The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

# Mutual Entrainment: Implicit Elicitation of Human Gestures by Robot Speech

Takamasa Iio, Masahiro Shiomi, Kazuhiko Shinozawa, Takaaki Akimoto, Katsunori Shimohara and Norihiro Hagita

*Abstract*— **Social robots that provide services to humans in real environments have been developed in recent years. Such a robot should appropriately recognize its users' orders through human-like communications because of user-friendliness. However, their styles of communicating are too diverse to achieve this goal. If the robot could shape their styles, its recognition ability would be improved. An entrainment, which is a phenomenon where human's behavior is synchronized with robot's behavior, can be useful for this shaping. Previous studies have reported the entrainment occurring in the same modality, but they have given little attention to entrainment across different modalities (e.g., speech and gestures). We need to consider this cross-modal effect because human-robot interaction is inherently multi-modal. In this paper, we defined "mutual entrainment" as the entrainment across different modalities and investigated the effect of it through a laboratory experiment. We evaluate how the frequency of human pointing gestures varies with the amount of information in robot speech, and as a result, we found that the gesture frequency increased as the amount of information decreased. The results suggest that smoother human-robot communications can be achieved by shaping human behavior through mutual entrainment.**

## I. Introduction

SOCIAL robots that provide services to humans in a real environment have attracted attention in recent years. For example, guide robots have been developed for use at museums [1]-[4]. Such a robot is expected to appropriately recognize its users' orders in the manner of communications between humans in order to smoothly communicate with humans [5]. However, their styles of communicating are too diverse; they may use a variety of words or pointing gestures to indicate objects. Therefore, it is difficult that the robot recognizes such users' orders precisely. If the robot could shape their styles as it wants, its recognition ability would be improved. An entrainment is regarded as a method to shape their styles by robot's behavior.

The entrainment is a phenomenon that human's behavior is synchronized with robot's behavior implicitly. When a robot converses with a human, for example, robot's gestures entrain the human's gestures [8] (Fig. 1); robot's speech also entrains the human's speech [9] (Fig. 2). Entrainment is believed to foster a sense of unity between the human and the robot and to moderate their communications [10].

Previously, entrainment within the same modality has been considered: between human's gestures and robot's gestures and between human's speech and robot's speech. However, entrainment across different modalities has been little studied; it remains unknown whether robot's speech entrains human's gestures or whether robot's gestures entrain human's speech. Therefore, we need to clarify the cross-modal effects because human-robot interaction is inherently multi-modal and cannot be fully developed by entrainment only within a single modality.

In this paper, we define "mutual entrainment" as entrainment across different modalities, and we investigate its effects through a laboratory experiment. As a first step in the study, we focused on an entrainment between human's gestures and robot's speech because of wide range of application. Robot's gestures have limitations in actual applications since they depend on its hardware structure; on the other hand, robot speech is basically applicable to any robots by reprogramming its software. In the experiment, we evaluated the percentage of human's pointing gestures, altering the amount of information in robot's speech and tried to answer the following questions: Does the amount of

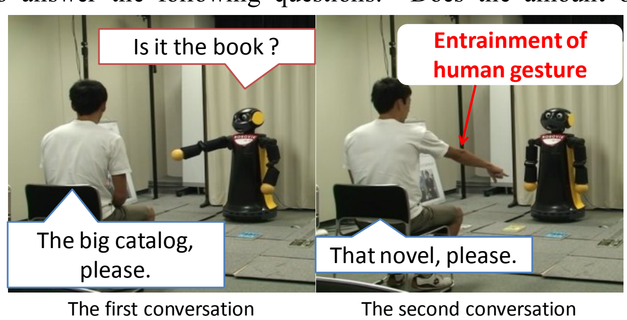


Fig. 1 Entrainment of human pointing gestures by robot pointing gestures

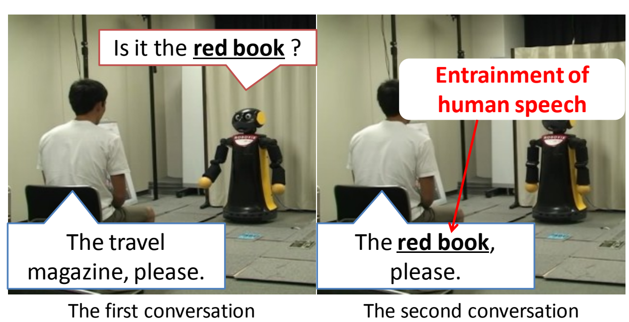


Fig. 2 Entrainment of human speech by robot speech

Manuscript received February 28, 2010. This research was supported by the Ministry of Internal Affairs and Communications of Japan.

T. Iio, M. Shiomi, K. Shinozawa, T. Akimoto, and N. Hagita are with the Advanced Telecommunications Research Institute International, Kyoto, JAPAN (corresponding author's phone: +81-774-95-2570; fax: +81-774-95-1401; email: iio@atr.jp)

K. shimohara is with Doshisha University, Kyoto, JAPAN.

information in robot's speech influence human's gestures? Is this influence maintained even if the robot does not make gestures?

