CIAIR Simultaneous Interpretation Corpus

Hitomi Tohyama
Koichiro Ryu
Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan
E-mail: hitomi @el.itc.nagoya-u.ac.jp

Shigeki Matsubara
Nobuo Kawaguchi
Information Technology Center,
Nagoya University,
Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan

Yasuyoshi Inagaki
Faculty of Information Science and Technology, Aichi Prefectural University,
Nagakute-cho, Aichi-gun, Aichi-ken, 480-1198, Japan

Abstract

This paper describes the design, analysis and utilization of a simultaneous interpretation corpus. The corpus has been constructed at the Center for Integrated Acoustic Information Research (CIAIR) of Nagoya University in order to promote the realization of the multi-lingual communication supporting environment. The discourse tag and the utterance time tag were given to the corpus. Therefore, the corpus is expected to be useful not only for the development of simultaneous interpreting systems but also for the construction of an interpreting theory.

1. Introduction

Recently, spoken language corpora have been constructed for the purpose of studying on speech processing in many research organizations (for example [1, 2]). The large-scale corpora are recognized to be important widely, and used in various research areas, such as speech recognition, natural language processing, linguistics, language education, and dictionary compilation.

At the Center for Integrated Acoustic Information Research of Nagoya University (following, CIAIR), a corpus of simultaneous interpretation between Japanese and English has been constructed over five years (from the 1999 fiscal year to the 2003 fiscal year). We aim at the realization of the multi-lingual communication supporting environment. The recording time is 182 hours in total. The speech data has been all transcribed and visualized. Furthermore, we have completed language analysis of the corpus. The size of transcribed data is about 1 million words, and the corpus would deserve to be called the simultaneous interpretation corpus of the largest-in-the-world class. Additionally, we have developed some software tools for corpus analysis in order to support the practical use of the corpus. They have been developed as software which can be performed on the Web server, and a user can refer to the corpus easily by using a browser.

This paper describes the design, collection, construction, analysis, and utilization of the simultaneous interpretation corpus. We discuss the application to other research fields beyond the area of computer science. These fields, for example, include cognitive science, linguistics and education, etc.

In the following section, we describe the purpose and the design of the simultaneous interpretation corpus. In Section 3, we describe the recording of the corpus. In Section 4, the construction of the corpus is explained in full detail. Section 5 discusses the use of the corpus.

2. Corpus Design

2.1. Aim of Data Collection

Machine interpretation has become one of the most important research topics with the advance of technologies for speech processing and language translation. Several experimental systems of spoken dialogue translation for specific task domains have been developed [3,4,5]. The interpreting style of them is within so-called consecutive interpretation. In order to provide an environment to support natural and smooth cross-lingual communication, to develop a technique for simultaneous machine interpretation has been awaited and tried out recently. Not only a generation of translation but an outputting timing of translation is required for a simultaneous interpreting system. It would be effective to investigate and analyze the interpreting process of professional simultaneous interpreters. The CIAIR is constructing and maintaining various types of speech and language database for the purpose of the advancement of robust speech information processing technology [6]. Moreover, a bilingual database of simultaneous interpretation has also been constructed as a part of this project. We aim to develop speech translation technologies and to construct interpreting theories.

2.2. Policy of Corpus Design

The large-scale corpus needs to be equipped with flexibility, because many researchers are expecting to utilize the corpus for their purposes. Therefore, we collected both monologue and dialogue data. The contents of the database are daily topics. The database targets English and Japanese. Table 1 shows the outline of the recording of the corpus.

Table 1. Outline of simultaneous interpretation corpus

Item	Contents
Speech type	monologue dialogue
Language	English Japanese
Interpretation style	simultaneous
Media	speech text

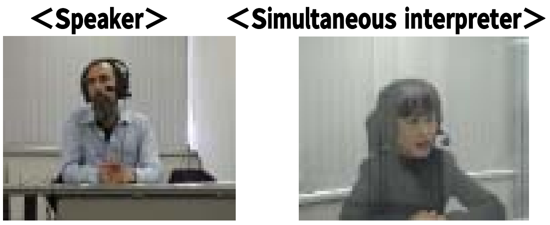


Figure 1. Recording environment

3. Recording of Speech Data

3.1. Recording Environment

One of the purposes of CIAIR is to collect a large quantity of speech data which were generated under natural circumstances, so the recording took place in a classroom. Facial expressions and conversational behaviors of speaker's are also important information for simultaneous interpretation, thus, the interpreters were in booths from which they could see the speakers (Fig.1). Both speakers and interpreters used the same cross-talking microphones. The speeches were digitized by sampling frequencies of 16 kHz and 16 bits, and recorded onto digital audio tapes (DAT) in multiple channel environments. All interpreters are professional interpreters who are active in the front lines.

