Influence of Pause Length on Listeners' Impressions in Simultaneous Interpretation

Hitomi Tohyama[†] *Shigeki Matsubara*[*]

[†]Graduate School of Information Science, Nagoya University, Japan
[*]Information Technology Center, Nagoya University, Japan
[†]hitomi@el.itc.nagoya-u.ac.jp

Abstract

We have been attempting to realize simultaneous machine interpretation. However, determining the interpreting utterance timing is as difficult as determining translation units. This remains a major concern for the development of such a speech translation system. It is also crucial for the system's users that the speech generated by the system is clear and easy to listen to. In this paper, we focus attention on the pauses that partly characterize simultaneous interpreters' utterances. We attempt to analyze the results of an experiment conducted using 31 subjects on the relationship between listener-friendliness and the length of pauses in speech, using the CIAIR simultaneous interpretation database as the data source. The results generated some knowledge about listener impressions of simultaneous interpretation, which will be helpful for realizing simultaneous machine interpretation.

Index Terms: cross-lingual, simultaneous interpretation, machine translation, bilingual speech corpus

1. Introduction

In speech, a pause is an essential element for producing the rhythmic aspect of spoken languages, with the rhythmic aspect being closely related to humans' listening skill and our understanding of the semantic contents of the speech [1]. It is known that humans instinctively feel that phenomena accompanied by rhythmic aspects are secure [2], and that we have a nature that is attracted by anything rhythmic and synchronized with it. However, since simultaneous interpreters take pauses in order to wait for a speaker's next input of necessary information before starting their interpretation, their utterances sometimes give us different impressions from usual utterances, and those pauses tend to influence listeners' impressions. Therefore, in this paper we investigate the characteristics of listener-friendly simultaneous interpretation.

This paper is presented as follows: Section 2 describes the importance of the investigation to realize simultaneous machine interpretation. Section 3 illustrates the process and results of an experiment performed using 31 human subjects. In Section 4, we analyze the relationship between listener impressions and the periods of pauses in simultaneous interpretation contexts based on the experimental results, and discuss our observations. Section 5 includes our conclusions and a brief mention of future work.

2. Importance of Controlling Pauses

We have conducted an experiment using 31 subjects to clarify the relationship between listener impressions and pauses in simultaneous interpretation contexts, with the aim of realizing a simultaneous machine interpretation system [3].

Controlling the length of pauses is important in determining the time required to generate translations and the timing of its outputs for a simultaneous machine interpretation system. For example, just at the moment the machine detects a part of an utterance that is possible to translate, it is one of the machine's roles to generate and output the translation simultaneously. Yet, a long pause is required for detecting parts of utterances that can be translated, and such a pause is likely to influence the subjects' listener impressions. Thus, it is crucial to control the length of pauses.

3. Experiment

The purpose of our experiment is to investigate the relationship between listener impressions and pauses in simultaneous interpretation contexts. We used the CIAIR Simultaneous Interpretation Database of Nagoya University [4] for our experiment. The size of the transcribed data is about one million words, a simultaneous translation corpus that deserves to be classed as one of the largest in the world. The transcription was produced based on the standard transcription rules of the Corpus of Spoken Japanese (CSJ) developed by the National Japanese Language Research Institute [5]. The transcribed data consists of utterance units, and each utterance unit is divided by silences of more than 200 msec. In this paper, we define a period of continuous silence exceeding 200 msec as a pause. We visualize the emergence of pauses using a time-charting tool (refer to Fig. 1).

3.1. Outline of Experiment

We extracted twelve interpretation cases from corpus data comprising freely spoken lectures (hereafter, "A-style lectures") and nine interpretation cases from corpus data consisting of lectures read from a script (hereafter, "B-style lectures"), twenty-one interpretation cases altogether. These cases consist of speech sounds from lectures on various themes such as politics, economics, and culture, given in English and those of their English-Japanese simultaneous interpretations.

