Physical Constraints on Human Robot Interaction

Michita Imai
ATR MI&C Labs.
Seika, Soraku, Kyoto,
619-0288, Japan
michita@mic.atr.co.jp

Kazuo Hiraki
Presto, JST and Electrotechnical
Laboratory, MITI
1-1-4 Umezono, Tsukuba,
Ibaraki, 305-8568 Japan
khiraki@etl.go.jp

Tsutomu Miyasato
ATR MI&C Labs.
Seika, Soraku, Kyoto,
619-0288, Japan
miyasato@mic.atr.co.jp

Abstract

This paper conducts an experiment to investigate the effect of a physical constraint on a subject's viewpoint when using spoken language to navigate a robot. In addition, a robot navigation environment named Spondia-II has been developed for the experiment with an actual autonomous mobile robot. It is well known that the meaning of an utterance, such as a demonstrative pronoun, depends on the viewpoint of the speaker or the hearer. In a conversation between people, the primary factor in determining viewpoint is the physical constraints that are mediated by their body movements. This paper notes that these physical constraints also have an effect on viewpoint even when people instruct a robot. Furthermore, it is argued that, the utterance process also would greatly improve if the robot were able to comprehend the constraints.

1 Introduction

According to the development of personal robots [Yamasaki and Anzai, 1992; Burgard *et al.*, 1998] and robot entertainment [Fujita and Kageyama, 1997], we can suppose that someday, all people will have the chance to interact directly with robots. In the near future, a robot house-maid wearing an apron may be responsible for cleaning your house. Such dreams have appeared to come true with the appearance of Honda's robot P-2/P-3 [Hirai, 1998]. However, the actual appearance of the personal robot will require considerably more breakthroughs in the sensors, actuators, and robotics technologies to be used. To contribute to technologies for face-to-face interaction, this paper studies a natural conversation between a human and an autonomous mobile robot.

Dialogues between people proceed with the people referring to the situations surrounding them. A natural dialogue system must have the same capabilities as people. For example, the interpretation and the generation of utterances requires information on the locations of the speaker and the hearer, the object positions surrounding them, or their actions.

There are several studies [Imai *et al.*, 1994; Nagao and Takeuchi, 1994; Neal *et al.*, 1988] on dialogue systems dealing with external world information. For example, Linta [Imai *et al.*, 1994] is one dialogue system for a robot, able to interpret and generate utterances by extracting external information from the robot's sensors. In addition to robots, multi-modal dialogue systems [Nagao and Takeuchi, 1994; Itoh *et al.*, 1997] have also been developed for computer applications using a camera or a touch panel for extracting nonverbal information from a user. In addition, Barwise and Perry formalized relationships between certain situations and the semantics of utterances with logical expressions [Barwise and Perry, 1983]. There has also been a study on language evolution between two robots [Steels, 1997]. In the system, words the robots have developed are connected to external information through the robots' cameras.

In addition to external information, humans also use physical constraints on their conversations. These physical constraints appear to be mediated by the body movements of people. For example, a relationship may emerge when people face each other. Another relationship may emerge when people walk together. In these situation, the meaning of a word like the demonstrative pronouns "here" or "there" varies according to changes in the physical constraints. To date, however, dialogue systems for computer applications have not dealt with physical constraints because computers do not have an actual body. On the other hand, a robot can handle such constraints because it can interact with people based on its movable body. If a robot were able to comprehend the physical constraints, this would greatly improve the utterance process. However, there are some questions in trying to apply physical constraints to human-robot interactions: What are the actual physical constraints between two people? In particular, what kinds of constraints emerge in human-robot interactions? The answers to these questions still need clarification.

This paper conducts an experiment that investigates the usage of a Japanese demonstrative pronoun under a physical constraints in order to examine the effect of such constraints on human-robot conversations. For the experiment, we have developed an experimental environment named Spondia-II, where the experimental subject

could navigate the robot using the demonstrative pronoun.

