Control

We prepared a simple utterance controller for controlling the robot's utterances. There are four contents: greeting, draw attention, guide, and explanation. A human developer pre-writes the sentences and the robot automatically uses them according to information from the state controller. The robot will choose to greet to the customer when both of their *participation-states* are *participating* and to draw attention when only the *visitor* is not *participating*. When both of them are participating in the conversation, if the *visitor* is paying attention to the target product, the robot will explain it, or it will guide the customer to the product first.

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 customer. It also receives input from the *utterance controller* to synchronize its pointing gesture with an utterance.

V. EVALUATION EXPERIMENT

We conducted an experiment to verify that our proposed model is useful for an information-presenting robot.

A. Scenarios

The experiment was conducted in a lab room, under the assumption that it was a small computer shop. There were three products in place (Fig. 1). A customer who visits this shop has an appointment with a salesperson robot to receive an explanation of one of the products. When he visits the shop, he freely waits for the salesperson robot to come. Then, when the

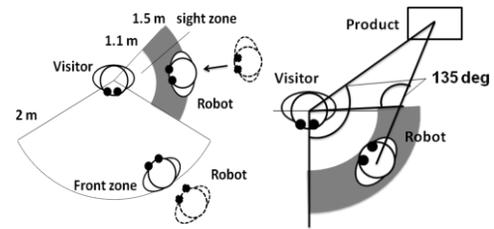


Figure 10 Proactive adjustment of spatial formation

robot arrives, they meet and initiate conversation. Finally, the robot explains the product. This setting is intended to place the focus of the evaluation on the interaction for initiating conversation.

We instructed participants to evaluate the interaction of the robot from the standpoint of a shop owner, who would choose one robot from candidates. They are asked to play the role of a customer in various ways so that they could fully judge the appropriateness of the behavior of each robot.

B. Conditions

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

a) Proposed Method (Proposed): The robot behaved based on the proposed model.

b) Always greet and guide (Guide): In this strategy, although the robot doesn't have a complicated model for the initiation of conversation, it behaves as politely as possible. It first tries to go to a location that is an appropriate distance for greeting the customer and then asks the customer to look at the product. It next goes to the location that is best for explaining the product, i.e., the location based on O-space, and explains the product.

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. Thus, when the robot finds a customer, it immediately stands at a location that is appropriate for explaining the product and starts to talk.

C. Procedure

A total of 15 native Japanese-speaking people (7 men, 8 women, average age: 27) participated in the experiment, for which they were paid.

The experiment was conducted in a 6 m x 10 m room. Due to the visibility of the motion capturing system, the experiment area was restricted to a 3 m x 4.5 m area. We used the robot and motion capturing system described in section 3.

In the experiment, participants first put on the markers of the motion capturing system, and an experimenter calibrated the motion capturing system.

Then, the scenario and instructions were provided to the participants. They were asked to evaluate three types of robots, as if they were the owner of the shop. Our intention was to let them evaluate various spatial formations for initiating

conversation, as each method would have its strengths and weaknesses.

For this purpose, they were asked to simulate the behavior of various types of customers, such as one who waits in front of the product and one who waits at the store entrance. Specifically, they were instructed to pretend to be five types of customers and to interact with a robot as each customer type, thus participating five times for each condition. In each condition, after interacting with the robot five times, pretending to be various types of customers for each interaction, they filled out a questionnaire designed for rating their impressions.

The experiment used a within-subject design and the order of conditions was counterbalanced.

D. Measurement

We asked participants to fill out a questionnaire for each condition (i.e., after five interactions). The measurement was a simple rating on a Likert scale of 1 to 7. The questionnaire had the following items:

- Appropriateness of the standing position when the robot greeted the customer
- Appropriateness of the standing position when the robot explained the target product
- Overall evaluation: How good or bad the participant thought the robot was

E. Hypothesis and Prediction

As discussed, we believe that it is important to be aware of the interlocutor’s *participation state*, which is estimated from the spatial configuration. Also, it is important for the robot to behave while considering the constraints for maintaining the *participation state*. While two of the methods try to make interaction natural and good, they lack the above points, which are implemented in the proposed method. Thus, our hypothesis states that if we were successful in implementing our ideas, the proposed method would produce the most appropriate interaction. Based on this consideration, we predicted the following: The proposed method will outperform the other two methods for the overall evaluation.