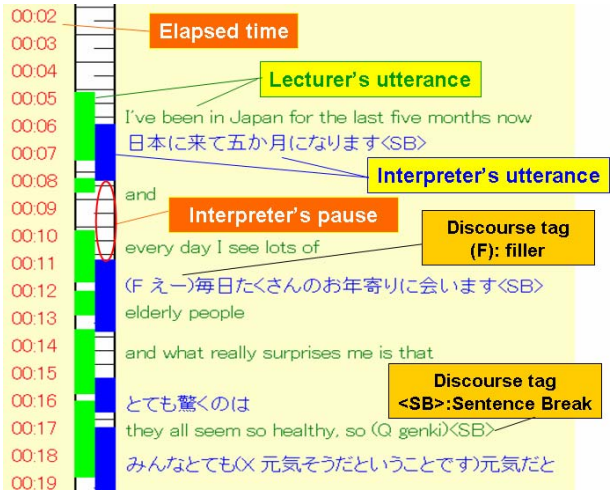


Figure 1 *Time chart tool of CIAIR simultaneous interpretation database*

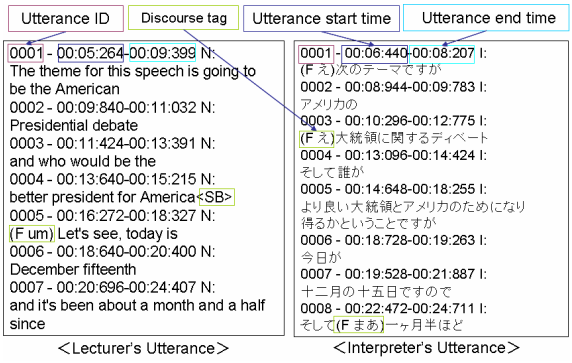


Figure 2 *Sample of text data (English- Japanese interpreting data)*

We added discourse tags to the described speech data and each utterance was marked with the time when the speech began and the time when it ended, by which we could calculate the length of each pause and the speed of each utterance (Fig. 2).

As the data of A-style lectures were data of simultaneous interpretation of lectures in which lecturers talked off the top of their head, the average of their speech speed was 9.1 mora/s. On the other hand, because the data of B-style lectures were data of simultaneous interpretation of lectures in which lecturers gave a talk while referring to their notes, the average of their speech speed in B-style lectures was 11.2 mora/s. This means the time pressure on interpreters is high in the case of B-style lectures.

We prepared 21 interpretation cases by extracting respectively speech sound of 60 sec from each speech beginning five minutes into the speech and ending six minutes into it passed as an interpretation case. The subjects evaluated the samples according to a five-grade system: very difficult to listen to, slightly difficult to listen to, undecided, slightly easy to listen to, and easy to listen to. All 31 subjects were native Japanese speakers.

3.2. Experimental Results

We evaluated each interpretation case using a five-point scale. The highest score was 5, and an interpretation case receiving this evaluation was easiest to listen to. On the other hand, a case receiving a score of 1 was considered most difficult to listen to. The listeners' evaluations toward both lecture styles A and B were scored and the mean scores for each interpretation case were calculated. The graphs in Fig. 3-(1) and 3-(2) show the ranking of the mean score of those simultaneous interpretation cases in descending order.