## II. RELATED WORKS

### A. Entrainment of gestures

Entrainment in human interaction has been reported in the area of cognitive psychology [22]-[28]. These studies found that the speaker's gestures entrained the partner's gestures in conversations. For instance, Charny reported that the postures of a patient and a therapist were congruent in psychological therapy [28]. Recent studies have applied this knowledge to designing robot gestures, models of joint attention [19]-[21], and embodied communication in human-robot interaction [7][8][10].

Joint attention is the process by which humans who communicate with each other have shared attention to the same object through gazing and pointing [18]. Breazeal et al. developed a robot with a joint attention model. The robot could identify the attended object by recognizing the human gaze direction through vision processing [19]. In contrast, Imai et al. proposed an approach to robot-centered joint attention [6], i.e., the process by which a robot draws a human's attention to the same object that the robot is giving its attention to.

The studies on embodied communication have shown that human gestures are entrained by robot gestures. Ogawa et al. developed a robot that was capable of synchronizing the nod of its head with human speech. Through a conversation with a human, entrainment of the human nod motion was observed [7]. Ono et al. investigated human-robot communications involving giving/receiving route directions. The results showed that through entrainment human gestures increased as robot gestures increased [8].

A robot has the capability to perform multimodal interaction by using its speech and gestures. However, most studies have been aimed at entrainment of gestures between a human and a robot. Consequently, there has been no study on the entrainment of specific types of human speech by robot gestures.

### B. Entrainment of speech

Entrainment of speech has been studied just like that of gestures. According to the studies, two persons come to use the same terms for an object when they repeatedly talk about the object [11]-[13]. This is called lexical entrainment, and it has been studied in not only human-human interaction but also human-computer interaction [14]-[16] and human-robot interaction [9]. For example, Brennan suggested that a human readily adopted the terms of a computer partner through Wizard-of-Oz experiments using a database query task [14]. The results showed that users of a spoken dialog system adapted their lexical choices to the system vocabulary. Iio et al. conducted experiments in which a human referred to several objects in a conversation with a robot. The results revealed that the human tends to choose the same terms and the same category of terms as the robot used [9].

However, previous studies of lexical entrainment have mainly discussed entrainment of speech between a human and a robot. Therefore, the effect of robot speech on human gestures remains unknown. In this paper, we discuss mutual entrainment of human gestures by robot speech.

## III. METHODOLOGY

We conducted a laboratory experiment to understand the mutual entrainment between human's gestures and robot's speech, using the Wizard-of-Oz method.

### A. Experimental design

In our experiment, a subject instructed the robot to move objects around the subject. Referring to an object and confirming the object comprise the basic interaction of a human with a social robot. We employed 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.

The conversational flow is as follows; the robot asked a subject what books the subject would like to move. After the subject chose a book, the robot confirmed the book. If the confirmation was correct, the subject indicated another book; otherwise the subject indicated the same book again. Table I shows a typical example of the conversation between the subject and the robot.

In the experiment, a subject referred to five books in any order. We defined a period of referring to all books in this way as a session. The subject repeated the session three times. The interval between sessions was about two minutes. We obtained 15 references per subject, since the subject referred to five books three times.

### B. Design of confirmation behavior

To investigate mutual entrainment, we focused on the confirmation stage by the robot. The interaction of one person referring to an object and then another confirming the object is typical in human conversations. The study of lexical entrainment in human-robot interaction has also employed the confirmation stage by a robot [9]. We are interested in

TABLE I
EXAMPLE DIALOGUE

| *Speaker* | *Dialogue* |
|---|---|
| Robot | Please indicate a book. |
| Subject | The travel magazine, please. |
| Robot | Is it that red book? |
| Subject | Yes. |
| Robot | Ok, please indicate the next book. |
| Subject | Hmm, carry this novel, please. |
| | (This task continues until all books are indicated.) |

TABLE III
THE VIEW OF CONDITIONS

| *Robot's body motion* | *The amount of verbal information of robot's confirmation* | | |
|---|---|---|---|
| | Information amount 0 | Information amount 1 | Information amount 2 |
| The gaze and pointing gesture | gp0 | gp1 | gp2 |
| The gaze only | g0 | g1 | g2 |

how subjects' gestures vary according to robot's speech in the confirmation stage. We focused on robot's gestures and speech, designing the robot behavior in the confirmation stage.

*1) Gestures in confirmation*

Designing robot gestures in the confirmation stage, we considered the following two scenarios:

- Scenario 1: The robot can use both its gaze and pointing gestures.
- Scenario 2: The robot can use only its gaze.

In the first scenario, the robot turns its face and gaze toward an object and points at the object, speaking a prescribed message corresponding to each book. This confirmation behavior is usual for a robot with a human-like body.

