USE OF NON-VERBAL INFORMATION IN COMMUNICATION BETWEEN HUMAN AND ROBOT

Masao Yokoyama, Kazumi Aoyama, Hideaki Kikuchi and Katsuhiko Shirai

Department of Information and Computer Science, Waseda University

3-4-1,Okubo,Shinjuku-ku, Tokyo 169-8555 Japan

Email:kazumi@shirai.info.waseda.ac.jp

ABSTRACT

In this research, we consider the use of non-verbal information in human-robot dialogue to draw the communication ability of robots close to that of human beings. This paper describes analysis of output timing of non-verbal information for the interactive dialogue between human beings. Moreover, we analyzed influences of output timing by controlling it in dialogue with a CG robot. As a result, we clarify the strength of constraint and naturalness of various types of non-verbal information. We also confirm that appropriate output timing of non-verbal information is the start of utterances. This is the same as in human-human dialogue. As a result, non-verbal information made speaker-change smooth for the CG robot.

1. INTRODUCTION

For the enhancement of human-robot interaction, handling of non-verbal information is an important consideration. It is important that human communication ability is implemented by the robot because it is common to unspecified users. In this case, it is difficult to implement all of human communication ability. So up to the present, it has been general practice to implement only speech as a means of communication. But communication with a robot is likely to become unnatural under conditions in which only speech is used. This is because nonverbal visual cues play a big role in human communication. Non-verbal information is important for improving the smoothness of communication. The types of non-verbal information that is used in usual communication between human beings includes extensive general ideas, for example, physical actions such as gesture, posture, distance, appearance such as dress and decoration, physical characteristics and so on. It has been shown that 65%[1] or 93%[2] of messages that are transmitted between humans are non-verbal in nature. It is important to incorporate non-verbal information in communication with robots in order to enhance information exchange and naturalness.

Ekman classified physical actions made by humans as is given in Table 1[3].

We think that it is possible for a robot to have communication ability, which is closer to that of a human by applying this non-verbal information to physical actions made by the robot.

Emblem	This can be translated into the speech words. This is also called “symbol” or “sign.”
Illustrator	This emphasizes and elaborates contents of an utterance.
Affect Display	This means face expressions and gestures with emotions.
Regulator	This controls turn-taking and makes flow of conversation smooth.
Adapter	This means actions to adapt situations.

Table 1: Classification of physical –action by Ekman

In this research, we focus on the model for use of non-verbal information, especially regulator control. The regulator controls the actions that decide turn of an utterance and serves make the flow of a conversation smooth. Head movement and eye-gaze, are examples of regulator control. It seems that there are some differences of the model between human-human dialogue and human-robot dialogue (for example, output timing, relation to turn-taking, pattern of expression, and so on all show some differences). In Section 2, we analyzed the tendency of output timing of non-verbal information in the dialogues between human beings. Similarly, in Section 3, we analyzed the tendency of output timing of non-verbal information in dialogues between humans and a CG robot. Furthermore, we proposed some methods of controlling various kinds of non-verbal information in human- robot dialogue.

2. OUTPUT TIMING OF NON-VERBAL INFORMATION BETWEEN HUMANS

2.1. Experiment

To analyze the output timing of non-verbal information in a conversation between human beings, we recorded subjects’ conversations as well as their concurrent physical actions. As mentioned above, Regulators in non-verbal information are important in natural turn taking, so we set up situations in which the “System” prompts the “User” to speak. In the experiment we used 5 kinds of pair utterances as stimuli (see Table 2) [6].

Subjects consisted of 10 college students. We asked the subjects to memorize 10 pair utterances indicated in Table 2. The task for each subject was to speak to the “User” with natural physical actions, for example, one subject as the role of the system moved one of his hands forward with the palm up to

indicate “Please”. We recorded each subject’s responses on VTR. Audio-data and image-data in VTR were transferred to our workstation digitally. This made it possible to analyze the appearance time of verbal information and non-verbal information using our dialogue analysis tool. The main non-verbal cues that were focused on included eye-gaze, eye-blinking, head movement, and hand movement. The reason for focusing on these cues was because of their regulatory role in turn-taking.

Kinds of pair utterances	Example
Greeting – Greeting	“Hello” - “Hello”
Request – Approval or reject	“Please raise your hands” - “Yes” or “No”
Call – Response	“Let’s begin” - “O.K.”
Yes/No question – Answer	“Are you Japanese?” - “No.”
Question – Answer	“Where’s your house?” - “My house is ***”

Table 2: Examples of pair utterances

2.2. Result and discussion

In order to discover the tendency of output timing of non-verbal information, we analyzed the frequency of appearance of non-verbal information in reference to the relative time since the start of the utterance (Results are depicted in Fig. 1). In addition, the frequency of appearance of non-verbal information was analyzed in reference to the relative time since the start of the keyword in the utterance. A keyword is the most important content word in a sentence. In the case of an illustrative sentence “Are you Japanese?”, the keyword is “Japanese”. The results of the experiment revealed that frequency of non-verbal information in reference to the start of the utterance had the clearest tendencies. The keyword is indicated to the subjects by the experimenter.