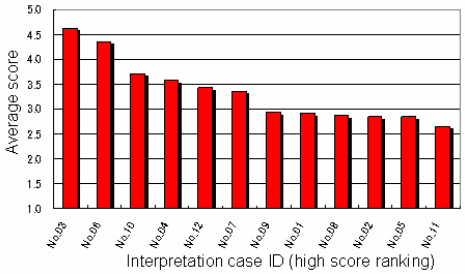


Figure 3-(1) *A-style lectures:* Evaluation result of each interpretation case

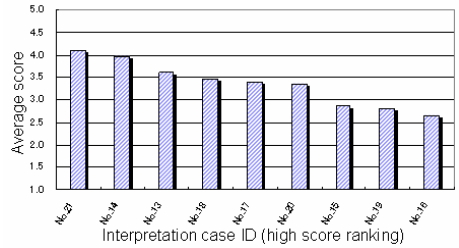


Figure 3-(2) *B-style lectures:* Evaluation result of each interpretation case

4. Analysis of the Experimental Results

4.1. Relationship between the Listeners' Impressions and the Average Pause Time

We have investigated the relationship between how the subjects feel about the interpreter's utterance -- whether it is comfortable to listen to -- and the respective average pause time that appears repetitively in each interpretation case of lecture styles A and B. In Figs. 4-(1) and 4-(2) below, the rank for each translation case is plotted on the horizontal axis and the average pause time for each case is on the vertical axis.

4.1.1. A-style lectures: the cases with the slow speech rate

Figure 4-(1) clarifies that the translation cases in which the average pause time was short acquired a high evaluation from the subjects (the correlation coefficient is shown in the lower right of the graph). The correlation coefficient of the length of the average pause time and the ranking of the evaluation is -0.65, indicating that there is indeed a correlation between those two factors. It is also clear that the cases with a high evaluation, that

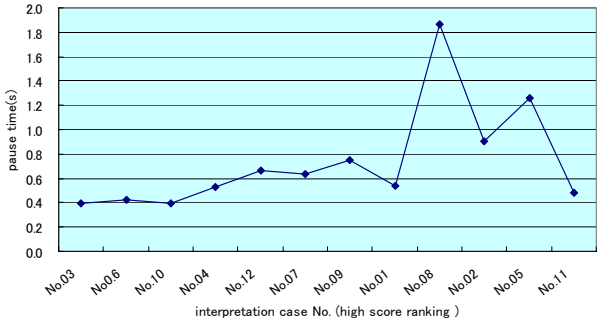
is, interpretations easy to listen to, tended to have a short pause time.

4.1.2. B-style lecture: the cases with fast speech rate

Figure 4-(2) displays the average pause time of each translation in lecture style B, showing that the length of the pauses of each translation tends to be long. The correlation coefficient of the evaluation ranking and the average pause time was -0.30; we could not identify enough evidence for a correlation between those two factors. Thus, we have found that the faster the lecturer speaks, the smaller the influence the speech rate has on the subjects' auditory impression. This indicates how comfortable it was for subjects to listen to the interpretation.

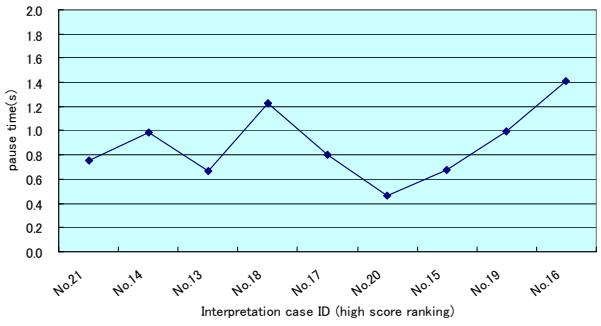
In this analysis, we used Spearman's rank correlation coefficient, which is a special case of the Pearson product-moment coefficient in which the data are converted to ranks before calculating. The following describes the degree of correlation of the correlation coefficient's absolute values.

Less than 0.2:	no correlation
0.2-0.4:	weak correlation
0.4-0.7:	moderate correlation
More than 0.7:	strong correlation



[Correlation coefficient −0.65]

Figure 4-(1) *A-style lectures: Listeners' Impression and the Average Pause Time*



[Correlation coefficient −0.30]

Figure 4-(2) *B-style lectures: Listeners' Impression and the Average Pause Time*

4.2. Aspect of the Interpreters' Utterance

To analyze the phenomena observed in Section 4.1, we selected the two highest- and lowest-ranking translation cases from both lecture styles A and B and observed the respective cycles of the speak-stop state (See Figs. 5 and 6). The horizontal axis represents the 60-sec period of the lecture t sampled between the five- and six-minute marks and the vertical axis indicates the speak-stop state.