3.2. Recording of Monologue Speech Data

Simultaneous interpreters go into a booth, and interpret the lecturer's speeches from the headphone. Although the lecturers face to the audience, they cannot hear the interpreter' speech. It enables them to speak at their own paces. The contents of speeches are economics, history, and, culture, etc. Moreover, each monologue speech is interpreted by two or four professional interpreters. Their degree of experience differs from one another (5 years over or not). The flexibility of a database is raised by using four interpreters. Therefore, it becomes possible to compare two or more interpretation examples in a sample of utterance. Moreover, we can compare interpreters' utterance speed, speaking timing, and strategy of interpretation. The speech was recorded for about 10 minutes per lecture.

3.3. Recording of Dialogue Speech Data

Travel conversation was selected as a domain of conversation, which includes popular topics during overseas travel at airports and hotels, and simulated conversations are recorded. To put it concretely, the following topics were selected: "airport check in", "hotel check-in/check-out", "booking of a room at a hotel", and "booking of a seat in an airplane", and so on.

```
0001 - 00:02:360-00:04:559 N:
Northwest Airlines, may I help you?<SB>
0002 - 00:14:600-00:15:399 N:
I see<SB>
0003 - 00:15:752-00:18:704 N:
Could you please tell me the name of the flight
0004 - 00:19:016-00:19:832 N:
and date?<SB>
0005 - 00:33:344-00:37:024 N:
December forteenth, Northwest (A three o two;302)<SB> Is that correct?<SB>
0006 - 00:40:520-00:42:456 N:
And may I have your name, please?<SB>
0007 - 00:48:296-00:50:624 N:
(R Osada Megumi)<SB> Is that correct?<SB>
0008 - 00:52:952-00:55:072 N:
And could you spell that please?<SB>
```

Figure 2. Sample of text data
Dialogue - English speaker's utterance

```
0001 - 00:03:464-00:05:016 I:
(F あ)ノースウエスト航空でございます<SB>
(F ア)ノースウエストコークーデゴザイマス<SB>
0002 - 00:15:400-00:15:680 I:
はい<SB>
ハイ<SB>
0003 - 00:15:768-00:16:607 I:
かしこまりました<SB>
カシコマリマシタ<SB>
0004 - 00:18:856-00:21:568 I:
(F あの)便名と(F あの)日付お願い致します<SB>
(F アノ)ビンメート(F アノ)ヒズケオネガイイタシマス<SB>
0005 - 00:34:392-00:36:032 I:
(F えー)十二月十四日
(F エー)ジューニガツジューヨッカ
0006 - 00:36:304-00:38:888 I:
(F えー)ノースウエストの(A 三百二; ３０２)便でございますね<SB>
(F エー)ノースウエストノサンビャクニビンデゴザイマスネ<SB>
0007 - 00:42:240-00:43:376 I:
お名前お願い致します<SB>
オナマエオネガイイタシマス<SB>
0008 - 00:49:104-00:51:464 I:
(R 長田恵)様でよろしいですか<SB>
(R オサダメグミ)サマデヨロシーデスカ<SB>
0009 - 00:54:504-00:55:840 I:
(F えーと)スペルお願い致します<SB>
(F エート)スペルオネガイイタシマス<SB>
```

Figure 3. Sample of text data
Dialogue - English-Japanese interpreter's utterance

In order to enhance the quality of interpretations for both English speakers and Japanese speakers, each speaker was accompanied by one interpreter. To ensure all the participants' speech pretension, speakers can listen only to the output from the other speaker's interpreter, and the interpreters can listen only to the speech that they are assigned to interpret.

Please note that these dialogues are simulative, in which the contents of the speeches can be limited. In attempting to collect utterances as unfettered as possible, such background information as the speaker's roles and conversational tasks were informed to speakers in advance. For a speaker who is a customer of a hotel, for example, the kind and number of rooms that should be reserved, and for a speaker as front desk clerk, rooms that can be reserved, etc. We set up "airport" and "hotel" as the typical situations of dialogue communications doing overseas travel. The recoding time per one dialogue was from 1 minute to 16minites, and dialogues of various types were collected.

