

Effect of Captioning Lecture Videos For Learning in Foreign Language

VERI FERDIANSYAH^{†1} SEIICHI NAKAGAWA^{†1}

Toyohashi University of Technology, Tenpaku-cho, Toyohashi 441-8580 Japan

E-mail: { veri, nakagawa }@slp.cs.tut.ac.jp

Abstract: This paper describes a study on the effect of using captions in lecture videos taught in English, to assist non-native English speaker in learning. Lecture videos used in this paper was taken from MIT OpenCourseWare. We divided captions into 5 different types: (1) full text captions, (2) important sentence captions, (3) important phrase captions, (4) important keyword captions, and (5) no captions. Each caption was created manually and embedded in the videos. Participants for the experiment were chosen randomly from non-English native speaking country and each participant must follow a set of tests. Experiment results showed that captions (1) or (2) helped participant the most in understanding the content of lectures.

Keywords: lecture speech, full text captions, important sentence captions, important phrase captions, important keyword captions, no captions

外国語（英語）講義映像に対する字幕提示の理解度効果

フェルディアンシャー ヴェリ^{†1} 中川 聖一^{†1}

豊橋技術科学大学 〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1

E-mail: { veri, nakagawa }@slp.cs.tut.ac.jp

あらまし 本稿では、英語による講義の、英語を母国語としない学習者のための字幕提示法の比較結果について述べる。比較した字幕は、(1)英語の全文、(2)重要文だけ、(3)重要句だけ、(4)重要語だけ、(5)字幕なし、の5種類である。各字幕は、与えられている講義のスク립トから、手作業で講義映像に貼り付けた。日本滞在の留学生と日本人学生を被験者として比較した結果、全文字幕と重要文の字幕が、同程度に、講義の理解に役立つことが分かった。

キーワード： 講義音声、字幕、全文字幕、重要文字幕、重要句字幕、重要語字幕

1. Introduction

The widespread of Internet and the openness of education has made learning more accessible to everybody. Massachusetts Institute of Technology (MIT), with its MIT OpenCourseWare site, is the pioneer of the open courseware movement. Many other universities have followed the open courseware movement and opened their educational materials (e.g., course textbooks, presentations, videos, assessments, simulations, etc.) that are either (a) licensed under an open copyright license (e.g., Creative Commons) or (b) in the public domain so everybody can gain access to the resources at no costs. However, many of those educational materials, particularly lecture videos, are available without any captions or separate transcriptions. Therefore, for example, lectures in English, for students in non-English speaking countries such as Japan, will be very hard to understand the content of the lectures without getting any help from either captions or transcriptions.

Automatic transcribing of lecture videos has got a lot of

attention from researchers in recent years due to the technological advances in Automatic Speech Recognition (ASR) and its associated technologies [1]. Most of recent researches are focusing towards broadcast news. In Japan, NHK (Nippon Hoso Kyokai, or Japan Broadcasting Corporation) has developed and used an automatic broadcast news captioning and have achieved accuracy greater than 95% [2]. Similar system, developed in Spain, achieved accuracy greater than 91% and has been used on the Aragonese Public Television from June 2008 without human assistance [3]. Compared with broadcast news, lectures have more variations in terms of speaking style, environmental conditions, fluency, and less constrained in syntax [4]. Recent research in [5] showed an error rate of 32.4% on the Translation English Database (TED) task. Beyond broadcast news and lecture speech, closed captioning has also been used in telemedicine or telehealth (videoconferencing for health care) to provide better access to telehealth system for people who are deaf or hard of hearing [6].

The purpose of this study is to see whether captioning lecture videos will benefit students for learning in foreign language while investigating which type of caption will give more benefit

^{†1} 豊橋技術科学大学
Toyohashi University of Technology

to the learners. We divided captions into five different types and each caption was created manually by the authors and embedded in the videos using video editing software. We randomly chose participant for the experiment which is students from non-English speaking country and each participant must follow a set of tests. During the tests, each participant will be shown 5 different videos with 5 different captions. This study showed that full text captions and important sentence captions helped participant the most in understanding the content of lectures.

