Development of a System to Assist Simultaneous Interpretation and Shadowing

Michiko Watanabe[1], Youichi Tokioka[2], Keikichi Hirose[3]

[1]Graduate School of Frontier Sciences, The University of Tokyo, Japan
[2] Straightword, Inc. Japan
[3]Graduate School of Information Science and Technology, The University of Tokyo, Japan
watanabe@k.u-tokyo.ac.jp, tokioka@straightword.co.jp, hirose@gavo.t.u-tokyo.ac.jp

Abstract

We have been developing a system to assist simultaneous interpretation on site by introducing merits of consecutive interpretation. While recording a speaker's voice, it can replay the voice at various speeds. Interpreters can stop and restart the replay at any given point in time so that they can concentrate on giving interpretation when necessary. This approach will free interpreters from simultaneously engaging in multiple tasks, such as listening to the original speech, comprehending it, translating it into another language and speaking out. We conducted experiments to examine the effects of the system on English-Japanese interpretation using four interpreters. With the system, the accuracy and the fluency of interpretations improved. However, it took interpreters more than twice as long as the interpretations in traditional methods. We report our efforts to reduce the time delay and discuss applying the system to shadowing.

Index Terms: simultaneous interpretation, quality of interpretation, delay, cognitive load, shadowing

1. Introduction

In recent years, the demand and need for interpretation have been increasing in the wake of the advancing globalization of society. In addition to conventional, typical venues like conferences, business negotiations, and broadcasting, the need for interpreting is arising in a diverse array of fields such as medical care, courtrooms, and sports. Although research is being conducted into automatic voice translation, its current performance falls far short of the required levels. Against this background there has come to be a pressing need for the fostering of interpreters.

While a number of classification approaches exist for interpretation depending on differences in perspective, the two methods listed below constitute the major ones and are selectively employed according to the circumstances and need [1].

In talks with "consecutive interpretation", the speaker and the interpreter speak alternately. While the speaker is talking, the interpreter makes notes on what is said. When the speaker comes to a stopping point in their discourse, the interpreter offers a translation of it all at once. Compared to simultaneous interpretation described below, it has the disadvantage of being time consuming. On the other hand, it is easier for interpreters to produce accurate interpretations than simultaneous interpretation because of less time pressure [2].

In "simultaneous interpretation" the interpreter provides a translation nearly concurrently with the original speech. This method is often employed at conferences. The interpreters are usually located in a small room called a booth, where they provide interpretation through microphone while listening to the speaker's voice through headphones.

Between these two methods, simultaneous interpretation has a greater increase in demand because of its time efficiency. Speakers can keep speaking at their pace without waiting for the interpreter's translation. Therefore, there is a pressing need for the fostering of interpreters capable of rendering high quality simultaneous interpretation. However, methods for effectively fostering interpreters do not seem to be established yet. Interpreters seem to be only imitating their superiors' techniques on-site to improve their professional ability [3].

Simultaneous interpretation requires interpreters heavy cognitive loads because they have to perform following five activities in parallel: 1) listening to what the speaker is saying; 2) comprehending the speech; 3) encoding the message into another language; 4) speaking out the interpretation; 5) hearing their own voice [1]. This can lead to a decline in the quality of the translation very easily. To cope with such difficulties, it is common to provide continuous interpretation by alternating several interpreters approximately every 15 minutes, which results in high labour costs for the employer.

Interpretation can be more difficult when the word order of the two languages radically differs. Let us consider interpretation between English and Japanese, for example. English is a SVO language whereas Japanese is SOV. Verbs are located at the end of the sentence in Japanese. Therefore, interpreters who are listening to English have to wait interpreting verbs or verb phrases into Japanese until all the adverbial phrases have been heard. In other words, they have to keep the meanings of verbs in memory while interpreting the other parts until the end of the sentence. On the other hand, when interpreting Japanese into English, interpreters cannot interpret verbs until the Japanese sentence comes to the end, which can cause a serious time loss. To cope with such difficulties, interpreters need to have a lot of expertise. As a consequence, the number of skilled interpreters is very limited. It is estimated that only about 100 Japanese-English interpreters are active on the front lines in Japan [1].

In order to reduce such difficulties in simultaneous interpretation, we have been developing a system which assists simultaneous interpretation rather than relying solely on the interpreters' abilities, as has traditionally been done. This system can play a previous part while recording ongoing speech, which enables interpreters to stop and resume playback at any point without missing the ongoing part of speech. Figure 1 is a schematic diagram of how the system can be used on site. Figure 2 illustrates user interface for the system.
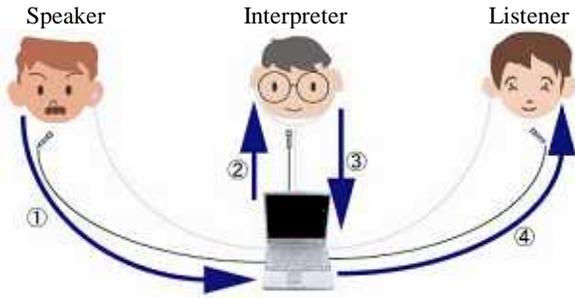
Figure 1: *Schematic diagramme of usage of the system*
Original speech is recorded and stored on PC.
An interpreter listens to the speech and interprets it at his own pace.
The interpretation is transmitted to the listener either directly or through the system.