The reason that Spondia-II focuses on employing the demonstrative pronoun is that the meaning of the pronoun varies according to the physical constraints. For example, the word "this" suggests an object in front of the speaker when the speaker and the hearer are facing each other. In contrast, if two people are walking together, "this" may suggest their shared destination. The meaning of "this" frequently differs depending on the physical constraints of the situation. Spondia-II uses the difference in the usage of the demonstrative pronoun to find out the effect of physical constraints.

The rest of this paper is organized as follows. Section 2 describes the features of demonstrative pronouns, and explains the effect of a viewpoint and physical constraints on the pronouns. Section 3 describes an experimental environment named Spondia-II and also explains the experimental conditions under Spondia-II. Section 4 describes an experimental method for Spondia-II and the outcome of trials involving the method. In section 5, we discuss the result of the experiment and the importance of physical constraints on conversations. Section 6 concludes the paper with a summary and future works.

2 Demonstrative Pronoun

2.1 Force of Viewpoint

The viewpoint that a speaker has is the deciding factor in the meaning of a demonstrative pronoun [McNeill, 1987]. A demonstrative pronoun suggests something directly, regardless of abstract matters or matters of substance. Such a function of the pronoun is called deixis in psycholinguistics. For example, "left" and "right" are deixis [Levelt, 1984]. To accurately understand the meanings of these type of pronouns, the hearer must notice the viewpoint of the speaker. Since a viewpoint is crucial to such meanings, we need to focus on the variety of viewpoints before examining how demonstrative pronouns are used.

The relationship between the meaning of a demonstrative pronoun and a speaker's viewpoint can be formalized as follows.

Rule 1 $p \leftarrow s \bigoplus d$

Here, p is the meaning of the pronoun, s is the situation around the speaker, $\bigoplus$[1] is the viewpoint of the speaker, and d is a demonstrative pronoun. **Rule 1** indicates that a robot must not only extract the information from a situation s but must also guess the user's viewpoint $\bigoplus$ in order to understand the demonstrative pronoun d. In addition, when generating the pronoun, the robot must be able to select a viewpoint that allows the user to easily understand the demonstrative pronoun generated.

Ullmer [Ullmer-Ehrich, 1982] has carried out an experiment in order to determine the movements of viewpoints when using deixis expressions. In his experiment, he made subjects describe objects in themselves' room.

[1] We use the same notation as in McNeill's book [McNeill, 1987].

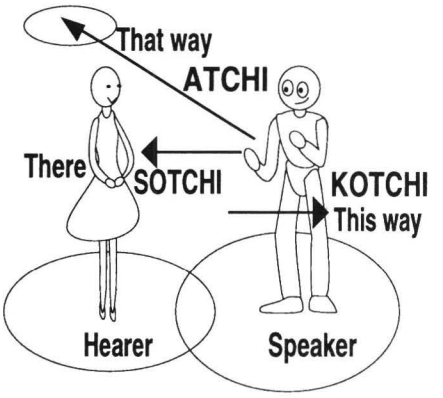

Figure 1: Default meaning of Japanese demonstrative pronouns.

As a result, almost all of the viewpoints stayed at the door to the room. Based on their viewpoint, they moved their focus of attention across the room as they described objects. Ullmer called each viewpoint a gaze tour (GT).

Klein [Klein, 1982] has also examined the transition of a viewpoint. He measured viewpoint by questioning pedestrians about a path to somewhere. As a result, the viewpoint moved along their walk. He called this viewpoint a walk tour (WT).

The studies of Ullmer and Klein indicate that the viewpoint of a speaker varies according to the condition of the speaker when he/she says a demonstrative pronoun. This fact can be formalized as follows.

Rule 2 $E \rightarrow \bigoplus$

This rule expresses that the speaker's viewpoint $\bigoplus$ is selected depending on condition E, as we mentioned in the explanations of GT and WT. For example, in Klein's experiment, E is the walk condition. According to **Rule 2**, the effect of the walk condition can be expressed as $Walk \rightarrow WT$, where $E = Walk$ and $\bigoplus = WT$.

The aim of this paper is to find the E in robot navigations. In particular, we would like to investigate the physical constraints of a situation as E.