In the second scenario, the robot turns its face and gaze toward the object only, speaking the message. Robot's arms remain stationary at that time. The scenario assumes a situation where the robot cannot move its own arms. Consider a situation where the robot carries bags in both hands, for example. In the situation, the robot has to use its gaze without pointing gestures so that the robot can quickly share attention to an object with the user. We think the behavior based on robot gaze is enough to communicate the meaning of confirmation.

In this paper, we investigate the effects of mutual entrainment in both scenario 1 and scenario 2.

*2) Speech in confirmation*

We manipulated the amount of robot's information in the confirmation stage. This paper defined the amount of information as the number of terms that uniquely identify a book. In particular, the robot used three types of referential expressions in the confirmation stage as follows.

- The robot used reference terms.
- The robot used a proper noun associated with each book.
- The robot used the proper noun and additional information about the book.

In the first type, a subject cannot recognize what books are confirmed by the robot without its gaze or pointing gestures. Thus, the amount of information is defined as 0.

In the second type, a subject can recognize the book confirmed by the robot without its gaze or pointing gestures, since the robot gives the proper noun that distinguished each book. Thus, the information amount is defined as 1. The reason for using not color or shape but the proper noun (title) is that humans tend to use titles to refer to books [32].

In the third type, the robot gives the proper noun and the additional information of the books, such as the book's color, size and category. Therefore, the subject can recognize the book confirmed by the robot as well as in the second type. Furthermore, the third speech is more prolix than the second. Thus, the amount of verbal information is defined as 2.

We list all types of robot speech in the confirmation stage in Table II. The reference terms of the robot speech with information amount 0 were changed due to positional relation. Here, Book 3 was the closest from both the subject and the robot; therefore, the reference term of Book 3 was "This".

## C. Conditions

We combined the three types of robot speech in the confirmation stage with the two types of robot gestures, and the six conditions are shown in Table III. Note that gp0, gp1 and gp2 mean the gaze and the pointing gestures with information amounts 0, 1, and 2, respectively; furthermore, g0, g1 and g2 mean gaze only with information amounts 0, 1, and 2, respectively.

TABLE II
LIST OF THE ROBOT CONFIRMATION SPEECH AT THE CONFIRMATION

| | *The information amount of the robot speech in the confirmation stage* | | |
|---|---|---|---|
| | 0 | 1 | 2 |
| *Book 1* | Is it that? | Is it "The book to learn English using Gundam - that famous speech is expressed like this."? | Is it the black book titled "The book to learn English using Gundam - that famous speech is expressed like this." and written by the Gundam English Study Group. |
| *Book 2* | Is it that? | Is it "Journal of experimental psychology, Leaning memory and cognition"? | Is it the large blue magazine titled "Journal of experimental psychology, Leaning memory and cognition"? |
| *Book 3* | Is it this? | Is it "Murphy's law - the intelligence in contemporary American society."? | Is it the small yellow book titled "Murphy's law - the intelligence in contemporary American society."? |
| *Book 4* | Is it that? | Is it "English as a tool - English conversations using well-known words."? | Is it the white book titled "English as a tool - English conversations using well-known words." of the additional volume in the TAKARAJIMA series? |
| *Book 5* | Is it that? | Is it "Walking in Kyoto."? | Is it the red travel magazine titled "Walking in Kyoto."? |

### D. Hypotheses

In scenario 1, entrainment of gestures tends to occur between a subject and the robot because the robot uses pointing gestures in the confirmation stage. Therefore, the subject would often use pointing gestures in the first place. We consider how to increase the frequency of the subject pointing gestures in such a situation. Humans often speak terms with information amount 0 like a reference term when pointing at an object. In other words, a subject would more use pointing gestures if the robot entrains the subject speech into terms with information amount 0. The knowledge of lexical entrainment implies that a human tends to adopt the same type of terms as a robot uses. Thus, if the robot uses terms with information amount 0 in the confirmation stage, the subject may use these terms also. As a result, the subject pointing gestures would increase.

Furthermore, in scenario 2, the robot uses only its gaze in the confirmation stage. Entrainment of gestures does not occur between a subject and the robot because the robot does not use its pointing gestures. In such a situation, can the robot still entrain the subject pointing gestures? Many studies have reported that human gestures increased as robot gestures increased, but few studies have discussed the entrainment of human gestures under restricted robot gestures.

From the above considerations, we developed the following two hypotheses:

- Hypothesis 1: A subject uses more pointing gestures as few amount of information in the robot speech.
- Hypothesis 2: The hypothesis 1 is correct even when the robot does not use pointing gestures.

### E. Experimental environment

Fig. 3 shows the experimental setup. The experiment was conducted in a rectangular room 7.5 m by 10 m. We used an area of 3.5 m by 3.5 m in the center of the room, due to the restricted area covered by the video camera. A subject was seated in front of the robot. Five different books were positioned between the subject and the robot so that the subject could identify these books by sight.

