DIFFERENCE IN VISUAL INFORMATION BETWEEN FACE TO FACE AND TELEPHONE DIALOGUES

Yuri Iwano *Yosuke Sugita* *Yusuke Kasahara* *Shu Nakazato* *Katsuhiko Shirai*

School of Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku, Tokyo 169, Japan
iwano@shirai.info.waseda.ac.jp

ABSTRACT

In this research, we analyzed conversations between a pair of subjects, under two conditions. One is face to face conversation that has a visual contact, and the other is conversation through telephone line that has not. From the recorded videotape we extracted the subject's actions especially focusing on the head movements. By comparing the dialogues under two conditions, it seems that there are two types of head movements, one is intended to give a response to his partner and the other is to send some signal. We are going to analyze how visual information contributes in spoken dialogue perceptions, and possibility of adopting it in a multi-modal human interface.

1. INTRODUCTION

Using the information not only from a speech signal but also knowledge of linguistic characteristics of spoken dialogue and dialogue management method are important for the accomplishment of natural dialogue on a computer. Dialogue is an interactive communication of information mainly based on speech. For instance, considering conversation using a telephone, we can have natural conversations without actually seeing each other. However, in practical conversations, visual information such as gesture, facial expression, and head movement clearly makes it much smoother and more natural. Therefore, in the more natural human interface that can use multiple modalities, visual information becomes important as well as voice information.

Most researches related to analysis of spoken dialogue are based on only auditory information. We are trying to clarify how human uses the knowledge of spoken dialogue management by dealing with more natural communication that includes visual information. Above all, by using visual information we can deal with a listener's attitude against the speaker that cannot be done by using only auditory information.

Unlike spoken language, non verbal behavior has no standard rule and social and individual differences are large. Among these human movements in a conversation, the head movement is clear and can be classified easily. Therefore it is one of the visual information that we can obtain a statistical result at least by dealing with a certain amount of data, even there are individual differences.

The research carried out here is all based on Japanese language. It is said that during conversation, Japanese speakers give more responses like nodding than English speakers. The way Japanese speakers answer questions are different from English speakers. Therefore, although the result of this research is culturally dependent, still the importance of visual information can be shown.

2. EXPERIMENT

2.1. Experimental conditions

In this research, we analyzed conversations between two subjects, under two conditions. One is face to face conversation that has a visual contact, and the other is conversation through telephone line that has not. The scene of the "Face to face conversation" is shown in Figure 1. To find out the role of visual information we had to analyze dialogues that are completely speech oriented. It is important to reduce an unnatural feeling of the experimental situation for the subject. Therefore we simulated a telephone line that should be familiar as we use it in our daily life. Two subjects were actually sitting in a same room but they cannot see each other (Figure2).

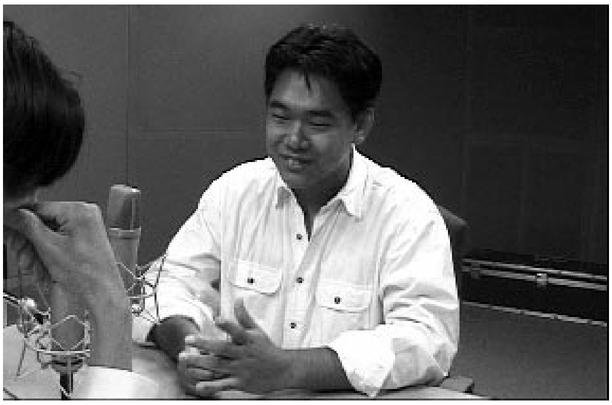

Fig 1. Face to face conversation

We asked 10 students who know each other to make a pair, and recorded five dialogues for each condition. The experiment was done in a recording studio. A task given to the subjects is to make a scenario of a comic strip. We asked them to decide not only the story as in detail as possible, but also the title and the copy of it. During the experiment we used two video cameras to record each subject's actions and DAT to record high quality audio data (16bit 48KHz) suitable for various speech processing.