2.2 Japanese Demonstrative Pronoun

There are three Japanese demonstrative pronouns: "ATCHI," "KOTCHI," and "SOTCHI." The meanings of these words are similar to "that way," "this way," and "there" in English, respectively. Figure 1 indicates the default meanings of these Japanese demonstrative pronouns. In the figure, the meanings depend on the viewpoint of the speaker. The utterance "KOTCHI" suggests a direction towards the speaker, the utterance "SOTCHI" suggests a direction near the hearer, and the utterance "ATCHI" suggests a direction away from both the speaker and the hearer.

Although speakers generally use the above default viewpoints for Japanese demonstrative pronouns, some speakers may use other viewpoints. Therefore, the hearer must notice changes in the speaker's point of view.

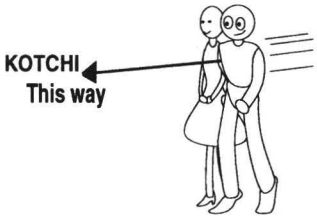

Figure 2: Effect of a physical constraint on the viewpoint.

2.3 Effect of Physical Constraints

Physical constraints are defined as constraints placed on the viewpoint when a speaker uses a demonstrative pronoun. These constraints emerge depending on the combination of the actions of the speaker and the hearer.

Figure 2 shows an example of a physical constraint when two persons are walking together. When people walk together, they frequently use "KOTCHI" (this way) to suggest a direction toward their destination. In this case, the meaning of "KOTCHI" differs from that in Figure 1. From the example, it is understood that the activity of walking together does affect the speaker's viewpoint. That is, a shared viewpoint emerges between the speaker and the hearer. As this example suggests, physical constraints have a significant effect on the viewpoint of the speaker and the hearer. Because of the importance of the viewpoint, we should not ignore such physical constraints when developing a utterance process system for demonstrative pronouns.

3 Spondia-II

We have constructed a robot navigation environment named Spondia-II that consists of an experimental field and an actual autonomous mobile robot. The aim of Spondia-II is to investigate the relationship between a viewpoint and a physical constraint. For the investigation, Spondia-II examines usage of demonstrative pronouns under two conditions: one is a condition with a physical constraint, and the other is a condition without any physical constraints. In the examination, the subjects of Spondia-II must navigate the robot from a start point to a goal point in the field using the demonstrative pronouns "ATCHI," "KOTCHI," or "SOTCHI."

3.1 Experimental field

The left side in Figure 3 shows the experimental field of Spondia-II. As can be seen in the figure, the start point is in one corner and the goal is in the opposite corner (from the start point). The physical viewpoint of the figure is the same as the subject's viewpoint. In short, the start point is in front of the subject. Thus, the start is nearer to the subject than the goal point. In Spondia-II, the situation s in **Rule 1** can be defined using the field of Figure 3. In short, s consists of the square field with the start point near the subject and the goal point far from the subject.

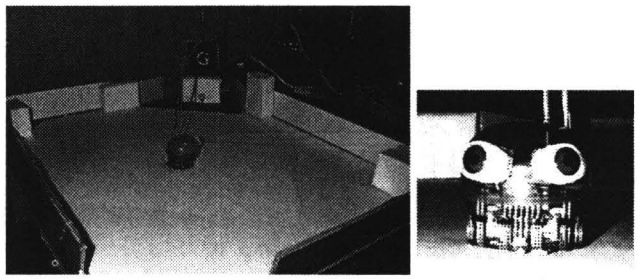

Figure 3: Spondia-II experimental environment and the mobile robot Khepera.

Spondia-II also uses a small mobile robot named Khepera. The right side in Figure 3 shows the body of Khepera. Since the viewpoint, as we mentioned previously, significantly affects the interpretation of a demonstrative pronoun, we fixed the plastic doll's eyes on its body to make the Khepera's viewpoint noticeable to the subject. Khepera also has eight distance sensors and illumination sensors around its body. There is a flashlight at the goal position, so that Khepera can adjust its orientation with respect to the goal, by using its illumination sensors. Using the distance and illumination sensors, Khepera is able to move around in the experimental field and change direction in response to the subject's commands of "ATCHI," "KOTCHI," or "SOTCHI."