#### 1) Video camera and microphone

In this Wizard of Oz experiment, an operator controls the robot remotely. We installed a video camera in the experimental room so that the operator could observe the body motion of the subject and robot, and we attached a microphone to subjects' body so that the operator could listen to subjects' speech. The video image and voice were all recorded for analysis after the experiment.

#### 2) Operator role

The experiment employed the Wizard-of-Oz method because the robot has difficulties in recognizing an object referred to by an untrained subject. The difficulties arose from the following two issues.

- When subjects refer to objects, they use not only speech but also gesture.

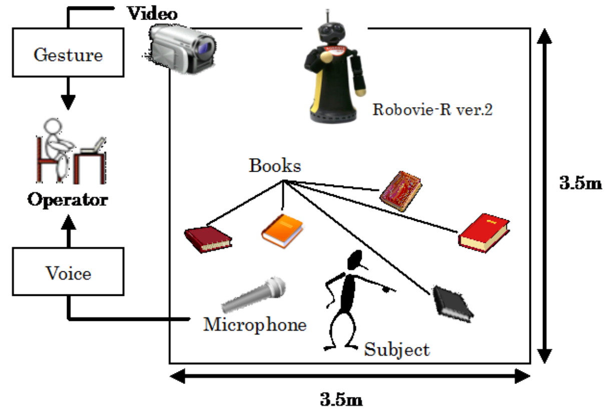


Fig. 3 Experimental environment.

- And also, they use many types of referential expressions: the object's name, color, shape, location and so on.

The first difficulty is that it is hard to precisely recognize the direction of their pointing gestures. For example, a subject often says, "Give me that," pointing at an object. In this case a robot has to recognize the pointing gesture, since subjects' speech does not include information about the object. The second difficulty is that it is nearly impossible to describe all of the associations between the object and the various referential expressions. To avoid these difficulties, an operator played the role of the recognition part of the robot.

The operator recognized subjects' reference in place of the robot and then initiated a behavior program of the robot on the basis of the voice and gesture of the subject.

The operator only recognized references meeting the following rules and rejected the other ways of reference because we assumed robot's cognitive ability expected in the future..

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

In the first rule, the bibliographical information includes a title, part of the title, author, and category. The information is generally open to the public; therefore, we assumed the robot could obtain the information.

In the second rule, the attribution is based on the supposition that a robot could pick up a characteristic color or size by an image recognition technique in the future. Therefore, the operator recognized the references by color and size terms predefined by the experimenter.

The third rule was also set on the basis of the feasibility of the pointing gesture recognition.

#### 3) Humanoid robot

Robovie-R ver.2 is a humanoid robot developed by the Intelligent Robotics and Communication Labs, ATR. It has a human-like upper body designed for communicating with humans. Fig. 4 shows its appearance. It has a head, two arms, a body and a wheeled-type mobile base. On its head, it has two CCD cameras for eyes and a speaker for a mouth. The

Fig. 4 Robovie-R2

speaker can output recorded sound files installed on the internal-control PC located in the body. Robot's degrees of freedom (DOFs) are as follows: two DOFs for the wheels, three DOFs for its neck, and four DOFs for each arm. Its body has sufficient expressive ability to perform human-like gestures. In addition, it has two wheels to move (forward-reverse travel and rotation). Its height is 1100 mm, its width is 560 mm, its depth is 500 mm and its weight is about 57 kg.

### *F. Experimental Procedure*

A total of 36 subjects participated in the experiment. All subjects were native-Japanese-speaking university students from Kansai area. The subjects were first given a brief description of the purpose and the procedure of the experiment. After this introduction, they were asked to review and sign a consent form. We obtained permission to record video and sensor data from the responsible authorities of the mall. The experimental protocol was reviewed and approved by our institutional review board.

The subjects moved to the experimental laboratory to learn the details of the task. We told them that we were developing a robot for recognizing an object and would like their help in evaluating the design, and then the subjects were assigned randomly to the six conditions. Since six subjects in each condition referred to five books, we obtained 90 references.. Fig. 5 shows images from the experiment.

### *G. Measurement*

To verify our hypotheses on the mutual entrainment in human-robot interaction, we manipulated the amount of information in robot's speech in the confirmation stage. We aimed at the following two kinds of measurements.

- Gesture: We captured subject gestures using a video camera. From the video, we measured the number of references with the pointing gestures.
- Speech: We recorded subjects' speech using a small microphone. From the audio data, we measured the number of references where the information amount was (i) 0, (ii) 1, and (iii) 2.

## IV. Results

We designed the two scenarios based on robot's gestures and conducted a between-subjects experiment in each scenario.We manipulated the amount of verbal information in robot's speech at the confirmation stage.

Fig. 5 Images of the experiment.

### *A. Scenario 1*

In scenario 1, the robot used its gaze and pointing gestures.