Table 2. The kind of main discourse tags

Tag	Usage
Type I: Tags that refer to the characteristics of linguistic message	
(F)	Filled pauses
(D)	Word fragment, repairs
(W)	Reduced or incorrect pronunciation
(O)	Foreign language, archaic Japanese etc.
(A)	Use of alphabet in orthographic transcription
Type II: Tags that refer to the existence of phonetic / non-verbal events	
<H>	Non-lexical lengthening of vowels
<Q>	Non-lexical lengthening of consonants
<FV>	Vowel whose phonemic status is not identifiable
<SB>	Sentence breaks

4. Corpus Construction

4.1. Transcription of Speech Data

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 [7]. Figure 2 and 3 show the sample data of English speaker's utterances and that of English-Japanese interpreter's utterances, respectively.

All the speech data (182 hours) were transcribed into a text. The standardization is shown as follows:

- **Utterance unit**

  Utterance units were set by 200ms-or-longer pauses in the speech of speakers and interpreters.

- **Notation**

  Recorded Japanese speech is transcribed in two different ways: orthographic and phonetic transcriptions.

- **Tag annotation**

  - **Utterance ID**

    A serial number was given to each utterance unit.

  - **Time tag**

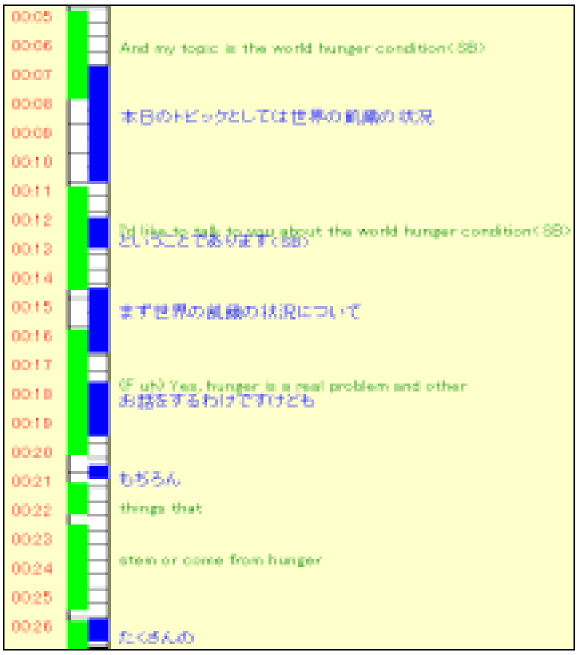
    The beginning time and end time of the utterance units were tagged.

  - **Discourse tag**

    Language tags were also added onto fillers, hesitations, and corrections. Table2 shows examples of the tags used in the transcription text.

4.2. Visualization of Speech Data

We developed a timing information visualization tool. The speaking time of English lecturers, English-Japanese interpreters, Japanese lecturers and Japanese-English interpreters and their overlapping relations are displayed as Fig4 shows. Thereby, we can visually observe a overlap of a lecturer and a translator utterance.



And my topic is the world hunger condition<SB>

本日のトピックとしては世界の飢餓の状況

I'd like to talk to you about the world hunger condition<SB>

ということであります<SB>

まず世界の飢餓の状況について

(F uh) Yes, hunger is a real problem and other

お話をするわけですけども

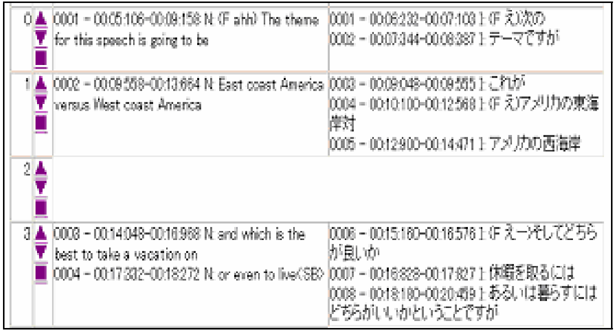
もちろん

things that

stem or come from hunger

たくさんの

Figure 4. Sample of time chart



	<Lecture's utterance>	< Interpreter's utterance>
0	0001 - 00:05:106-00:09:158 N (F ahh) The theme for this speech is going to be	0001 - 00:06:232-00:07:103 J (F え)次の 0002 - 00:07:344-00:08:387 J テーマですが
1	0002 - 00:09:569-00:13:664 N East coast America versus West coast America	0003 - 00:09:049-00:09:565 J これが 0004 - 00:10:100-00:12:568 J (F え)アメリカの東海岸対 0005 - 00:12:900-00:14:471 J アメリカの西海岸
2		
3	0003 - 00:14:048-00:16:968 N and which is the best to take a vacation on 0004 - 00:17:302-00:18:272 N or even to live<SB>	0006 - 00:15:160-00:16:576 J (F えー)そしてどちらが良いか 0007 - 00:16:828-00:17:827 J 休暇を取るには 0008 - 00:18:180-00:20:459 J あるいは暮らすにはどちらがいいかということですが