3.2 Viewpoint in the field

The viewpoints which subjects have in the field of Spondia-II are grouped into three families (Figure 4). The first viewpoint, shown in Figure 4(A), is the same viewpoint as that of Figure 1. The second viewpoint, shown in Figure 4(B), is a shared viewpoint between the subject and Khepera, which emerges by having them walk together as in Figure 2. The third viewpoint, shown in Figure 4(C), is a bird's eye view of the field. In the bird's eye view, the subject selects a demonstrative pronoun according to the distance between him/herself and the target to be mentioned because the subject only looks at the targets in the field directly, regardless of the location of Khepera.

Although we proposed three kinds of viewpoints, Spondia-II actually combines these into two viewpoints: the shared viewpoint of Figure 4 (B) and a viewpoint merged from in Figure 4 (A) and (C). The two viewpoints will be used in the rest of this paper. The merged viewpoint of Figures 4(A) and (C) is called a daily viewpoint (DV). The viewpoint in Figure 4(B) is called a shared viewpoint (SV). The reason for the division is that there is the difference of a dependency on physical constraints between the two viewpoints. More specifically, the viewpoint SV is dependent on physical constraints while the viewpoint DV emerges under almost all situations regardless of physical constraints. As a result, the two viewpoints are summarized as $DV \in \{A, C\}$ and $SV \in \{B\}$.

In Spondia-II, the subject will take one of the two viewpoints. Therefore, the viewpoints of the subject and Khepera ($\oplus_s$ and $\oplus_k$ respectively) are formalized as $\oplus_i \in \{DV, SV\}$. Here, i denotes k or s. The reason this paper uses two viewpoints $\oplus_s$ and $\oplus_k$ is that the viewpoint of the subject is not necessarily the same as Khepera's.

It is essential for Khepera to adjust its viewpoint $\oplus_k$ to the subject's $\oplus_s$ in order for the given demonstrative pronoun to have an accurate meaning. To determine the conditions for the adjustment, Spondia-II examines what factors affect the subject's viewpoint $\oplus_s$. The examination will help Khepera gain knowledge for guessing a subject's viewpoint. Moreover, the knowledge will give Khepera the ability to control its own viewpoint $\oplus_k$ so the subject can also easily understand a demonstrative pronoun from Khepera.

Before the explanation of the factor in the change in the viewpoint, let us consider **Rule 1** in order to get into specifics of the effect of viewpoint on a demonstrative pronoun. The situation s in **Rule 1** can be fixed on the field of Spondia-II, as we mentioned in 3.1, because we use the same field throughout the experiment. In addition, the demonstrative pronoun d is given by the subject directly. Since Khepera certainly has information on s and d, Khepera need only decide on the subject's viewpoint $\oplus_s$ in order to use **Rule 1**. If Khepera can guess $\oplus_s$, the meaning p of d will be determined.

What is the main effect E in **Rule 2** on the subject's viewpoint $\oplus_s$? This paper proposes three possible effects for E. 1) The viewpoint $\oplus_s$ varies depending on the situation, like positions of objects. 2) The viewpoint $\oplus_s$ varies depending on Khepera's viewpoint $\oplus_k$. 3) The viewpoint $\oplus_s$ varies depending on a physical constraint.

Spondia-II investigates the effects of 2) and 3). It does not deal with effect 1) because there have already been many studies on this type of situation [McNeill, 1987].

Effects 2) and 3) are formalized as follows.

Rule 3 $\oplus_k \rightarrow \oplus_s$

Rule 4 $BI \rightarrow \oplus_s$

Rule 3 represents the effect of Khepera's viewpoint on the subject's viewpoint. In addition, **Rule 4** represents the effect of a physical constraint BI on the subject's viewpoint. Spondia-II makes a comparison between **Rule 3** and **Rule 4**; i.e., it decides which is most effective: Khepera's viewpoint $\oplus_k$ or the physical constraint BI.