#### *1) Subject's pointing gestures*

Fig. 6 shows the numbers of pointing gestures made by subjects in each condition. The numbers were 81 in gp0 but decreased to 68 in gp1 and, moreover, to 63 in gp2.

We compared the percentages of the numbers among the conditions using the chi-square test because the conditions were unpaired and the numbers were nonparametric. As a result, the percentages were significantly different among the conditions ($x^2(2) = 11.374$, $p < 0.01$, Cramer's V = 0.205). We conducted a residual analysis to determine in what conditions the number has a significant difference from the expected frequency. The results showed that the percentage in gp0 was significantly high (adjusted residual = 3.248, $p < 0.01$) and the percentage in gp2 was significantly low (adjusted residual = -2.410, $p < 0.05$).

The results showed that subjects' pointing gestures decreased as the amount of verbal information in robot's speech increased.

#### *2) Amount of information in subjects' speech*

The number of the amount of information in subjects' speech is shown in Fig. 7. First, we look at the numbers of the amount of information 0 in subjects' speech. The numbers were 36 in gp0 but then decreased to 17 in gp1 and, moreover, to 9 in gp2. Next, we look at the numbers of the amount of information 2 in subjects' speech. The numbers were 1 in gp0 but then increased to 8 in gp1 and, moreover, to 16 in gp2.

We compared all percentages of these numbers among the conditions using the chi-square test. The percentages were significantly different among the conditions ($x^2(4) = 33.707$, $p < 0.01$, Cramer's V = 0.250). As a result of the residual analysis, a significant difference was found for the amount of information 0 (adjusted residual = 4.706, $p < 0.01$) and the amount of information 2 (adjusted residual = -3.266, $p < 0.01$) in gp0 and for the amount of information 0 (adjusted residual = -3.581, $p < 0.01$) and the amount of information 2 (adjusted residual = 3.415, $p < 0.01$) in gp2. The results showed that the amount of verbal information in subjects' speech increased with that in robot's speech.

In short, these results support our hypothesis 1: Human's pointing gestures increase when the robot gives a smaller amount of information in its speech. The reason for such mutual entrainment might derive from lexical entrainment: The subject tended to use more reference terms similar to those of the robot with the amount of information 0 in robot's speech; the subject tended to use more prolix expressions similar to those of the robot with the amount of information 2 in robot's speech.

### *B. Scenario 2*

In scenario 2, the robot used only its gaze.

#### *1) Subjects' pointing gestures*

Fig. 8 shows the number of pointing gestures made by subjects. The numbers were 69 in g0, 54 in g1, and 59 in g2.

We compared the percentages of the numbers among the conditions using the chi-square test. As a result, the difference of the percentages was marginally significant among the conditions ($x^2(2) = 5.767$, $0.05 < p < 0.1$, Cramer's V = 0.147). The results of residual analysis showed that the number in g0 was significantly high (adjusted residual = 2.127, $p < 0.05$) and the number in g2 was significantly low (adjusted residual = -2.042, $p < 0.05$). The number was not significantly different in g1.

#### *2) Amount of information in subjects' speech*

The number of the amount of information in subjects' speech is shown in Fig. 9. First, we look at the numbers of the amount of information 0 in subjects' speech. The numbers were 40 in g0, 8 in g1, and 19 in g2. Next, we look at the number of the amount of information 1 in subjects' speech. The numbers were 47 in g0, 78 in g1, and 64 in g2.

We compared the percentages of the numbers using the chi-square test. The percentages were significantly different among the conditions ($x^2(4) = 33.180$, $p < 0.01$, Cramer's V = 0.248). The results of the residual analysis were as follows: The percentage of subjects' references of the amount of information 0 was high in g0 (adjusted residual = 5.280, $p < 0.01$) and it was low in g1 (adjusted residual = -4.507, $p < 0.01$), and in contrast, the percentage of subjects' references of the amount of information 0 was low in g0 (adjusted residual = -4.284, $p < 0.01$) and it was high in g1(adjusted residual = 4.226, $p < 0.01$). However, the percentage of subjects' references of the amount of information 2 was not significantly different in g2.

These results support our hypothesis 2: Even when the robot does not use its pointing gestures, human pointing gestures increase as few amount of information in its speech. The reason for the increase in human pointing gestures for information amount 0 of the robot speech might derive from lexical entrainment, as in the case of scenario 1. However, there were more human pointing gestures for information amount 2 than for information amount 1. The results are different from those of scenario 1: Human pointing gestures occur less when the robot increased the amount of information. The reason for the increase in human pointing gestures for information amount 2 in the robot speech is that the subject used more reference terms than prolix expressions.