Figure 5. Alignment support tool

4.3. Construction of the Parallel Corpus

For a detailed analysis of interpreters' speech, such as extraction of temporal characteristics of interpretation, the acquisition of translation patterns, the detection of translation units and so forth, it is necessary to align utterances of speakers and interpreters with relatively small units[7]. An alignment support tool (Fig. 5) which works on the internet has been developed by using CGI script. The users can align the utterance units by the clicking the mouse on the bilingual text displays. The aligned data can be used for analysis of interpreting units and timing. We have aligned the corpus using the tool according to the following conditions:

Table 3. Statistics of speech database

Items		No. of words/morphemes	No. of utterance	Recording time[min]
Speaker	English	90249	8422	695
	Japanese	84278	6529	597
	Total	174527	14951	1292
Interpreter	E-J	266050	25507	1639
	J-E	127991	16083	1265
	Total	394041	41590	2904
Sum Total		568568	56541	4196

Table 4. Statistics of dialogue database

Items		No. of words/morphemes	No. of utterance	Recording time[min]
Speaker	English	107850	14223	1678
	Japanese	106258	16485	1678
	Total	214108	30708	3356
Interpreter	E-J	116776	15286	1678
	J-E	91743	13719	1678
	Total	208519	29005	3356
Sum Total		422627	59713	6712

＜Main rules＞

- The unit should be the minimum alignment unit
- Utterances should be aligned as small as possible
- Utterance units such as fillers or non-language phenomena and the utterances with no appropriate counterparts can have no correspondence.

As stated above, a detailed alignment standard was established, so, the annotation work was made uniform.

4.4. Providing Environment Information

A large-scale corpus has various availabilities. Information that doesn't appear at speech data and text data might be important. Therefore, we provided the Environment information of recording data file for every session. Environment Information are date, location, recording time, audio-video equipment, topic, type of speech, the speaker's roles and conversational tasks, the information on speaker, the information on interpreter (years of experience, etc).

4.5. Statistics on the Corpus

The large-scale corpus involving 1 million words have been developed; we have finished recording the data of 182 hours of speech, transcribing it into text, attaching discourse tags, and matching source utterances to their target utterances, so far. Statistics on lecture data and the dialogue data are shown in table 3 and 4, respectively.

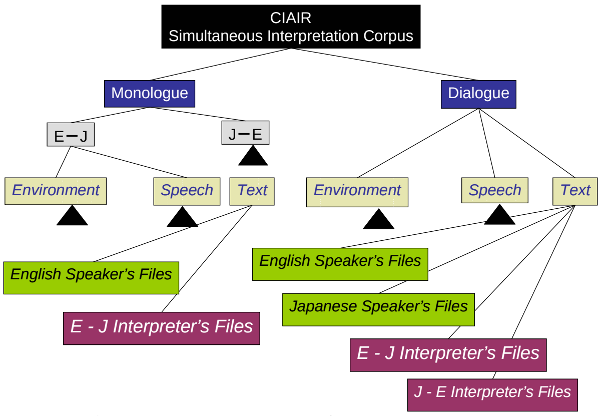


Figure 7. Structure of the CIAIR Corpus

4.6. Structure of the CIAIR Corpus

This corpus consists of three parts, the speech data, the text data, and environment data of recording. (Fig. 7).

- **Transcription of speech data**

  A speaker and an interpreters' speech are stored as a multiplex wave file.

  - **Speech database**

    This database consists of one multiplex voice file of English and Japanese per session.

  - **Dialogue database**

    The speaker data and his/her interpreter data are stored together in one file. Therefore, two stereo files were made for each dialogue.

  - **Transcript data**

    There are four type of data file: English speaker's transcript file, E-J interpreter's transcript file, Japanese speaker's transcript file, J-E interpreter's transcript file.

  - **Environment data**

    The recording environment information files were made for each of lecture or dialogue (the foregoing Fig. 6).