3.3 Viewpoint model of Khepera

Khepera has two different viewpoints DV and SV that are used to examine the effect of its viewpoint on the subject's viewpoint (**Rule 3**). Khepera's viewpoints give it two models for interpreting demonstrative pronouns. However, Khepera's viewpoint $\oplus_k$ cannot be communicated to the subject directly. Instead of directly communicating its viewpoint, Khepera expresses the viewpoint by acting in response to the subject's utterance. There

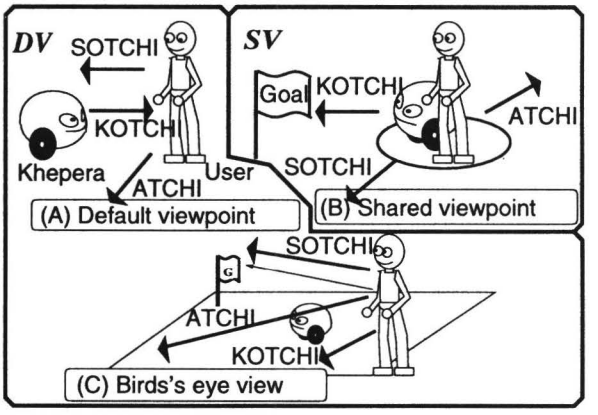

Figure 4: Viewpoints between a user and Khepera.

are two kinds of possible responses for Khepera. One is an action in which Khepera proceeds towards the goal when the pronoun "ATCHI" or "SOTCHI" is given. The action corresponds to the daily viewpoint DV. In the other action, which corresponds to the shared viewpoint (SV), Khepera proceeds towards the goal only when the utterance "KOTCHI" is given. The actions and viewpoints are formalized as follows.

Rule 5

$$\oplus_k = DV \quad : \quad \begin{cases} ATCHI, SOTCHI & \rightarrow \quad Goal \\ KOTCHI & \rightarrow \quad random \end{cases}$$

Rule 6

$$\oplus_k = SV \quad : \quad \begin{cases} ATCHI, SOTCHI & \rightarrow \quad random \\ KOTCHI & \rightarrow \quad Goal \end{cases}$$

3.4 Physical Interaction

To investigate the effect of a physical constraint (**Rule 4**), Spondia-II prepares two experimental conditions: one is an experiment with a physical constraint, and the other is an experiment without physical constraints.

In the actual condition for a physical constraint, the subject carries Khepera to the goal on his/her hand instead of controlling Khepera with vocal instructions. Therefore, the following two conditions BI for the physical constraint exist.

Condition 1 $BI = Voice$

Condition 2 $BI = Hand$

Spondia-II uses the "*hand*" condition to give the subject a shared viewpoint. In actuality, the shared viewpoint appears when they walk side by side. However, Khepera is too small to walk with the subject. Therefore, Spondia-II selects the "*hand*" condition to adjust for the size of Khepera.

4 Experiment

The purpose of Spondia-II is to examine the effect of a physical constraint on the subject's viewpoint when

Figure 5: Experimental scene: in the left picture, the subject navigates Khepera with vocal instructions. In the right picture, the subject carries Khepera to the goal by hand.

he/she navigates Khepera using a demonstrative pronoun. Since the speaker usually takes the daily viewpoint DV with a Japanese demonstrative pronoun, we want to note whether the shared viewpoint SV appears depending on the physical constraints or not. In short, Spondia-II examines the occurrence of the following rule.

Rule 7 $\bigoplus_s = DV \rightarrow \bigoplus_s = SV$

However, we cannot also observe the subject's viewpoint $\bigoplus_s$ directly, like Khepera's viewpoint. Thus, in order to determine the subject's viewpoint, Spondia-II observes the demonstrative pronoun the subject uses to suggest the goal. If the subject suggests the goal with the pronoun "ATCHI" or "SOTCHI," it is determined that he/she has the daily viewpoint DV. In contrast, if the subject uses the pronoun "KOTCHI," it is determined that he/she has the shared viewpoint SV. The decision is represented as follows.

Rule 8 $ATCHI, SOTCHI \rightarrow DV$

Rule 9 $KOTCHI \rightarrow SV$

These rules are the reverse of **Rules 5** and **6**.