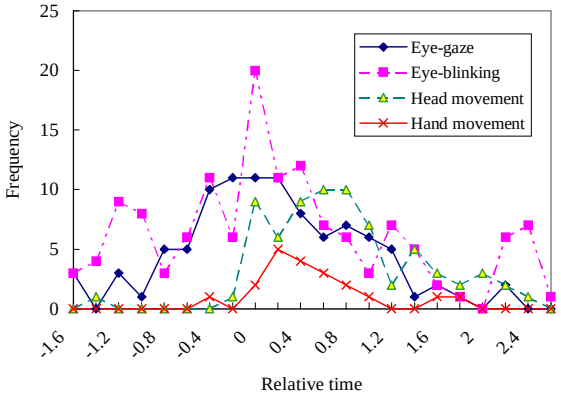

Fig.1: Frequency of non-verbal information against start of utterance

Fig.1 shows that the frequency peaks of non-verbal information occur at just or immediately after the start of utterances. Moreover, detailed observation of the figure reveals that the occurrence of the frequency peaks depends on the type of non-verbal information. Among the four types of non-verbal information, eye-gaze and eye-blinking were shown to be most likely to occur at the start of an utterance, whereas, head movement and hand movement are most likely to occur immediately after the utterance. In the next section, an additional experiment was conducted in which subjects talked to a CG robot. Methods of controlling various types of non-verbal information are proposed.

3.CONTROLLING OF NON-VERBAL INFORMATION

3.1. Constraint on speaker change

It is very important for users that they can speak in free timing to a dialogue system. Under conditions in which free timing is restricted users can experience stress and decrements in efficiency can occur. Conversely, in the case in which there are no constraints on free timing, speaker change is not smooth. In this research, the strength of the constraint on speaker change was evaluated for various types of non-verbal information. We used the length of pause for evaluation of each non-verbal information. The length of pause is measured from the system’s end of utterance to the user’s start of utterance. A shorter pause may not always be good. But, a longer pause results from needlessly long silence. It is thought to occur when users do not know when to speak. So it seems that length of pause effects strength of constraint.

3.2. Experiment

(1)System

It is very difficult to apply control mechanisms for the various types of non-verbal information for a real robot, so in this experiment the effects of various types of non-verbal information was examined using a simulated robot by 3D CG, “Doraeman” (see Fig.3). The system used in the experiment can display 3D CG (indicating non-verbal information) as well as present and record acoustic speech signals. This provides the system with a multi-modal interface (see Fig.2).

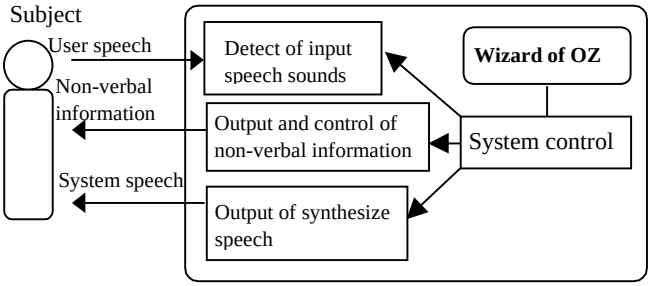

Fig.2: Outline of system used for experiment

(2) Method

The experiment was set up in such a way that the system would prompt the user to speak under various conditions. The signs used for speaker-change are given in Table 3. In this experiment 6 different types of non-verbal information were considered (see Table 3 and Fig.3).

In the experiment, the variety of non-verbal information (given in Table3) and the variety of output timing (given in Fig.4) expressed by the CG robot was manipulated. Subjects consisted of 20 college students. In this experiment, combinations of 6

kinds of non-verbal information and five kinds of output timing, in total 30 conditions are compared. Subjects were told to do two types of tasks; responding to and evaluating CG robot utterances. Subjects were told to respond to the CG robot utterances freely. The dialogue was recorded for later analysis. 10 pair utterances (Table2) were used to avoid biases that may be caused by certain types of utterances. Each utterance was assigned randomly to each subject with the same probability. Therefore, for each utterance, 2 of the 20 subjects had the same condition. Subjects were later asked to evaluate the naturalness of speaker change and the easiness of answering to the CG robot utterances. In conditions blocked by non-verbal information type, subjects evaluated 5 different random timing patterns. Subjects were told non-verbal information would be presented with various timing but were not told the order of the presentation. Five step (1. Poor – 5. Good) measurement was applied for the evaluation.