Figure 2: *User Interface for the system.* The speech waveforms of an original speaker and an interpreter are respectively shown in the upper and lower windows. The part being played is indicated by a green cursor in the upper window. The replay speed can be changed between the rates of 0.6 times and 2.0 times at intervals of 0.2 times. The users can speed up or slow down the replay at any given point in time.

This system can replay the original speech at various speeds while recording it. The interpreter can stop the replay and restart it at any given point in time. With such functions, the system allows interpreters to concentrate on listening or speaking out when necessary. By introducing such a system we aim at improving the quality of interpretation to the level of consecutive interpretation without wasting so much time as consecutive. This system has a merit from the speaker's point of view as well in that they can keep talking at their own pace without waiting for the interpreter to finish their interpretation.

In the remaining of this paper we report on experiments to compare the quality of interpretations with and without the system. We also evaluated interpretations with the system from the point of time efficiency.

2. Method

2.1. Subjects

Four trained interpreters took part in the experiment. They are all female speakers of Japanese in their 30s or 40s. Two of them have high scores (upper range of the 900s) in TOEIC (Test of English for International Communication) and rich experience in interpretation. We call them experts (or "E"). The other two have lower TOEIC scores (upper range of the 800s) and less experience than the experts. We call them intermediate interpreters (or "I").

2.2. Material

We created 13 English passages. Nine of them were news report style passages and the rest were dialogues. Each passage had 20 items to be checked when interpreted into Japanese. The items can be grouped into the following four categories: 1) numbers with units; 2) proper nouns; 3) keywords of the topic; 4) syntax (concordance of subjects and predicates). Each category contained five items per passage. Thus, when an interpreter correctly translated all the items into Japanese, she would obtain 20 points for the passage. The following is an example of a passage. Underlined words are the items to be checked.

> Competitive swimmer Nakane, who raced in the 800-meter women's free-style swimming event at the 2004 Athens Olympic Games, announced she will not participate in the relay at this year's Beijing Olympics, but will focus on individual events. On the 7th of this month, at a competition in Sydney, Australia, she took second place in the 800-meter free-style and set a new Japanese national record of eight minutes and 19.28 seconds, which was 0.23 seconds faster than the previous record.

A female native speaker of American English read out the passages and the reading was recorded. The mean duration of the passages was 30 seconds.

2.3. Procedures

Each of the four interpreters interpreted the 13 passages in three ways: 1) in a conventional way of simultaneous interpretation ("conventional condition"); 2) using the system with the replay speed same as the original speech ("with-system_1.0 condition"); 3) using the system with the replay speed being 1.2 times faster than the original speech ("with-system_1.2 condition"). In the conventional condition interpreters heard the English passages through headsets and simultaneously interpreted them in a booth. In the second condition the interpreters practiced using the system with five practice passages before starting the experiment. They were instructed that they could stop and restart the speech as often as they wanted.

To counterbalance the order effect of conditions, one intermediate (Ia) and one expert (Ea) interpreters first interpreted the passages in the conventional way and did it with the system the second time. The other intermediate (Ib) and expert (Eb) interpreted the passages in the opposite order, i.e. first with the system and second in the conventional way. There was three days' interval between the two trials.

The third trial was conducted after analysing the results of the first two trials to examine the effects of accelerating the replay on the quality and the duration of interpretations. There was about half a year's interval between the first two trials and the third one because of the need of system development. The replay speed can be changed between the rates of 0.6 times and 2.0 times at intervals of 0.2 times. The users can speed up or slow down the replay at any given point in time by pressing a key. However, in order to keep conditions for interpretation constant for the experiment, we prepared the English speech accelerated by 1.2 times. Thus, the English speech to be interpreted has become 17% shorter than the original. The mean duration of original speech is 30 seconds while that of the accelerated speech is 24 seconds. The interpreters listened to the accelerated speech and interpreted it into Japanese. As in the second condition, the interpreters were instructed that they could stop and restart the speech as often as they wanted.

All the interpretations were recorded. The interpretations were scored against the 20 check points per passage (accuracy scores) by one scorer and checked by another. Four native speakers of Japanese listened to all the interpretations and gave impression evaluations in five rating scales. The evaluation criteria for the five scales are given in Table 1.