Fig 2. Conversation using telephone

Same pair did "Face to face" and "Telephone" dialogues experiment on a different day. To avoid the effect of the experiment's sequence of two conditions, it was counterbalanced across all experiments. We did not assign any limitation on their conversations. They can speak whatever and as long as they want. When the story and the title are settled the experiment ends. The dialogue length varies from 15 minutes to 41 minutes and the total length of "Face to face" experiment was approximately 117 minutes for five pair and the other was 93 minutes.

2.2. Preparation for analysis

All recorded audio data are transferred to our workstation digitally. Each dialogue is transferred as one file so we can deal with time information. All the dialogue information are written in a text file containing who is speaking, beginning and ending time of it and the content of the utterance. As the time information of the utterance has a common time scale, we can deal with analysis using time information. Head movements were extracted visually from the recorded videotape.

3. ANALYSIS OF HEAD MOVEMENTS

Head movements such as nodding, occur in our conversation even under the situation that we cannot see each other. Therefore, we can say that these movements are not always intend to give a response to his partner. We can assume that there are two different types of head movements. As mentioned above, one is an unconscious movement or a movement for confirmation of one's perception. The other is more communicative actions intend to give some sort of response to his partner. By analyzing the head movements of "Face to face" and "Telephone" dialogues we classify these two types of head movements and discuss about the difference of them. The results suggest that the head movements have several kinds of roles in the dialogue which assist the communication and make the way of thinking more smooth among subjects. We define the head movements here as vertical, horizontal and inclined movements. We ignore the movements accompanied by the whole body movements.

Table 2. Appearance freq. of head movement labels

Labels	Face to Face		Telephone	
	with*	without	with	without
Acknowledgment	423	271	238	31
Agreement	248	42	143	8
Request	322	1	122	0
Emphasis	429	4	156	3
Repeat	32	0	14	0
Denial	19	3	5	0
Skeptical	23	8	4	0
Total	1496	329	682	42

*with or without utterance

4. RESULT

We analyzed the head movements of five 'Face to face" and five "Telephone" dialogues. We will show the appearance frequency of head movements in Table 1. We also categorized it by whether the head movement accompanied utterance or not. As expected, the head movement appeared more when the subject could see each other, even considering the difference of the dialogue length between "Face to face" and "Telephone" situations. It is interesting that approximately 18% of the head movements of "Face to face" dialogue did not accompany utterances compared with 5.6% of "Telephone" dialogue. As head movements can be used to express one's intention to his partner, this result implies the importance of visual information that could potentially affect the dialogue without voice information.

Table 1. Appearance frequency of Head movements

Categories		Appearance frequency	
Movement	Utterance	Face to face	Telephone
Vertical	with	1451	673
Vertical	without	314	42
Horizontal	with	21	4
Horizontal	without	3	0
Inclined	with	24	5
Inclined	without	12	0
Total		1825	724

Now we define a further category for head movements. Head movements can be classified according to the meaning of its movement. We defined the following categories and we will show the number of appearances in Table 2.

- "Acknowledgment" Response of acknowledgment.
- "Agreement" Represents clear affirmative response such as YES.
- "Request" Movement that appears at the end of utterance requesting some response from his partner. For example, asking a question.
- "Emphasis" Vertical movements to emphasize the utterances.
- "Repeat" Movement appeared when the subject repeated his partner's utterances.
- "Denial" Movement to express denial.
- "Skeptical" Movement to express skeptical feelings.

Most of the horizontal head movements are intended to give negative response and inclined head movements to give skeptical response. As the horizontal and inclined head movements have quite clear meanings and the appearance rate is low, we will focus on vertical head movements in this paper.

From Table 2, head movements that do not accompany utterances especially appear when they mean “Acknowledgment”. In “Face to face” dialogue, 39.0% (271/694) of the head movements did not accompany utterance. 12.9% of the vertical head movements representing “Agreement” did not also have utterances. Even in “Telephone” dialogue 5.3% of the “Agreement” was without utterance. In these cases, only visual information channel was available to express one’s intention to his partner.

4.1. Differences among individuals

Here we will see the individual differences of probability of involving head movements. Table 3 shows the functional labels of the dialogue, and its number of appearances. Functional labels show utterance’s role in the dialogue.