Categories			Output method
Non Human	Beep	Auditory	Sound
	Illustration	Vision	Indicate illustration of mouth or characters
Human	Eye-gaze	Vision	Turn his eyes to something
	Eye-blinking	Vision	Open and Close his eyelids
	Head movement	Vision	Move head vertically
	Hand movement	Vision	Move one of his hands forward with a palm up

Table3: Non-verbal information of turn giving from system to user

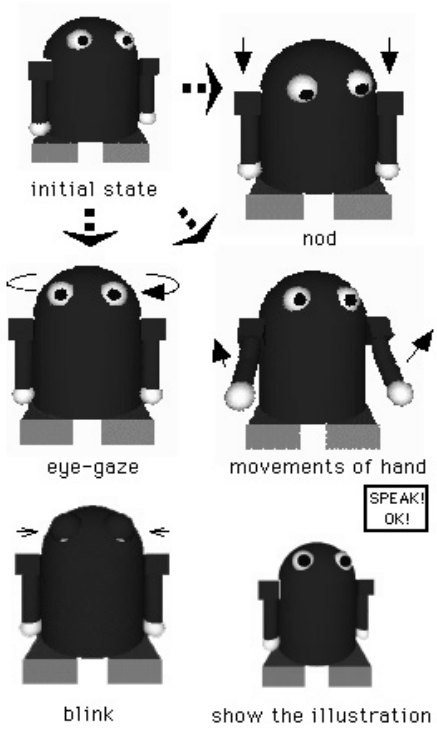

Fig.3: Variety of non-verbal information expressed by CG robot

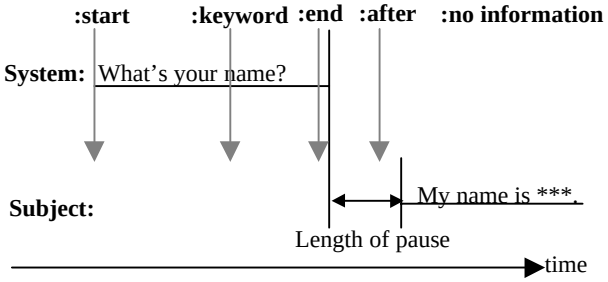

Fig.4: Output timing of utterance and non-verbal information

In the experiment, the robot uttered with non-verbal information in various output timing and then he uttered without non-verbal information.

3.3. Result and Discussion

3.3.1. Comparison of strength of constraint by length of pause

Fig.5 shows the length of pause after the end of the system's utterances with and without each type of non-verbal information. In the case of utterances without non-verbal information, the length of pause shows a tendency of increasing generally so that unnatural silence was easy to bring about. This means that subjects repeated the exchange of the pair of utterance which is the same so that they often waited for the robot's action, especially, in the case of utterances without beep and illustration. For these cases the length of pause is longer than other cases. So we can say that these kinds of non-verbal information, that are not used by humans, have a strong constraint on speaker-change. Also we can say that the kinds of non-verbal information which human beings use such as gaze have a week constraint on speaker-change. The order of types of non-verbal information in the strength of constraint is hand-movement, eye-blinking, gaze, and head-movement. The result of the five-step evaluation on the naturalness of speaker-change indicates that beep and illustration are lower than other kinds of non-verbal information as is shown in Fig.6. The order of types of non-verbal information in the naturalness of speaker-change is eye-blinking, gaze, hand movement, and head movement.

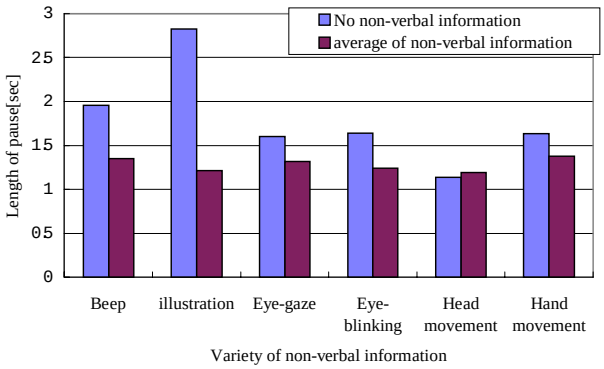

Fig.5: Length of pause and controlling non-verbal information

From these results, we can propose the following method of controlling non-verbal information. For the improvement of smoothness of speaker-change in conversation, the robot should use a type of non-verbal information, which is not used by humans, to prompt the user to utter by force. For the

improvement of naturalness of speaker-change, the robot should use a type of non-verbal information, used by humans, that has the most suitable strength of constraint on speaker-change.