Duration from the beginning of English speech to the end of the interpretation was measured for each speaker in each condition. The ratio of duration including interpretation to the duration of the original speech was calculated. We call it "time expansion ratio" of interpretation. The ratios were compared between conditions.

Table 1: *Criteria for five scale impression evaluation.*

rating	criteria
5	Sounds like normal Japanese with no sense of incongruity
4	Seems slightly strange, but largely sounds like normal Japanese with no sense of incongruity
3	Seems strange, but there are no places where the meaning cannot be grasped in terms of the Japanese
2	Seems strange and there are several places where the meaning is unclear in terms of the Japanese
1	Seems strange and painful to listen to. There are numerous places where the meaning is unclear in terms of the Japanese

3. Results and Discussion

Figure 3 shows the mean accuracy scores for passages by the four interpreters in three conditions. As mentioned in Section 2.3, "I" stands for intermediate level interpreters and "E" for experts. Ia and Ea first interpreted in a conventional way and second with the system. On the other hand, Ib and Eb interpreted with the system first and in a conventional way second. A perfect score is 20. As observed from the figure, every interpreter obtained higher scores by using the system. A one-way repeated-measures ANOVA shows a significant difference among the three conditions, $F(2,6) = 46.24$, $p < .001$. Paired comparisons adjusted by Bonferroni revealed

that there were significant differences between the original condition and the with-system_1.0 condition, and between the original condition and the with-system_1.2 condition, $t(3) = 7.19$, $p < .05$; $t(3) = 26.14$, $p < .05$, respectively. However, there was no significant difference between the with-system_1.0 condition and the with-system_1.2 condition, $t(3) = .16$, $p = 1.00$. The accuracy scores were higher when the interpreters used the system than when they did not. The results indicate that the system helps interpreters improving accuracy of interpretation. Speeding up the replay by 1.2 times that of the original speech did not seem to deteriorate the quality of interpretation.

Figure 4 displays mean impression scores for the interpretations in three conditions by the four evaluators. A one-way repeated-measures ANOVA shows a significant difference among the three conditions, $F(2,6) = 6.52$, $p < .04$. However, post-hoc tests showed no significant difference between any conditions. The evaluation tended to be higher for the performance with the system than without, suggesting that the system contributes to improving fluency, as well as accuracy, of interpretations.

Figure 5 indicates the mean time expansion ratios of the interpretations in three conditions. A one-way repeated-measures ANOVA shows a significant difference among the three conditions, $F(2,6) = 23.60$, $p < .001$. Paired comparisons adjusted by Bonferroni revealed that there was a significant difference between the conventional condition and the with-system_1.0 condition, $t(3) = 8.42$, $p < .05$. There was no significant difference between the other conditions. When the interpreters stopped and replayed the speech at the same speed as the original using the system, it took them more than twice as long as the conventional method to complete the interpretation. The time expansion ratio for the interpretation is roughly as large as the ratio for consecutive interpretation [2]. However, when the replay was speeded up by 1.2 times, the expansion ratio became smaller particularly with the expert interpreters. We infer that accelerating the replay can be one of the effective ways to reduce the delay of interpretation, though the proposed method still needs improvements in terms of time efficiency.

As a future work we plan to examine the effects of further speeding up the replay on translation. We had five interpreters listen to replays of the English speech at various speeds between the rates of 1.2 times and 2.0 times at intervals of 0.2 times and asked which rate they found optimal for interpretation. All the interpreters answered that the rate of 1.4 times was the best. Based on this survey, we have started training interpreters to interpret speech accelerated by 1.4 times.

Some users of the system report that pressing a key to stop and restart replay can be another cognitive burden in interpretation. We expect that the time can be shortened by further training interpreters to become accustomed to using the system.

The optimal accuracy, fluency and time efficiency differ depending on the purpose of interpretations. We believe that interpretation using this system can be alternative to simultaneous or consecutive interpretation.

4. Applying the System to Shadowing Practice

Shadowing is originally a way of training simultaneous interpreters: Trainees are required to simultaneously listen to and repeat incoming speech as quickly and accurately as
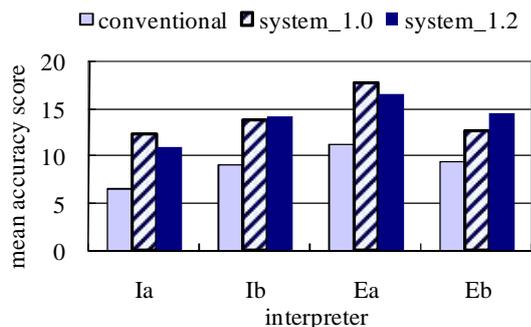
Figure 3: *Mean accuracy scores obtained by the four interpreters in three conditions* (I: intermediate; E: expert, a perfect score is 20)
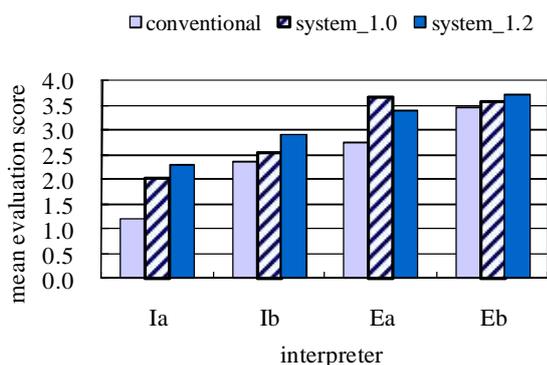