## V. Discussion

### *A. Physical modalities and mutual entrainment*

The number of subjects' pointing gestures was the highest

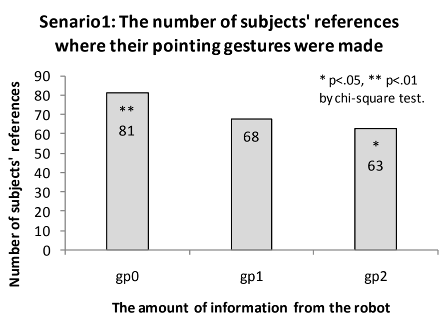


Fig. 6 The number of subject's references where their pointing gestures made in the scenario 1 where the robot used its gaze and pointing gestures.

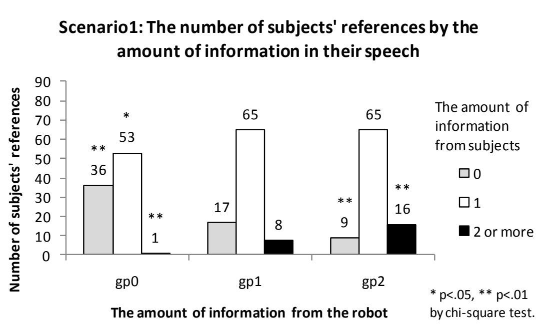


Fig. 7 The number of subject's references by the amount of information 0, 1 and 2 in their speech in the scenario 1 where the robot used its gaze and pointing gestures.

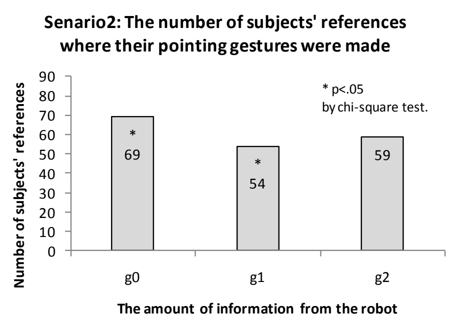


Fig. 8 The number of subject's references where their pointing gestures made in the scenario 2 where the robot used its gaze only.

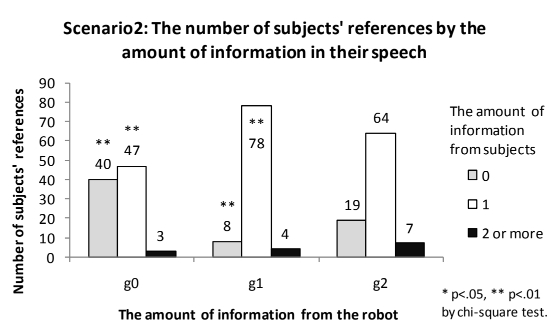


Fig. 9 The number of subject's references by the amount of information 0, 1, and 2 in their speech the scenario where the robot used its gaze only.

for the amount of information 0 of robot's speech in scenario 1 and scenario 2. This tendency of mutual entrainment was the same in both scenarios even if the physical modalities were different. This is important because the robot could entrain human's pointing gestures by decreasing information, even if the robot cannot make the pointing gestures.

However, the existence or nonexistence of the physical modalities of the robot partially brought different results. In scenario 1, where the robot used its gaze and pointing gestures during confirmation, the number of subjects' pointing gestures decreased as the amount of information in robot's speech increased. However, in scenario 2, where the robot used its gaze only during confirmation, the number of subjects' pointing gestures did not linearly decrease as the amount of information in robot's speech increased. In particular, the number of subjects' pointing gestures was higher in the amount of information 2 (g2) in robot's speech than the amount of information 1 (g1).

The different results possibly arise from a gap between robot's gestures and the amount of information in robot's speech. The gap indicates that robot's gestures were restricted to its gaze, whereas the amount of information in robot's speech was rich. The gap might bring discomfort to subjects, and as a result, the tendency of subjects' pointing gestures was different in g2. However, this paper could not investigate that possibility because the data were insufficient to support that supposition; therefore, we still need to conduct experiments by designing other scenarios, e.g., the robot making pointing gestures without gaze during confirmation and the robot remaining stationary during confirmation.

### *B. Effect of length of robot speech*

Our experimental results show that human pointing gestures can be implicitly induced by manipulating the amount of information in robot's speech. However, there remains a question: Does the amount of information really affect human pointing gestures? That is to say, we have to consider the possibility that the length of the robot speech has an effect on human's pointing gestures. In this paper, we cannot precisely discuss the effect of the length of robot's speech on mutual entrainment. We need to conduct an experiment in a situation where the robot talks about any topic unrelated to confirmation of the topic under discussion.

However, we believe that not so much the length of robot's speech as the amount of information affected subjects' pointing gestures. This is because subjects' pointing gestures were related to lexical entrainment, and the lexical entrainment, in turn, was related to the amount of information. According to the experimental results, when robot's speech was the amount of information 0, subjects' speech was also the amount of information 0; in other words, when the robot spoke reference terms, the subject also spoke similar reference terms and made pointing gestures. In contrast, when robot's speech was the amount of information 2, subjects' speech was also the amount of information 2; in other words, when the robot spoke prolix expressions, the subject also spoke similar prolix expressions without the pointing gestures. The above suggests that the number of subjects' pointing gestures varied by incorporating subjects' speech into robot's speech. Therefore, we believe that the amount of information in robot's speech more strongly affects subjects' pointing gestures than the length of robot's speech.