4.1 Subjects of Spondia-II

There were 43 subjects, and their ages ranged from 18 to 39 years old.

4.2 Experimental story

In Spondia-II, we gave the subjects a task in which they must navigate Khepera from the start point to the goal. The story of the task has the following steps.

1. The subjects tell Khepera a demonstrative pronoun to move it towards the goal.
2. Khepera proceeds for a while, then stops and looks for the way to go by turning left or right quickly. The quick movement makes the subjects notice the unsettled feeling Khepera has when it becomes lost.
3. At this point, the subjects must take Khepera to the goal with one of the following two methods.
   (a) The subjects give Khepera a demonstrative pronoun again (left in Figure 5).
   (b) The subjects carry Khepera by holding it in their hand while giving the goal's location with a demonstrative pronoun (right in Figure 5).

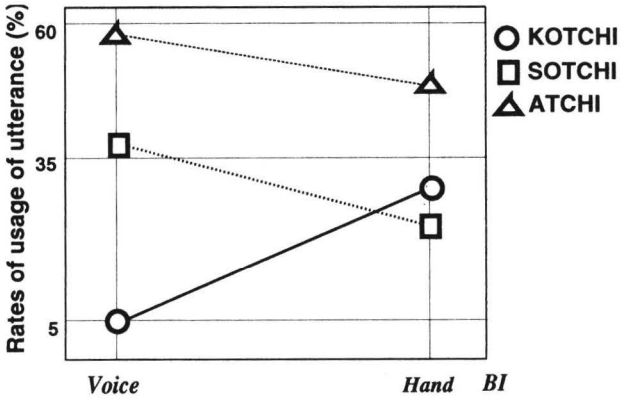

Figure 6: Rates of usage of each utterance under DV, $\chi^2(2) = 34.09, p < .01$.

Table 1: The value of $\chi^2(1)$ between two utterances under DV, $*p < .01$.

	KOTCHI	SOTCHI	ATCHI
KOTCHI		28.85*	24.39*
SOTCHI			1.07

According to the story, we recorded the demonstrative pronouns used by the subjects in 1, 3.a, and 3.b of the story. For step 1, we examined the subjects' default viewpoint (DV or SV). Step 3.a involved an examination under **Condition 1**, and step 3.b involved an examination under **Condition 2**. Throughout steps 3.a and 3.b, we investigated the effect of the physical constraint on the subjects' viewpoint.

Furthermore, we also studied the effect of Khepera's viewpoint on the subjects' viewpoint. For the investigation, we gave each subject one of two kinds of Khepera: one with a daily viewpoint DV and the other with a shared viewpoint SV. In other words, each subject interacted with Khepera that had only one type of response method, that is, **Rule 5** or **6**. In total, 27 subjects were given Khepera with the DV viewpoint, and 16 subjects were given Khepera with the SV viewpoint.

4.3 Outcome

From step 1 of the experimental story, we found the following usage rates for the demonstrative pronouns: 97.7% of the subjects used "ATCHI" to refer to the goal, and the remainder (2.3%) used "SOTCHI."

Figures 6 to 8 show the usage rates of each of the demonstrative pronouns in step 3 of the experimental story.

Figures 6 shows the data recorded using Khepera with DV, and Figures 7 shows the result of Khepera with SV. The horizontal axis gives the experimental conditions "*voice*" and "*hand*." The vertical axis represents the percentage of the use of each demonstrative pronoun. The results of the statistical analysis on the usage data

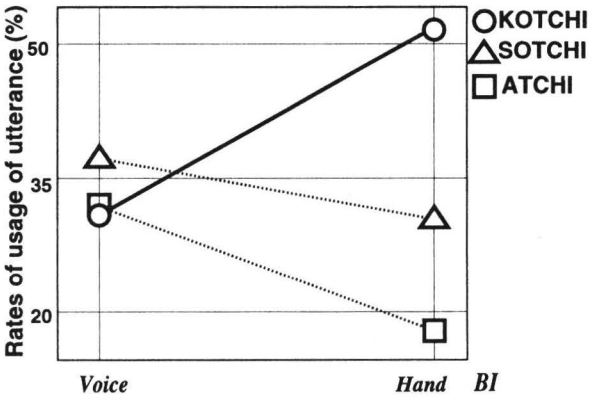

Figure 7: Rates of usage of each utterance under SV, $\chi^2(2) = 7.06, p < .05$.

