Cross-modal effects between gestures and words in human robot interaction

Takamasa Iio[1,2], Masahiro Shiomi[1], Kazuhiko Shinozawa[1], Takaaki Akimoto[1],
Katsunori Shimohara[2] and Norihiro Hagita[1]

[1] *Advanced Telecommunications Research Institute International, Kyoto, JAPAN*
[2] *Doshisha University, Kyoto, JAPAN*
(Tel : 81-774-95-1405; Fax : 81-774-95-1408)
(iio@atr.jp)

Abstract: This paper reports a new finding of a phenomenon that person's gestures or words are implicitly modified by robot's gestures or words. Previous researches focused on an implicit effect of robot's gestures on person's gestures or an implicit effect of robot's words on person's words, but they did not focused on an implicit effect of robot's gestures on person's words or an implicit effect of robot's words on person's gestures. We supposed that there was such an effect as to arise between different modalities, and we defined it as a cross-modal effect. In order to verify hypotheses about the cross-modal effect, an experiment was conducted, in which a pair of a pointing gesture and a deictic word was focused on. This result showed that participants used a pointing gesture more often when a robot used a deictic word, and participants used a deictic word more often when the robot used a pointing gesture. Therefore, person's pointing gesture was implicitly modified by robot's deictic word, and also person's deictic word was implicitly modifie d by robot's pointing gesture. The cross-modal effect is expected to be applied to robot's dialog design to elicit comprehensible behavior from a person.

Keywords: Entrainment, Cross-modal effects, Multi-modal interaction, Human-robot interaction.

I. INTRODUCTION

Social robots that support people in daily life are expected to communicate with them in humanlike manners such as body movements or voices. That is because even people who do not use computers well could smoothly converse with the robots as if they converse with other people. For achieving natural communications, it is important to research human robot interaction [1]. One of purposes of the research is to understand how people communicate with the robots.

When a person communicates with a robot, an interesting phenomenon called entrainment often arises. This is the phenomenon that the person's gestures are synchronized with the robot's gestures or the person's words are synchronized with the robot's words as shown in Fig.1. For example, Ono et al. reported that person's gestures became similar with robot's gestures in a route direction conversation [2]. Iio et al. showed that when a person and a robot repeatedly referred to the same objects, the person tend to use the same words as the robot used [3]. By using the entrainment, a robot could implicitly elicit a certain behavior from a person. Such elicitation would make it possible for the robot to improve its recognition capability because the robot could elicit comprehensible gestures or words from the person. Therefore, we should understand how to modify person's behavior by robot's behavior.

The effect between each similar modality like the entrainment has been researched in human robot interaction. However, we have to consider not only the effect but also the effect between each different modality because all modalities are not independent respectively but some modalities, especially gestures and words are dependent on each other [4, 5]. If there is the effect, the robot could effectively elicit a certain gesture or a word from the person by using the effect with the entrainment.

We defined the effect between different modalities as cross-modal effect. Based on the entrainment between similar modalities and the mutual dependence of gestures and words, we supposed that there was the

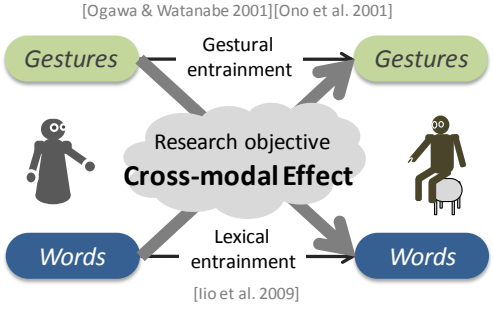

Fig.1. Entrainment and cross-modal effect.

following cross-modal effect: We assume that a gesture and a word are used together at the same time. If a robot often uses the gesture, a person would often use the similar gesture too; this phenomenon is gestural entrainment. Since the gesture tends to be used with the word, the person is also likely to use the word often. This means a cross-modal effect of the robot's gesture on the person's word. Based on the above logic, we can consider a cross-modal effect of the robot's word on the person's gestures.

In this paper, we supposed that there was the effect between different modalities. We defined it as cross-modal effect. The hypotheses about the cross-modal effect were made and verified through a laboratory experiment. Finally, a process of the cross-modal effect was discussed.

II. EXPERIMENT

1. Target gesture and word

This experiment aimed at a pair of a pointing gesture and a deictic word because they are coupled with each other. The deictic word cannot identify an object by itself; therefore they are likely to be used with a pointing gesture.

2. Experimental conversation

In order to introduce a pointing gesture and a deictic word, we employed an object reference conversation in the experiment. The conversation was as follows; the robot asked a participant to choose an object and the participant chose an object. Then, the robot confirmed the object. If the confirmation was correct, the participant indicated another object; otherwise the participant indicated the same object again.

We adopted books as the objects because books are found in many households; moreover, books involve the various referential expressions, such as title, color, category, author and location.

Table 1. The experimental conditions.

		Gestural factor	
		Pointing	No Pointing
Lexical factor	Deictic	*PD*	P_nD
	No deictic	PD_n	P_nD_n

Fig.2. The pictures of our experiment.