F. Results

Verification of prediction: For the overall evaluation score (Fig. 11), a repeated measures ANOVA was conducted. A significant main effect was found ($F(2,28)=9.125$, $p=.001$, partial $\eta^2 = .395$). Multiple-comparison with the Bonferroni method revealed that the score for the *proposed* condition is significantly higher than both *guide* ($p=.021$) and *best-location* ($p=.002$) conditions. No significant difference was found between *guide* and *best-location* conditions ($p=.5$). Therefore, our prediction was supported.

Analysis of details: To analyze the source of the difference, other scores were analyzed. For “appropriateness of standing position when it greeted” (Fig. 12), a repeated-measures analysis of variance revealed a significant main effect ($F(2,28)=4.697$, $p=.017$, partial $\eta^2=.251$), but multiple-

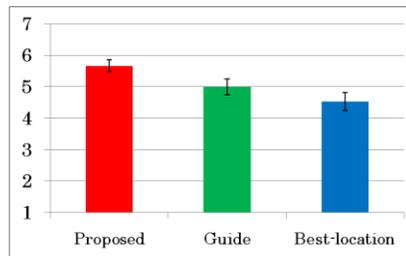


Figure 11 Overall evaluation

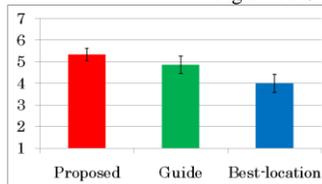


Figure 12 Appropriateness of standing position when it greeted

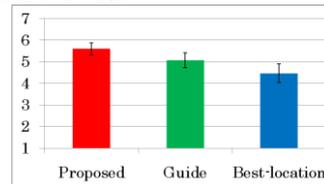


Figure 13 Appropriateness of standing position when it explained

comparison with the Bonferroni method only showed non-significant or almost-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 explained the target product” (Figure 13), 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 ($p=.003$), while other comparisons were not significant (*proposed* vs. *guide*: $p=.209$, and *guide* vs. *best-location*: $p=.111$).

In summary, the *proposed* system was evaluated as the best method overall among those compared. The effect of the proposed system in the overall evaluation could be partially 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. Participants would perhaps perceive the merit of the proposed method in other factors such as the appropriateness of timing or adequateness of behavior rather than the standing positions.

VI. DISCUSSION

A. When will this Capability be used?

We believe that the capability of a robot to naturally initiate conversation is one of the major capabilities to be implemented in future social robots. Although in many other research projects it is assumed that people and robots have already met and started interaction, in the real world this is not generally the case. Or, perhaps at an early deployment phase robots might not need to initiate interaction by themselves, as people would be interested in novel robots and walk over to the area in front of the robot. In such a case, robots don’t need to know the constraints of spatial configuration in the initiation of interaction.

In contrast, when robots actually start to work in the real world without attracting so much curiosity, people will often

not initiate interaction by themselves. In such a case, robots often fail to initiate interaction [21]. This will be more serious if the robot has a concrete role, e.g., shopkeeper, while working in our future daily lives. The shopkeeper scenario used in this study is one of the future situations where a robot is expected to take such a role. There are many other situations where a first meeting is involved, such as the role of 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.

B. Limitations

The proposed model was tested in a specific scenario, so its generalizability is limited. It is possible that context affects the preference of a robot's behavior. For example, in a busy business scenario, the condition of *always start interaction at the best location to explain* would work better than the proposed model. We believe that our shopkeeper scenario is rather neutral, so it could reflect interaction in many daily use scenarios, but this needs further verification.

The parameters in our model deal with Japanese people and our own robots. Therefore, when it is adapted, one would need to consider adapting parameters. For instance, the height of the people and robots would affect the distance parameter in the model.

VII. CONCLUSION

We studied natural interaction at the moment of initiating conversation. In a shopkeeper scenario where a salesperson meets a customer, we modeled natural human interaction. The model was implemented in a humanoid robot and tested with an evaluation experiment. We compared the proposed model with two baseline models. From the experiments' results, the proposed model was evaluated as the best.

VIII. ACKNOWLEDGEMENTS

This work was supported by KAKENHI 21118008.