### *C. Application of mutual entrainment*

Mutual entrainment could be useful for designing robot behaviors to support the recognition capability of the robot. We can consider applying this concept to choosing the proper sensors in a sensor network.

Sensor networks have been studied in recent years [2]-[4]. A robot connected to such networks gains an advanced recognition capability by integrating various sources of information that could not be achieved by any single sensor [29]-[31]. In sensor networks, choosing the proper sensors for a purpose is a fundamental problem. This problem derives from the difficulty of searching for the proper sensors from a huge number of sensors. Mutual entrainment could resolve the problem and improve the recognition capability of the networked robot.

We assume the conversation of a human with a robot in the situation where the sensors of speech recognition are unable to function normally due to some noise. This situation could occur for a robot moving in a real environment like a museum [1]-[4].

When the human refers to a nearby object, the robot has to use other sensors besides those related to speech recognition and recognize the object referred to. If the human points at the object at that time, the robot becomes able to recognize the object easily. However, the robot asking the human to point at the object at first might lead the human to feel stress because he or she would be always very conscious of his or her gestures.

The robot could implicitly entrain the human pointing gestures by decreasing the amount of information in its own speech, according to the results of our experiment. Assumptions about the object referred to could be made more successfully. Furthermore, such mutual entrainment is possible even if the robot cannot move its arms. It is significant that human pointing gestures are entrained by robot's speech without robot's pointing gestures.

### *D. Limitations*

We discussed mutual entrainment in a limited situation where subjects conversed face to face with the robot in a laboratory room and referred to a book. However, we believe that mutual entrainment can be effective for referring to various objects in a real environment. According to a previous work, lexical entrainment can occur in a real environment having various objects with more unfamiliar names than books with titles [9]. Mutual entrainment can also occur and provide a useful tool in a real environment, since it relates closely to lexical entrainment.

## VI. Conclusion

This paper focused on mutual entrainment between human gestures and robot speech. The previous studies of entrainment in human-robot interaction have discussed entrainment between the same modalities of a human and a robot. However, there has been little study on mutual entrainment across different modalities. Therefore, we considered whether robot speech entrained human gestures.

We conducted an experiment where a subject and a robot referred to objects alternately. We designed two scenarios involving robot gestures and manipulated the amount of information in robot's speech when it confirmed the object referred to in each scenario. We measured the number of subjects' pointing gestures to the reference object. As a result, the following findings were obtained:

- In the scenario 1 where the robot used its gaze and pointing gestures, subjects' pointing gestures were most at the amount of information in robot's speech; furthermore, the subject pointing gestures decreased as the amount of information in robot's speech increased.
- In the scenario 2 where the robot used its gaze only, subjects' pointing gestures were also most at the amount of information 0 in robot speech; however, the difference of the percentages was marginally significant among the conditions.

In short, human pointing gestures tended to slightly decrease as the amount of information in robot's speech increases. We believe that this knowledge is useful for designing new conversational strategies for a robot. This might include, for instance, the robot implicitly entraining human pointing gestures through its speech to support its recognition capability.

## References


[1] M. Shiomi, T. Kanda. H. Ishiguro, and N. Hagita,"Interactive Humanoid Robots for a Science Museum", IEEE Intelligent systems (Special Issues on Interactive with Autonomy), vol. 22, no. 2, pp. 25-32, Mar/Apr, 2007.

[2] T. Kanda, M. Shiomi, L. Perrin, T. Nomura, H. Ishiguro, and N. Hagita, "Analysis of people trajectories with ubiquitous sensors in a science museum", Proc. 2007 IEEE Int. Conf. Robotics and Automation (ICRA'07), no. FrE8.4, pp. 4846-4853, 2007.

[3] Y. Koide, T. Kanda, Y. Sumi, K. Kogure, and H. Ishiguro, An Approach to Integrating an Interactive Guide Robot with Ubiquitous Sensors, Int. Conf. on Intelligent Robots and Systems (IROS), pp. 2500-2505, 2004.

[4] W. Burgard, A. B. Cremers, D. Fox, D. H¨ahnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "The Interactive Museum Tour-Guide Robot", Proc. National Conference on Artificial Intelligence (AAAI), 1998.

[5] H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, and R. Nakatsu: "Robovie: an interactive humanoid robot," Int. J. Industrial Robotics, vol. 28, no. 6, pp. 498-503, 2001.

[6] M. Imai, T. Ono, and H. Ishiguro. "Physical relation and expression: Joint attention for human-robot interaction," IEEE Transaction on Industrial Electronics, Vol. 50, No. 4, pp. 636-643, 2003.