Table 3. Appearance freq. of dialogue labels for all utterances with and without head movements

Labels	Face to Face	Telephone
Acknowledgment	985	1036
Agreement	409	387
Request	748	647
Repeat	153	96
Denial	38	38
Total	2333	2204

Utterance corresponding to “Acknowlegdement” is more frequent and uses speech channel in “Telephone” dialogue. However, in “Telephone” dialogue case as shown in Table 4, the utterance accompanying head movements is less.

Figure 3 to 6 show the probability of involving head movement within the same category of utterances. In this analysis we exclude head movement that has “Emphasis” label since it does not have one to one correspondence to functional label of dialogue. Each plot represents the probability of each speaker. Height of the diamond represents 95% confidence interval and the top and bottom of the quantile box represent the 75th and 25th quantiles. The lines above and below each box shows the 10th and 90th quantiles.

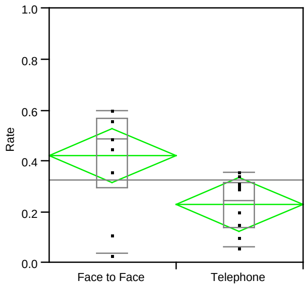

Fig 3. Acknowledge-

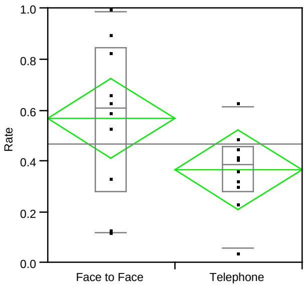

Fig 4. Agreement

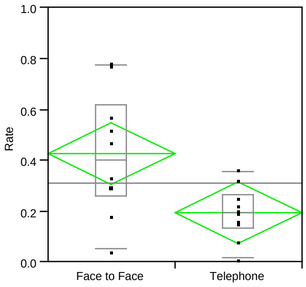

Fig 5. Request

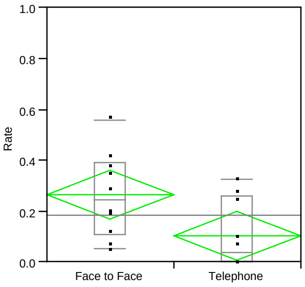

Fig 6. Repeat

Table 4. Prob. of involving head movements

Labels	Face to face		Telephone	
	Mean	Stddev	Mean	Stddev
Acknowledgment	43%	0.20	23%	0.10
Agreement	57%	0.30	37%	0.16
Request	42%	0.24	19%	0.10
Repeat	26%	0.17	10%	0.13

Table 4 shows the mean value and the standard deviation of the probability of involving head movements among 10 subjects. From these figures, we can confirm the probability of the utterance involving vertical head movement gets lower when we cannot see each other. Furthermore The variance of the probability of each speaker is larger in “Face to Face” compared with “Telephone” dialogue. Differences in the way of speaking among individuals appear more distinctively in the head movements that increase by the fact that one can see each other.

4.2. Transition of the dialogue

In this section, we will see whether the existence of the head movement affects the transition of the dialogue. We focused on the utterance intended to request a response, and see whether the following utterance is a response to its request or not. Table 5 shows the probability of getting some sort of response to the prior utterance on both “Face to Face” and “Telephone” dialogues, depending on whether the utterance of request involved vertical head movement or not.

Table 5. Transition of the dialogue

Head mov.	Face to face(%)		Telephone(%)	
	Response	Other	Response	Other
With	61.3	38.7	62.0	38.0
Without	50.7	49.3	50.4	49.6

At a glance, “request of response” with vertical head movement can obtain more response than without having head movement. Since the characteristics between “Face to face” and “Telephone” dialogues are almost the same, visual information has few effect on the transition. This result shows the request utterance with or without vertical head movement has different nature and the subjects behave quite independently to the existence of visual channel. The request involving head movement requires some response more strongly.

More precise analysis of the transition may be done by constructing a transition model by using Ergodic HMM[4].

4.3. Analysis of overlap

We analyzed whether vertical head movements are used to obtain a timing to speak. In Table 6 we counted all the utterances overlapped previous partner's one that accompanied vertical head movement.