Figure 4: *Mean impression scores by the four evaluators for the four interpreters (5 scale evaluation)*
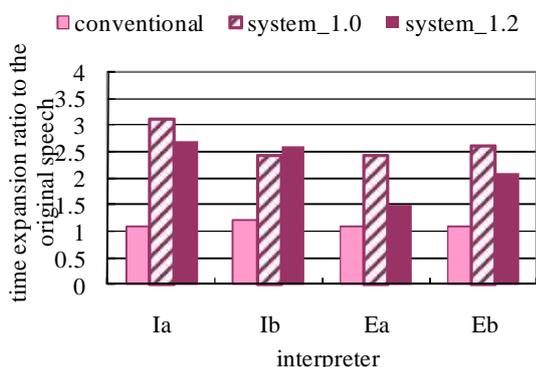

Figure 5: *The mean time expansion ratios of "conventional" and "with system" interpretations by the four interpreters*

possible while listening [4]. Shadowing and simultaneous interpretation have common features in that trainees have to be engaged in multiple tasks, namely listening and speaking, at the same time. Recently shadowing has been attracting attention as a method to improve listening and speaking abilities of second language learners [5]. While some positive effects of shadowing practice on second language learners have been reported, lack of support-systems for self-study has also been pointed out [6]. Our system, which can record and replay speech at various speeds in parallel, can be

helpful not only for interpreters but also for second language learners practising shadowing. With the system, learners can record and play speech that they want to shadow, record their own voice, and listen to it later to compare with the original speech. Learners can adjust replay speeds of the original speech according to their needs. The system is now being used to develop shadowing practice programmes for second language learners.

5. Conclusions

We have been developing a new system to assist with simultaneous interpretation. The system can record speech and play a previously recorded part at various speeds in parallel. Interpreters can stop playing the recorded speech and start it again at their own pace while the speaker continues talking. We expected that this system would reduce the difficulties and improve the quality of interpretations because it allows interpreters time to concentrate on a single task such as listening to original speech or producing interpretation when necessary. Both accuracy and fluency of interpretations improved with the system regardless of the interpreters' proficiency levels. However, the time needed to complete interpretation was longer than the time for conventional simultaneous interpretation. We plan to improve the time efficiency of interpretation with the system by speeding up the playback as well as by training interpreters to gain expertise in using the system. We expect that such a system enables less experienced interpreters to perform simultaneous interpretations of reasonably high quality and makes simultaneous interpretations more accessible in the international communities. This system can be used not only by interpreters but also by second language learners practising shadowing.

6. References

[1] Japan Institute for Labour Policy and Training, "Occupational Database," http://cmx.vrsys.net/JS/CCS_js_01.php?sId=g& accessed on 1 May 2010 (in Japanese).

[2] Ohara, M., Matsubara, S., Ryu, K., Kawaguchi, N., Inagaki, Y., "Temporal Features of Cross-Lingual Communication Mediated by Simultaneous Interpreting: An Analysis of Parallel Translation Corpus in Comparison to Consecutive Interpreting." *Interpretation Studies*, No.3, 35-53, 2003 (in Japanese).

[3] Torikai, K. "Consideration on contemporary interpretation", Journal of Interpretation and Translation, vol. 7, 2002, http://wwwsoc.nii.ac.jp/jais/html/tu_hon_journal/html/15_torikai2.html, accessed on 1 Mar 2010 (in Japanese).

[4] Mochizuki, M., "On the application of shadowing method to Japanese language education: with special reference to relevance of learners' Japanese competence to their self-assessment of the effects of shadowing," *Audio-Visual Education*, vol.29, 37-53, 2006 (in Japanese).

[5] Tamai, K., "On the effects of shadowing on listening comprehension -- keynote lecture at the 3rd Annual Conference of JAIS," *Interpretation Studies*, No.2, 178-192, 2002 (in Japanese).

[6] Tanaka, M., "Current pedagogical issues in teaching interpreting at the undergraduate level," *Interpretation Studies*, No. 4, 63-82, 2004 (in Japanese).