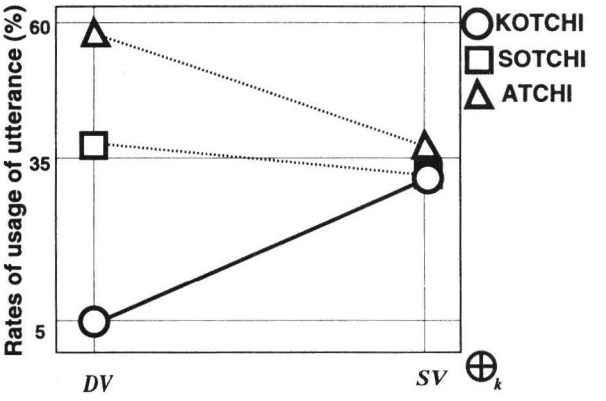

Figure 8: Rates of usage of each utterance (voice navigation), $\chi^2(2) = 34.61, p < .01$.

Table 2: The value of $\chi^2(1)$ between two utterances under SV, $*p < .05$.

	KOTCHI	SOTCHI	ATCHI
KOTCHI		5.18*	2.70
SOTCHI			1.32

Table 3: The value of $\chi^2(1)$ between two utterances (voice navigation), $*p < 0.01$.

	KOTCHI	SOTCHI	ATII
KOTCHI		21.17*	30.85*
SOTCHI			0.58

are given under each figure. We used a chi-square test for the analysis because the recorded data included the frequency of each demonstrative pronoun. Tables 1 and 2 show the results of a multiple comparison between the demonstrative pronouns based on their respective viewpoints.

Figure 8 shows the effect of Khepera's viewpoint $\oplus_k$. The horizontal axis shows each of Khepera's viewpoints. The vertical axis is the same as in Figures 6 and 7. Table 3 gives the results of a multiple comparison between the pronouns.

5 Discussion

In the outcome of step 1 of the experimental story, almost all of the subjects used the pronoun "ATCHI" to suggest the goal. This fact indicates that the subjects' default viewpoints were the same as DV.

Figure 6 and Table 1 show the significance of "KOTCHI" between the "*voice*" and the "*hand*" conditions. The usage rates of "KOTCHI" actually increased in the "*hand*" condition. According to **Rule 9**, such an increase indicates that the subjects' viewpoints $\oplus_s$ in the "*hand*" condition frequently became SV. On the other hand, DV appeared more frequently in the "*voice*" condition. As a result, this shows that the physical constraint is effective in the occurrence of **Rule 7**.

We also compared the effect of Khepera's viewpoint between DV and SV. Figure 8 and Table 3 show the comparison under the "*voice*" condition. The results show that the usage rates of each pronoun varied according to the change in Khepera's viewpoint. Moreover, it should be noticed that the usage rates of each pronoun were almost the same under SV.

From the same usage rates, we can understand that the subjects were confused about how to use the demonstrative pronouns. This is because they were unaccustomed to the action corresponding to **Rule 6**. Because of the confusing action, the subjects seem to use the demonstrative pronouns as nonsense symbols such as A, B, or C that have no meaning. In contrast to Khepera with SV, the subjects could navigate Khepera correctly in **Rule 5** because the action was based on the default viewpoint, DV.

Figure 7 and Table 2 compare the "*voice*" condition with the "*hand*" condition when Khepera has SV. The figure indicates the significance of "KOTCHI" versus "SOTCHI". The data for the "*voice*" condition is the same as the data for SV in Figure 8. From the discussion on Figure 8, the subjects also appeared to be confused about the usage of the pronouns. In contrast to the "*voice*" condition, there was an increase in the use of "KOTCHI" in the "*hand*" condition.