[7] H. Ogawa and T. Watanave, "InterRobot: speech-driven embodiment interaction robot," Advanced Robotics, Vol. 15, No. 3, pp. 371-377, 2001.

[8] T. Ono, M. Imai and H. Ishiguro: A Model of Embodied Communications with Gestures between Humans and Robots, Proc. of Twenty-third Annual Meeting of the Cognitive Science Society, pp. 732-737, 2001.

[9] T. Iio, M. Shiomi, K. Shinozawa, T. Miyashita, T. Akimoto, N. Hagita, "Lexical Entrainment in Human-Robot Interaction: — Can Robots Entrain Human Vocabulary? —", The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS2009), pp.3727-3734, 2009.

[10] D. Sakamoto, T. Kanda, T. Ono, M. Kamashima, M. Imai, and H. Ishiguro: Cooperative embodied communication emerged by interactive humanoid robots, International Journal of Human-Computer Studies, Vol. 62, pp. 247-265, 2005.

[11] S. Garrod and A. Anderson. Saying what you mean in dialog: A study in conceptual and semantic co-ordination. Cognition 27, 1987.

[12] S. E. Brennan and H. H. Clark: "Lexical choice and conceptual pacts in conversation," Journal of Experimental Psychology: Learning, Memory,and Cognition, vol. 22, pp. 1482-1493, 1996.

[13] R. Porzel et al. "How Entrainment Increases Dialogical Effectiveness," in Proceedings of the IUI'06 Workshop on Effective Multimodal Dialogue Interaction, 2006.

[14] S. E. Brennan: "Lexical Entrainment in spontaneous dialog." In Proceedings of the International Symposium on Spoken Dialogue, pp. 41-44, 1996.

[15] J. Gustafson, A. Larsson, R. Carlson, and K. Hellman: "How Do System Questions Influence Lexical Choices in User Answers?" In Proceedings of Eurospeech, Rhodes, Greece, pp. 2275-2278, 1997.

[16] S. Tomko and R. Rosenfeld: "Shaping spoken input in user-initiative systems," In The 8th International Conference on Spoken Language Processing, pp. 2825-2828, 2004.

[17] D. McNeill,: "Psycholinguistics: A new approach." Harper & Row. (1987)

[18] C. Moore and P. J. Dunham, Joint Attention: Its Origins and Role in Development. Hillsdale, NJ: Lawrence Erlbaum, 1985.

[19] C. Breazeal and B. Scasselati, "A context-dependent attention system for a social robot," in Proc. IJCAI'99, pp. 1146–1151, 1999.

[20] R. A. Brooks, C. Breazeal, M. Marjanovic´, B. Scassellati, and M. M. Williamson, "The Cog project: Building a humanoid robot," in Computation for Metaphors, Analogy and Agents, C. Nehaniv, Ed. Berlin, Germany: Springer, 1998, vol. 1562, Springer Lecture Notes in Artificial Intelligence.

[21] H. Kozima and A. Ito, "Toward mindreading by an attention-sharing robot," in Proc. Third Int. Symp. Artificial Life and Robotics (AROBO III'98), 1998, pp. 478–481.

[22] Morris, D.: Manwatching: A field guide to human behaviour. London: Jonathan Cape. 1977.

[23] J. D. Matarazzo and A. N. Wiens, The Interview: Research on its anatomy and structure, Aldine-Atherton(Chicago), 1972.

[24] A. Kendon, Movement coordination in social interaction: Some examples described, Acta Psychologica, 32, pp. 101-125, 1970.

[25] W. S. Condon and W. D. Ogston, A segmentation of behavior, Journal of Psychiatric Research, 5, pp. 221-235, 1967.

[26] W. S. Condon and L. W. Snder, Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. Science, 183, pp. 99-101, 1974.

[27] A.E. Scheflen, The significance of posture in communication systems, Psychiatry, 27, pp. 316-331.

[28] E. J. Charny, Psychosomatic Manifestations of Rapport in Psychotherapy, Psychosomatic Medicine, 28, pp. 305-315 (1966).

[29] K. Nakadai et al. Real-Time Auditory and Visual Multiple-Object Tracking for Robots, Int. Joint Conf. on Artificial Intelligence (IJCAI'01), pp. 1425-1432, 2001.

[30] Y. Sumi, T. Matsuguchi, S. Ito, S. Fels, and K. Mase, Collaborative capturing of interactions by multiple sensors, Int. Conf. on Ubiquitous Computing (Ubicomp), pp. 193-194, 2003.

[31] D. Schulz, D. Fox, and J. Hightower, "People Tracking with Anonymous and ID-Sensors using Rao-Blackwellised Particle Filters," Int. Joint Conference on Artificial Intelligence (IJCAI), pp. 921-926, Aug. 2003.

[32] K. Shinozawa, T. Miyashita, M. Kakio and N. Hagita: "User specification method and humanoid confirmation behavior", IEEE International Conference on Humanoid Robots, 2007