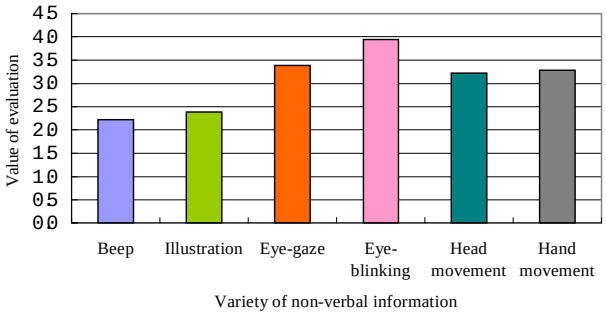

Fig.6: Result of questionnaire on naturalness of speaker-change

3.3.2. Output timing of each non-verbal information

Fig.7 indicates the length of pause of output timing for each type of non-verbal information. In the case in which analysis is limited to non-verbal information employed by humans, the length of the pause after an utterance that accompanies output of non-verbal information at "the start of an utterance" or "the start of a keyword" is the shortest (Only for eye-gaze is the length of the pause at "the end of the utterance" shorter than the one at "the start of the utterance"). For eye-gaze and eye-blinking, the length of pause in the case of "the start of an utterance" is shorter than the one in the case of "the start of a keyword". On the other hand, in hand-movement and head-movement, there is an opposite tendency. As indicated in section 2.2, in human-human dialogues, there is some tendency that gaze and eye-blinking frequently appear at the start of an utterance and hand-movement and head-movement frequently appear a little bit after the start of an utterance. This suggests that it is effective to apply a model of human non-verbal communication to a robot.

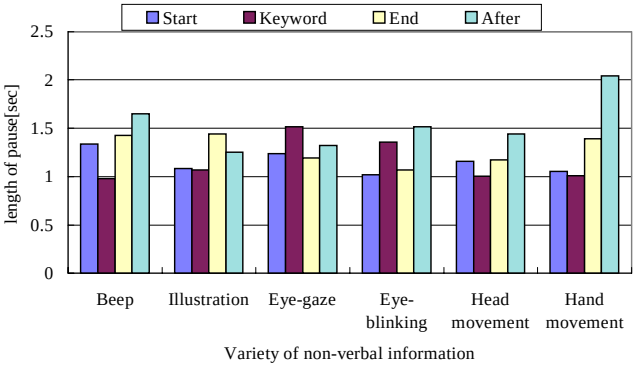

Fig. 7: Length of pause against output timing of each type

Table 4 gives a summary of evaluation on non-verbal information. In the case of non-verbal information that humans use, the evaluation result of output in fast timing at the start of an utterance or keyword was good. There were some comments made by subjects that the robot's action after an utterance disturbed their utterance. In the case of non-verbal information that is not used by humans, such as beep and illustration, the evaluation result of output after an utterance was good. This result differs from the result of comparison of the length of pause. This may be due to the daily experience with systems that prompt users to speak (eg. answering the telephone). Subjects commented that the robot's action during an utterance often confuses them to which direction to look. From these results, we can propose the following method of controlling non-verbal information.

In the case of types of non-verbal information that humans do not use, output at the start of an utterance is suitable for the improvement of smoothness of speaker-change in conversation, and output at the end of an utterance is suitable for the improvement of naturalness of speaker-change. On the other hand, in the case of types of non-verbal information that humans use, output at the start of an utterance is suitable in all cases.

Non-verbal information	Constraint	Evaluation
Beep (auditory)	Strong	Timing : end of utterance Interference with utterance.
Illustration (vision)	Strong	Timing : end of utterance Difficulty focusing on the communication signal.
which humans use Eye-gaze, etc. (vision)	Weak	Timing : start of utterance or keyword Natural, Human like.

Table 4: Summary of evaluation on non-verbal information

4.CONCLUSION

In this research, we aim at the realization of the human-friendly robot to human beings by using non-verbal information in communication. We focused on the regulator control, that controls turn-taking and makes flow of conversation smooth. We did 2 experiments, one was research concerned with the output timing of non-verbal information between human beings, the other one was concerned with an evaluation of output timing and the various types of non-verbal information useful in human-Robot communication. In these experiments, we clarified the strength of the constraint on speaker change and naturalness for each type of non-verbal information. Moreover, we confirmed that appropriate output timing of non-verbal information that humans use is the start of utterances, which is the same as in human-human dialogue. As a result, non-verbal information made speaker-change more smoothly for the CG simulation robot.

5. REFERENCES

1. Birdwhistell,R.L.: Kinesics and Context, Univ. of Pennsylvania Press(1970).

2. Mehrabian,A., Williams,N.: Non-verbal concomitants of perceived and intended persuasiveness, J.of Personality and Social Psychol. 13, pp.37-58(1969)

3. Ekman,P., Friesen,W.V.:the repertoireof nonverbal behavior semiotica,1,pp.49-98(1969)