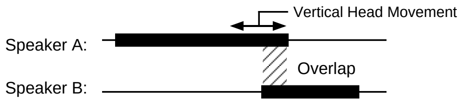

Fig 7. Overlap situation

Table 6. Rate of overlap to the utterance including vertical head movement

Labels	Face to Face	Telephone
Number of appearance (All overlap)	280 (644)	88 (435)
Rate	43.5%	20.2%

From Table 6, within all the overlap cases, 43.5% and 20.2% of overlap occured when the previous utterance involved vertical head movement in "Face to face" and "Telephone" dialogues respectively. This type of overlap occupies large part of whole overlaps in "Face to face" case.

If we see the probability of an overlap after a vertical head movement, (in "Face to Face" and "Telephone" dialogues) they are 19.3% (280/1451) and 13.1% (88/673). Therefore, the occurrence probability of overlap after the type of utterance which accompany the vertical head movement is not so much different for both cases. However, it is clear that in "Face to face" case, vertical head movement plays an important role to induce overlap. We can say that we are using our partner's vertical head movement to obtain the timing to speak earlier than our partner stops speaking.

From the other point of view, in Table 7 we will show the time information of the dialogues we used in this experiment. The numbers in the table represent the percentage of the time. "Utterance" is the rate of the speaker's speaking time. "Overlap" is the rate of time that both speakers are speaking. "Silence" is the rate of the time that both speakers remain silent. That is "Utterance A" + "Utterance B" + "Silence" - "Overlap" = 100. From this table, comparing the same pair's "Face to face" and "Telephone" dialogues, "Silence" time rate is similar. Thus, there is little difference in speaking time depending on whether we can see each other or not. On the other hand, the rate of overlap time gets bigger when we can see each other. This fact gives support to the previous result.

5. CONCLUSION

We analyzed the head movements in dialogues done face to face and on telephone line. We confirmed that head movements appear more when they face each other.

Table 7. Comparison of the time information

Speaker	Face to Face(%)			Telephone (%)		
	Utter-ance	Over-lap	Sile-nce	Utter-ance	Over-lap	Sile-nce
A	28.6	2.9	47.9	32.7	1.5	47.6
B	26.5			21.2		
C	27.7	2.3	46.0	29.3	1.2	45.4
D	28.6			26.5		
E	36.6	4.8	43.2	23.0	4.6	43.6
F	25.0			37.9		
G	30.7	3.9	40.3	31.2	2.7	43.2
H	32.9			28.4		
I	37.4	3.8	39.3	41.5	4.0	31.3
J	27.2			31.3		
Average	30.1	3.7	43.6	30.2	2.6	42.3

We assumed at the beginning that there might be two types of head movements according to whether it is a movement intended to give a certain response or not. From the view point of the person who is moving his head, some of the head movements are intentionally made. On the other hand as there exist head movements that does not accompany utterances, there is an opposite case that information is acquired by mainly visual information. From the view point of the person who is watching head movements, visually perceived head movements might not only transmit partner's affirmative, negative or skeptical response, but also it seems that it is used as a chance to start speaking even by overlapping to the previous utterance.

We found out several differences in the dialogue, depending on whether we can see each other or not. Considering head movement as not only a signal but also an interactive action, we can recognize the effect of visual information to the dialogue. In constructing an effective man machine interface, in conjunction with speech information, the use of visual information will become important. And the head movements in conversation are useful element to study the role of each utterance and the dynamical structure of the dialogue.

REFERENCES

[1] R. A. Bolt: "The Integrated Multi-Modal Interface", Trans. of the IEICE(Japan), Vol.J-70-D, No.11, pp.2017-2025, 1987.

[2] Senko K. Maynard: "Kaiwa Bunseki", Kuroshio publisher, 1993 (in Japanese)

[3] Keiko Watanuki, Fumio Togawa: "Some signals of emotional arousal: Analysis of conversations using a multimodal interaction database", Eurospeech 95, pp.1165-1168, 1995.

[4] Katsuhiko Shirai: "Modeling of Spoken Dialogue with and without Visual Information", Proc, of ICSLP96, vol1,pp188-191, 1996.

[5] Yuri Iwano, Shioya Kageyama, Emi Morikawa, Shu Nakazato, Katsuhiko Shirai: "Analysis of head movements and its role in spoken dialogue", Proc. of ICSLP96, vol4, pp.2167-2170, 1996.