4.2.1. The Transition of Interpreters' Utterances: The Case where the Lecturers' Utterance Rate was Slow (A-style lectures)

Regarding the A-style lectures, the two top-rated interpretation cases were characterized by the short and stable length of the pauses between utterance units. This means the cycle of the speak-stop state for the interpreters' utterances is also stable. In the two lowest-ranking translation cases, long pauses appear in the middle of utterances, meaning that there was an uneven distribution of speak- and stop-states.

A-style lectures: the two top-rated interpretation cases

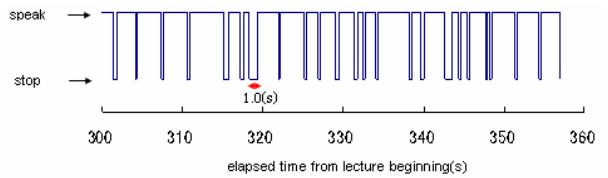


Figure 5-(1) *The Transition of Interpreters' Utterance (Interpretation case No.3)*

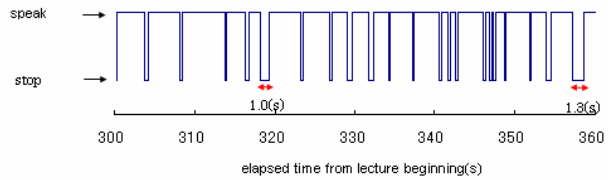


Figure 5-(2) *The Transition of Interpreters' Utterance (Interpretation case No.6)*

A-style lectures: the two lowest-ranking interpreting cases

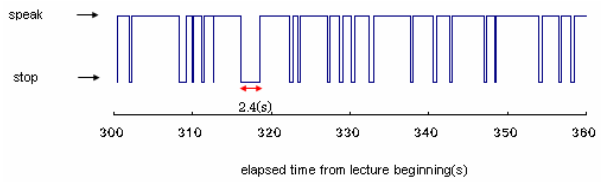


Figure 6-(1) *The Transition of Interpreters' Utterances (Interpretation case No. 11)*

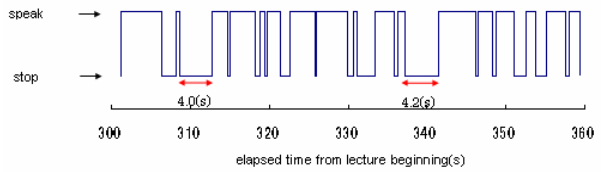


Figure 6-(2) *The Transition of Interpreters' Utterances (Interpretation case No.5)*

4.2.2. . The Transition of Interpreters' Utterances: Cases where the Lecturers' Utterance Rate was Fast (B-style lectures)

In lecture style B, where the time pressure on the interpreters was high, the pause time generally tended to be long. However, when we compare the interpretation cases that received a high evaluation with those receiving a low evaluation, it is clear that there was a difference in the stability of the speak-stop state cycles in the interpreters' utterances. In those two lowest-ranking cases, long 6-sec pauses appeared, and those pauses disturbed the stability of the speak-stop state cycle.

B-style lectures: the two top-rated interpretation cases

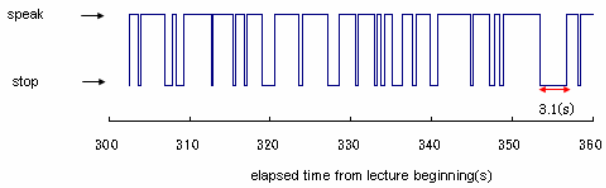


Figure 7-(1) *The Transition of Interpreters' Utterances (Interpretation case No. 21)*

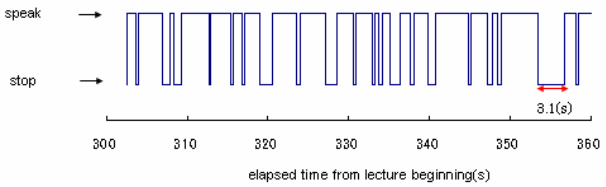


Figure 7-(2) *The Transition of Interpreters' Utterances (Interpretation case No. 14)*