3. Hypotheses

We made the following hypotheses about a cross-modal effect during the object reference conversation.

H1: When the robot uses a deictic word, the rate of his or her pointing gesture is high.

H2: When the robot uses a pointing gesture, the rate of his or her deictic word is high.

4. Conditions

We controlled robot's pointing gesture and robot's deictic word used to confirm a book. The experiment had four conditions listed in Table 1. The detail of each condition was as follows.

PD: The robot turned its face on a book and pointed at the book, saying "Sore desuka?" (In English, That one?)

$\boldsymbol{P_nD}$: The robot only turned its face on a book but kept its arms stationary in the side of its body, saying "Sore desuka?"

$\boldsymbol{PD_n}$: The robot turned its face on a book and pointed at the book, saying the book title.

$\boldsymbol{P_nD_n}$: The robot only turned its face on a book but kept its arms stationary in the side of its body, saying the book title.

5. Measurement

We measured the pointing gesture rate and the deictic word rate to verify our hypotheses.

The pointing gesture rate: We counted how many times a participant did references with a pointing gesture and verified the rate is changed by the lexical factor.

The deictic words rate: We counted how many times participants did references with a deictic word and verified the rate is changed by the gestural factor.

6. Experimental environment

Fig. 2 depicts our experiment. A participant was seated in front of the robot. The robot was Robovie-R ver.2, which is a humanoid robot developed by the Intelligent Robotics and Communication Labs, ATR. There were five different books between the participant and the robot.

7. Procedure

A participant was first given a brief description of the purpose and the procedure of the experiment. We told the participant that we were developing a robot for

recognizing an object and would like their help in evaluating the design. Then, the participant was assigned randomly to the four conditions. A participant referred five books three times, that is to say, the participant did 15 references. Therefore, we totally obtained 90 references in each condition.

The robot was controlled remotely by an operator, that is to say, our experiment employed the Wizard of Oz method. That was because the difficulty of recognizing participant's pointing gesture and participant's voice automatically. Considering robot's recognition capability expected in the future, the operator rejected participant's references which did not pass the following rules.

- Reference by bibliographical information.
- Reference by attributions able to identify each book.
- Reference by pointing a finger at a book.

8. Participants

There were 24 participants in the experiment. They were native-Japanese-speaking university students from Kansai area. Since they were assigned randomly to the four conditions, each condition has six participants.

III RESULTS

In order to analyze main effects and interaction of the gestural factor and the lexical factor, we did the analysis of variance using chi-square distribution based on the arcsine transformation method. This analysis can test the difference between proportions of unpaired two factors.

1. Pointing gesture rate

The pointing gesture rate of each condition is shown in Fig.3. The analysis is as follows.

Main effect of the gestural factor: The pointing gesture rate of *PD* and P_nD was higher than that of PD_n and P_nD_n. The difference was significant ($x^2(1)$ = 11.017, $p < 0.01$). Therefore, when the robot used a pointing gesture, participants tended to use a pointing gesture. The result says there was gestural entrainment of robot's pointing gesture on participant's pointing gesture.

Main effect of the lexical factor: The pointing gesture rate of *PD* and P_nD was higher than that of the PD_n and P_nD_n. The difference was significant ($x_2(1)$ = 12.719, $p < 0.01$). Therefore, when the robot used a deictic word, participants tended to use a pointing gesture. The result says there was the cross-modal effect of robot's deictic word on participant's pointing gesture. That is to say, the result supports our hypothesis 1.

Interaction between the gestural factors and the lexical factors: There was no significant interaction between these factors.

2. Deictic word rate

The deictic word rate of each condition is shown in Fig.4. The analysis found out there was significant interaction between the gestural factor and the lexical factor ($x^2(1)$ = 7.209, $p < 0.01$). We describe the detail of the interaction.

Simple main effect of the gestural factor: The deictic word rate of PD_n was significantly higher than that of P_nD_n ($x^2(1)$ = 12.216, $p < 0.01$), but the deictic word rate of *PD* was similar with that of PD_n. Therefore, when the robot used a pointing gesture without a deictic word, participants tended to use a deictic terms. The result says there was partially the cross-modal effect of robot's deictic word on participant's pointing gesture. That is to say, the result partially supports our hypothesis 2.

Simple main effect of the lexical factor: The deictic word rate of P_nD was significantly higher than that of P_nD_n ($x^2(1)$ = 14.417, $p < 0.01$), but the deictic word rate of *PD* was similar with that of P_nD. Therefore, when the robot used a deictic word without a pointing gesture, participants tended to use a deictic terms. The result says there was partially lexical entrainment of robot's deictic word on participant's deictic word.