5. Utilizing the CIAIR Corpus

To develop the simultaneous machine interpretation, it is necessary to analyze the interpreting process of professional simultaneous interpreters. We have researched simultaneous interpreters' utterance timing, interpretation unit, and generation of translation and compared the simultaneous interpretation with the consecutive one in cross-lingual communication. We have proved the effectiveness of simultaneous interpretation technology [12]. Furthermore, if the data for analysis is large-scale, it is also possible to verify a qualitative analysis result still more a quantitative one.

This section describes the main research results performed by using the corpus. In the end of this section, the application possibility of this corpus in various fields is described.

5.1. Construction of Interpreting Theory by Corpus Analysis

5.1.1. Analysis of Interpreter's Speaking Timing

Simultaneous interpretation may overlap with the corresponding native speech. It is expected that an interpreter recognizes a part of a lecturer's utterance as an interpreting unit and interprets it at an early stage [8]. We have investigated the interpreting units and speaking timing of professional interpreter by analyzing the aligned corpus [9]. The summary of results is shown below:

- Since a subject appears at the beginning of a sentence in both Japanese and English, the subject can be interpreted immediately.
- By controlling the out putting speed of system based on the quantity of the input utterance. It is possible to reduce the difference between the beginning time of the interpreter's utterance and that of the lecturer's utterance.

5.1.2. Temporal Features Analysis of Simultaneous Interpreting

Interpretation has two styles: consecutive interpretation and simultaneous interpretation. One of major differences between them is whether an interpreter starts to speak after the speaker completed his/her utterances, which is consective interpretation, or before, which is simulteneous interpretaion. Simulteneous interpretaion occurs that the speaker's utterance and interpreter's utterance temporally overlap each other; however in the consective interpretaion, those utterances doesn't. We have done the further research on these two styles of interpretation [10, 11].

We have compared simultaneous interpretation with consecutive one in cross-lingual communication and we have proved the possibility that simultaneous interpretation technology performs more effectively. Ohara's study [12] proved how effective the simultaneous interpretation is by analyzing the actual simultaneous interpretation data. Ohara focused the on the efficiency and the smoothness of cross-language conversation through simultaneous interpretation. The summary of results is shown below:

- The growth of average dialogue time growth rate on simultaneous interpretation was twice as much as consecutive interpretation, which indicates that conversational efficiency through simultaneous interpretation has been raised considerably in comparison with consecutive interpretation. This tendency can be seen in English-Japanese interpretation.
- The average of speakers' waiting time on conversations through interpreting is 4.4 seconds for English speakers on simultaneous interpretation, and 15.4 seconds on consecutive interpretation. That is 4.3 seconds for Japanese speakers on simultaneous interpretation, and 14.5 on consecutive interpretation, which indicates that smoothness increases drastically.

The result proved the usability of simultaneous interpreting technology as a support environment for cross-language dialogues because in the dialogues through simultaneous interpretation.

5.1.3. Collecting the Strategies for Simultaneous Interpretation

Simultaneous interpretation is advanced language processing activities of human. Simultaneous interpreters have to generate their translations simultaneously with original speech. However, they have the restrictions on speaking timing (when-to-say) and how to translate speaker's utterance (how-to-say). However, they have various kinds of strategies to raise simultaneity. In this investigation, the interpreting patterns used frequently and having both/either high flexibility and simultaneity were extracted from a bilingual spoken monologue corpus [13]. The CIAIR corpus has as many of four interpreter data per one monologue section. Therefore, it is possible to collect two or more interpretation patterns from one speaker's utterance. Those extracted interpretation patterns can be used as a rule of the interpretation system. It is possible to develop the system by using the rule.

5.2. Application to Other Fields

Today, in various research fields, a simultaneous interpretation has been studied. Most of those researches are qualitative researches that elaborately analyze few examples of a simultaneous interpretation. The examples of those researches are shown below:

- **Cases of research area**
  - ➢ **Cognitive Science and Cognitive Linguistics**

    A simultaneous interpretation is the advanced language processing that converts one language into the other language while maintaining the caught utterance. In addition, the converted content is passed on to the listener. In the study field of cognitive science and cognitive linguistics, the mechanism of simultaneous interpretation is researched in order to investigate how human *working memory* works [14].
  - ➢ **Linguistics**

    The effect which interpreter training method introduces into foreign language study is verified [15,16]. For instance, there are *shadowing*, *sight translation*, and *slash reading*, etc. The simultaneous interpreter generates a translation according to the word order of the speaker’s utterance. It is similar to the process that human understands his/her mother tongue. Therefore, it is said that such the method for training interpreters has effectiveness in the second language acquisition.