IX. REFERENCES

- [1] H. H. Clark, *Using Language*, Cambridge University Press, 1996.
- [2] A. Kendon, 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] H. Kuzuoka, Y. Suzuki, J. Yamashita and K. Yamazaki, Reconfiguring Spatial Formation Arrangement by Robot Body Orientation, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp. 285-292, 2010.
- [4] M. P. Michalowski, S. Sabanovic and R. Simmons, A Spatial Model of Engagement for a Social Robot, *IEEE International Workshop on Advanced Motion Control*, pp. 762-767, 2006.
- [5] M. Shiomu, T. Kanda, H. Ishiguro and N. Hagita, 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] C. L. Sidner, C. D. Kidd, C. Lee and N. Lesh, Where to Look: A Study of Human-Robot Engagement, *International Conference on Intelligent User Interfaces (IUI 2004)*, pp. 78-84, 2004.
- [7] H. Kuzuoka, S. Oyama, K. Yamazaki, K. Suzuki and M. Mitsuishi, GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions, *ACM Conference on Computer-supported cooperative work (CSCW2000)*, pp. 155-162, 2000.
- [8] B. M. Scassellati, *Foundations for a Theory of Mind for a Humanoid Robot*, ed: Massachusetts Institute of Technology, 2001.
- [9] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman and M. Berlin, 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.
- [10] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki and H. Kuzuoka, Museum Guide Robot Based on Sociological Interaction Analysis, *ACM Conference on Human Factors in Computing Systems (CHI2007)*, pp. 1191-1194, 2007.
- [11] Mutlu, B., Forlizzi, J. and Hodgins, J., 2006, A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior, *IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids'06)*, pp. 518-523.
- [12] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro and N. Hagita, 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.
- [13] Y. I. Nakano and R. Ishii, Estimating User's Engagement from Eye-gaze Behaviors in Human-Agent Conversations, *International Conference on Intelligent User Interfaces*, pp. 139-148, 2010.
- [14] C. Rich, B. Ponsler, A. Holroyd and C. L. Sidner, Recognizing Engagement in Human-Robot Interaction, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010)*, pp. 375-382, 2010.
- [15] E. T. Hall, *The Hidden Dimension: Man's Use of Space in Public and Private*, The Bodley Head Ltd, 1966.
- [16] E. Goffman, *Behavior in public place: Notes on the Social Organization of Gatherings*, The Free Press, 1963.
- [17] A. Kendon, Features of the structural analysis of human communicational behavior, in *Aspects of Nonverbal Communication*, W. v. R. Engel ed., 1980.
- [18] Y. Katagiri, M. Bono and N. Suzuki, Conversational Inverse Information for Context-Based Retrieval of Personal Experiences, *Lecture Notes in Computer Science*, vol. 4012, pp. 365-376, 2006.
- [19] H. Hüttenrauch, K. S. Eklundh, A. Green and E. A. Topp, Investigating spatial relationships in human-robot interactions, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2006)*, pp. 5052-5059, 2006.
- [20] K. Dautenhahn, M. L. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, E. A. Sisbot, R. Alami and T. Siméon, 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.
- [21] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro and N. Hagita, How to Approach Humans?: Strategies for Social Robots to Initiate Interaction, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2009)*, pp. 109-116, 2009.
- [22] F. Yamaoka, T. Kanda, H. Ishiguro and N. Hagita, A Model of Proximity Control for Information-Presenting Robots, *IEEE Transactions on Robotics*, vol. 26, pp. 187-195, 2010.
- [23] E. Pacchierotti, H. I. Christensen and P. Jensfelt, Evaluation of Passing Distance for Social Robots, *IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN2006)*, pp. 315-320, 2006.
- [24] R. Gockley, J. Forlizzi and R. Simmons, Natural Person-Following Behavior for Social Robots, *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2007)*, pp. 17-24, 2007.
- [25] N. Bergström, T. Kanda, T. Miyashita, H. Ishiguro and N. Hagita, Modeling of Natural Human-Robot Encounters, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2008)*, pp. 2623-2629, 2008.
- [26] K. Yamazaki, M. Kawashima, Y. Kuno, N. Akiya, M. Burdelski, A. Yamazaki and H. Kuzuoka, 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.
- [27] T. Matthew Ciolek, A. Kendon, Environment and the Spatial Arrangement of Conversational Encounters, *Sociological Inquiry* Vol. 50, pp. 237-271, 1980.