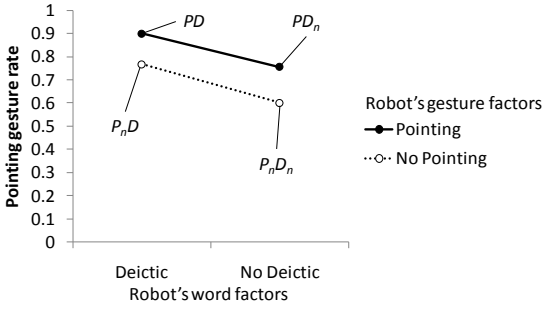

Fig.3. Pointing gesture rate of each condition.

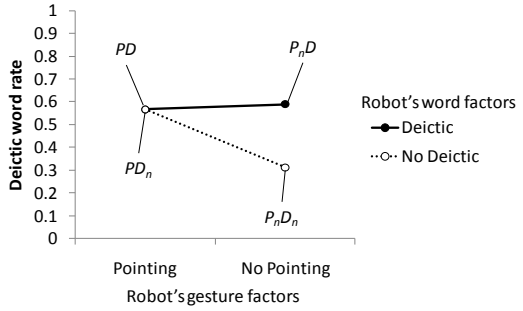

Fig.4. Deictic word rate of each condition.

IV. DISCUSSION

1. The process of the cross-modal effect

At first, we investigate whether a pointing gesture and a deictic word was used with each other. We grouped person's references into (A) a reference with both a pointing gesture and a deictic word, (B) a reference with a pointing gesture and without a deictic word and (C) a reference with a deictic word and without a pointing gesture. These rates were A = 0.625, B = 0.346 and C = 0.029, and they were significantly different ($x^2(1)$ = 149.621, $p < 0.01$). According to a multiple comparison, the rate of A was higher than that of B or C. Therefore, a pointing gesture and a deictic word are more likely to be used together.

Next, we consider the process of the cross-modal effect of robot's deictic word on participant's pointing gesture. According to the results in the section 4.2, there was partially lexical entrainment of robot's deictic word on participant's deictic word. Considering that a pointing gesture and a deictic word tended to be used together, we can say that, if only partially, the cross-modal entrainment arose from the mutual dependence of the pointing gesture and the deictic word and lexical entrainment.

Finally, we consider another process of the cross-modal effect of robot's pointing gesture on participant's deictic word. According to the results in the section 4.1, there was gestural entrainment of robot's pointing gesture on participant's pointing gesture. Like the above consideration, considering that a pointing gesture and a deictic word tended to be used together, we can say that the cross-modal entrainment arose from the mutual dependence of the pointing gesture and the deictic word and gestural entrainment.

2. Application of the cross-modal effect

We can consider to applying the cross-modal effect to a design of robot's behavior. For example, a robot could raise the possibility that a person use a pointing gesture by the gestural entrainment and the cross-modal effect. This helps the robot to recognize a referred-to object correctly because there are various objects in the real environment and sometimes the robot cannot identify each object only with words.

Although the experiment aimed at a pointing gesture and a deictic word, the results suggest that the cross-modal effect would arise from the mutual dependence of a gesture and a word and the entrainment. Therefore, if a gesture and a word are coupled, it might be possible for the cross-modal entrainment to arise in the other pair.

V. CONCLUSION

This paper defined an effect of one robot's modality on another person's modality as cross-modal effect. We supposed that if a gesture and a word are more likely to be used together, the cross-modal effect would arise through gestural entrainment or lexical entrainment, and we investigated that through a laboratory experiment.

The experiment focus on a pointing gesture and a deictic word during a conversation that a person and a robot refer to objects and investigated how the rate of person's pointing gestures and that of person's deictic words were changed by the combination of robot's pointing gestures and robot's deictic word. The experimental results were as follows; (1) a person used pointing gestures more often when the robot used deictic words; (2) a person used deictic words more often when the robot used pointing gestures.

In the cross-modal effect, a robot's modality elicits a similar person's modality by entrainment and the person's modality riggers another coupled modality. Therefore, when we design robot's behavior for maneuvering person's behavior, we should take account of not only the entrainment but also the cross-modal effect. We believe that this knowledge is useful for designing new conversational strategies for a robot.

REFERENCES

[1] Ishigro H, Ono T, Imai T, Maeda T, Kanda T and Nakatsu, R (2001), Robovie: An interactive humanoid robot, International Journal Industrial Robotics, 28(6):498-503.

[2] Ono T, Imai M and Ishiguro H (2001), A Model of Embodied Communications with Gestures between Humans and Robots, Proc. 23rd Annual Meeting of the Cognitive Science Society 732-737.

[3] Iio T, Shiomi M, Shinozawa K, Miyashita T, Akimoto T and Hagita N (2009), Lexical Entrainment in Human-Robot Interaction: - Can Robots Entrain Human Vocabulary? -", IEEE/RSJ International Conference on Intelligent Robots and Systems, 3727-3734.

[4] Kita, S (1998), Expressing turns at an invisible location in route direction: the interplay of speech and body movement, Sign-space Raum zeichen, 160 -172.

[5] Kita, S (2000), How representational gestures he lp speaking, Language and Gestures: Window into Thought and Action, 162-185.Casti JL (1995), Bell curves and monkey languages. Complexity 1:12-15