The fact should be noted that the subjects, in spite of the confusion, frequently used the utterance "KOTCHI" under the physical constraint "*hand*". In accordance with **Rule 9**, it is noticed that the subjects' viewpoints $\oplus_k$ were attracted to SV naturally. This fact indicates that the effect of a physical constraint is bigger than Khepera's viewpoint $\oplus_k$. Therefore, physical constraints are most useful for controlling a user's viewpoint towards a desired position.

The controlling method is particularly useful when the dialogue system interprets or generates a demonstrative

pronoun. For example, when a user walks along with the robot, the dialogue system can turn a viewpoint into a shared one to generate a pronoun, and it can interpret the pronoun based on the shared viewpoint. As a result, physical constraints should be used in human-robot conversations.

6 Conclusion

This paper investigated the effects of a physical constraint on a viewpoint for demonstrative pronouns for a user navigating a robot. For the experiment, we developed an experimental field named Spondia-II, which used a small mobile robot called Khepera.

In the experiment, we analyzed the subjects' usage of demonstrative pronouns under two conditions. One condition had a physical constraint: the subjects held Khepera in their hand and carried it to the goal. The other condition had no physical constraints: the subjects used only vocal instructions to navigate Khepera. These results showed that the subjects frequently had a shared viewpoint with Khepera under the physical constraints, and almost all of the subjects had a daily viewpoint under the vocal interaction conditions. The results indicate that physical constraints have an actual effect on human-robot interaction.

Since the experiments of Spondia-II have just begun, we have not completely examined all aspects of physical constraints. In future work, we plan to investigate the effect of a robot's size, a robot's action, and the type of subjects. Moreover, a very interesting question is whether a user will actually say a demonstrative pronoun to a robot. We plan to tackle these questions to improve the future of human-robot interactions.

References

[Barwise and Perry, 1983] J. Barwise and J. Perry. *Situations and Attitudes.* MIT Press, 1983.

[Burgard *et al.*, 1998] W. Burgard, A. B. Cremers, D. Fox, D. Hahnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. The interactive museum tour-guide robot. In *AAAI-98*, pages 11–18, 1998.

[Fujita and Kageyama, 1997] M. Fujita and K. Kageyama. An open architecture for robot entertainment. In *Proc. of Autonomous Agent'97*, pages 435–442, 1997.

[Hirai, 1998] K. Hirai. Research of Honda humanoid robot. In *Proc. of JSAI '98*, pages 1–4, 1998.

[Imai *et al.*, 1994] M. Imai, K. Hiraki, and Y. Anzai. Human-robot interface with attention. *Tranc. of IEICE (D-II)*, J77-D-II:1447–1456, 1994.

[Itoh *et al.*, 1997] T. Itoh, A. Denda, and S. Nakagawa. Evaluation of a sightseeing guidance dialogue system based on spoken Japanese with multi-modal interface and cooperative response. In *Interaction'97*, pages 135–142, 1997.

[Klein, 1982] W. Klein. *Speech, place, and action*, pages 161–182. Wiley, 1982.

[Levelt, 1984] W. J. M. Levelt. *Limits on perception.*, pages 323–358. VNU Science Press, 1984.

[McNeill, 1987] D. McNeill. *Psycholinguistics: A New Approach.* Harper & Row, 1987.

[Nagao and Takeuchi, 1994] K. Nagao and A. Takeuchi. Social interaction: Multimodal conversation with social agents. In *AAAI'94*, pages 22–28, 1994.

[Neal *et al.*, 1988] J. G. Neal, Z. Dobes, K. E. Bettinger, and J. S. Byoun. Multi-modal references in human-computer dialogue. In *AAAI-88*, pages 819–823, 1988.

[Steels, 1997] L. Steels. The origins of syntax in visually grounded robotic agents. In *IJCAI-97*, pages 1632–1641, 1997.

[Ullmer-Ehrich, 1982] V. Ullmer-Ehrich. *Speech, place, and action*, pages 219–250. Wiley, 1982.

[Yamasaki and Anzai, 1992] N. Yamasaki and Y. Anzai. Architecture for personal-robot. In *ROBOMEC '92 (A)*, pages 57–62, 1992.