B-style lectures: the two lowest-ranking interpretation cases

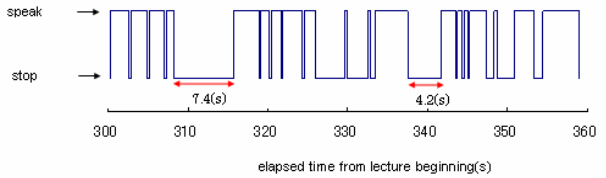


Figure 8-(1) *The Transition of Interpreters' Utterances (Interpretation case No. 16)*

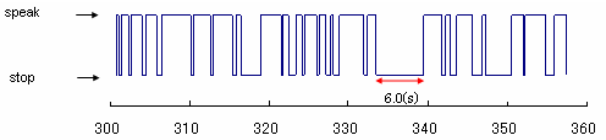


Figure 8-(2) *The Transition of Interpreters' Utterances (Interpretation case No. 19)*

5. Conclusion

In this paper we have examined the relationship between the pauses in simultaneous interpretation and listeners' impressions of them in order to collect data based on which the translation timing of simultaneous machine interpretation system being developed is determined. In the experiment conducted for this research, we used the CIAIR simultaneous interpreting database, which includes two different types of English-Japanese simultaneous translation (lecture style A and B), and had 31 subjects listen to those translations. The conclusions drawn from the results of this experiment are shown below. What we discovered from this experiment is beneficial toward realizing a machine-based simultaneous interpretation system:

a) It was found that, in lecture style A where the speech rate of the lecturers was rather slow, interpretations that received a high evaluation had short pauses between utterance units. On the other hand, in lecture style B, where the speech rate of each lecturer was fast, the influence of pause length on the listeners' impression was small.

b) By closely examining the aspects of simultaneous interpreters' utterances, we found that the translation cases receiving a high evaluation in both lecture styles A and B had the characteristic that the speak-stop state in those translation was stable and rhythmic.

c) What can be derived from (a) and (b) is that pause length is a crucial factor for listeners' impression, that is, how comfortable it is to listen to those interpretations. However, in situations where where the time pressure on the simultaneous interpreters was high, we have found that the influence of pauses on the listeners' impression is small if the interval and the distribution of those pauses are stable.

Future work will involve analyzing fillers (voiced sounds) such as "ahhh" and "ehhh" (≠ pause in this paper) so as to clarify the relationship between fillers and listener's impressions.

6. Acknowledgements

The authors would like to thank Prof. Toyohide Watanabe, Assoc. Prof. Nobuo Kawaguchi, and all members of Matsubara Laboratory at Nagoya University for their valuable advice. This work is partially supported by the Grand-in-Aid for Exploratory Research (No. 17652040) of JSPS.

7. References

[1] Sugito, M., Kunihiro, T., Yokose, H., and Kawano, M. Accent, Intonation, Rhythm and Pause, SANSEIDO, Japan, 1997.

[2] Sugiura, M. Vom Wesen des Rhythmus *(In Japanese)*, MISUZU SHOBO, Japan, 1994

[3] Matsubara, S., Takagi, A., Kawaguchi, N., and Inagaki, Y. "Bilingual Spoken Monologue Corpus for Simultaneous Machine Interpretation Research," *Proceedings of 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Vol. 1, pp. 153-159, 2002.

[4] Tohyama, H. and Matsubara, S. "Construction and Utilization of Bilingual Speech Corpus for Simultaneous Machine Interpretation Research," *Proceedings of **9th European Conference on Speech Communication and Technology** (Eurospeech-2005),* pp.1585-1588, 2005.

[5] Maekawa, K. "Corpus of Spontaneous Japanese: Its design and Evaluation," *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, 2003.