- **Cases of education**

  Recently, some universities has been opening the curriculum for training a translator in order to raise talented people with special skill. For example, there are courses of interpreter theory and an interpreter technical theory [17].

The CIAIR corpus will make it possible to analyze database quantitatively. So, those researchers in various fields can use this corpus for their researches in various ways. For example, they can collect a lot of samples data from this corpus.

6. Conclusion

This paper has described the design, analysis and utilization of the CIAIR simultaneous interpretation corpus of Nagoya University. We expected that the corpus will be used for not only the development of simultaneous interpreting systems but also the construction of an interpreting theory.

We are going to develop the CIAIR corpus further by giving it more detail tags and constructing alignment data. In those days, the spoken language technology has progressed. Therefore, the demand for large-scale spoken database is rising in not only speech processing but also cognitive science, phonology, and linguistics.

We hope that CIAIR corpus will be used in the various research fields. It is preferable to exchange the opinion between different areas and to progress overall.

The CIAIR corpus has been already exhibited. For more details, please refer to the following URL:

http://www.el.itc.nagoya-u.ac.jp/sidb/

Acknowledgment

The authors would like to thank Prof. Dr. Toyohide Watanabe and all members of SLP Group for their precious advices. This work is partially supported by the Grand-in-Aid for Young Scientists (B) (No. 15700127) of the Ministry of Education, Science, Sports and Culture, Japan and The Nakajima Foundation.

References

[1] K. Maekawa, H. Koiso, S. Furui, and H. Isahara, “Spontaneous Speech Corpus of Japanese”, *Proc. 2nd LREC*, 2000, pp.947-952.

[2] T. Takezawa, A. Nakamura and E. Sumita, “Database for Conversational Speech Translation Research at ATR” *the Phonetic Society of Japan, ver.4.* 2000, pp.16-23.

[3] H. Mima, H. Iida, and O. Furuse, “Simultaneous Interpretation Utilizing Example-based Incremental Transfer”, *Proc. 17th COLING and 36thACL*, 1998, pp.855-861.

[4] T. Watanabe, A. Okumura, S. Sakai, K. Yamabana, S. Doi, and K. Hanazawa, “An Automatic Interpretation System for Travel Conversation”, *Proc. 6th ICSLP, Vol. IV* , 2000, pp.44-48.

[5] J. Amtrup, “Incremental Speech Translation”, *LNAI, Vol. 1735*, 1999.

[6] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, “Multi-Dimensional Data Acquisition for Integrated Acoustic Information Research”, *Proc. 3rd LREC*, 2003, pp. 2043-2046.

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

[8] C. Funayama, “The Processing Unit in Simultaneous Interpretation”, *Journal of the Interpreting Research Association of Japan*, Vol.6, No.1, 1996, pp.4-13. (in Japanese)

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

[10] R. Shinzaki, “Information Processing in Simultaneous Interpretation and Consecutive Interpretation”, *Journal of the Interpreting Research Association of Japan,* Vol.14, No.2, 1994, pp.40-46. (in Japanese)

[11] D. Gile, “Consecutive vs. Simultaneous: which is more accurate?”, *The Journal of the Japan Association for Interpretation Studies, No.1*, 2001, pp.8-20.

[12] M. Ohara, S. Matsubara, K. Ryu, N. Kawaguchi, and Yasuyoshi Inagaki, “Temporal Features of Cross-Lingual Communication Mediated by Simultaneous Interpreting: An Analysis of Parallel Translation Corpus in comparison to Consecutive Interpreting” *The Journal of the Japan Association for Interpretation Studies,* No.3, 2003, pp.35-52. (in Japanese)

[13] H. Tohyama and S. Matsubara, “Corpus-based Analysis of Simultaneous Interpreting Patterns”, *IEICE Technical Report, Vol.10*3, *No.487*, 2003, pp.13-18. (in Japanese)

[14] M. Osaka, *Working memory : the sketchpad in the brain* Shinyousya , 2003. (in Japanese).

[15] S. Nagata, “Application of Interpreter Training Techniques to General Language Enhancement: with Emphasis on Listening and Speaking”, *Journal of the Interpreting Research Association of Japan,* Vol.6 No.2, 1996, pp.45-54.

[16] K. Tamai, “Strategic Effect of Shadowing on Listening Ability”, Proceedings of the FLEAT IV Conference in CD, 2001, pp. 620-625.

[17] K. Torikai, “Possibility of Interpreter Training in Japan: Forom the Perspective of English Education”, *Journal of the Interpreting Research Association of Japan,* Vol.7 No.1, 1997, pp.39-52.