This paper is organized into the following five sections. In Section 2, previous works related to this study is explained. Methodologies are described in Section 3. Section 4 presented the results of our experiments and in Section 5 a conclusion is made and we explained our future work.

2. Related Works

Similar studies have been done by other researchers prior to this work. Research in [7] investigated the use of captioned video materials in advanced foreign language learning. In his experiment, the author only used full text captions and no captions. He measured students' comprehension of the video by creating a 10 item multiple choice comprehension tests. 140 students conducted the experiment and divided into two groups, one group viewed the captioned video and the other viewed caption less video. Comparison of the comprehension test scores of both groups revealed that the group who viewed the videos with captions achieved higher score than those who viewed the videos with no captions.

In [8], the authors attempted to investigate the effects of captioning during video-based listening activities for second language learners. Each learner watched three short videos with and without captions in randomized order. After the second showing of the video, learners took comprehension and vocabulary tests, based on the video. The results of the research showed that captioning was more effective than no captioning. Further interviews with the learners revealed that learners used captions to increase their attention and reinforce their previous knowledge regarding the topic.

While in [9], the author tested a keyword captioning method based on a hypothesis that keyword captions, while presenting learners with lesser material to read, didn't weaken their comprehension in the spoken message. For the experiment, the author used three different amount of text on the video: (1) full text, (2) keyword, and (3) no text. The results of the experiment showed that keyword captions are an effective method for transmitting content in video materials. Keyword captions require less reading for learners and the appearance of keywords in captions triggers particular attention to specific content in the video.

Recent research in [10] distinguished three types of subtitling: (1) bimodal subtitling: from English dialogues to English subtitles, (2) standard subtitling: from English dialogues to subtitles in the learner's mother tongue, and (3) reversed subtitling: from dialogues in the learner's mother tongue to English subtitles. Those subtitling types were then used in [11] where the author investigated the effect of bimodal, standard,

and reversed subtitling on L2 vocabulary recognition and recall. The experiment results showed that bimodal and standard group performs significantly better than the reversed subtitling group.

For speech translations for lectures, researchers in [12] have attempted to create a simultaneous German-English lecture translation system and achieved a word error rate (WER) of about 15% and translation performance of 23.65 BLEU, while other researchers have collected a corpus of German lectures in [13].

Our purpose is to investigate how to present captions or translated captions. Our final goal is to present the speech translated captions based on this study.

3. Methodology

We used two methodologies for this study: (1) using English captions and targeting both foreign students and Japanese students and (2) using Japanese captions and targeting only Japanese students. The preparations and experiments conducted for both methodologies are the same and explained in the following sub sections.

3.1 Video Preparations

In our experiment, we used videos from MIT OpenCourseWare lectures. While screening and selecting the video materials that would be used, several criteria were maintained. The first and the most important criteria was that the video materials must have a transcription for our experiments, because manually transcribed videos needed were very impractical and time consuming. Thus, pre-defined transcription was very important. The topics of the videos were also important criteria. Many topics are available but not all of them are common topics in other universities. It has to be a general topic, taught in every Computer Science department in any universities. With these criteria in mind, the lecture for video materials was selected.

From over 200 courses available in Electrical Engineering and Computer Science Department in MIT OpenCourseWare, one lecture was selected: Introduction to Computer Science and Programming (Fall 2008). From this course, we selected five different lectures to be used as our video materials: (1) Introduction and Goals; Data Types, Operators, and Variables, (2) Branching, Conditionals, and Iteration, (3) Common Code Patterns: Iterative Programs, (4) Abstraction through Functions; Introduction to Recursion, and (5) Lists and Mutability, Dictionaries, Introduction to Efficiency.

Once the video test materials were chosen, a shorter video about five minutes in duration was made from the original 60 minutes video for each lecture.

3.2 Captions

(a) English captions

The chosen lecture video comes with a separate transcription. We utilized this transcription to create our captions. We created five different caption types manually: (1) full text captions, (2) important sentence captions, (3) important phrase captions, (4) important keyword captions, and (5) no captions.

Full text captions were created directly from the available



Figure 1 Type of captions

transcription. We just need to select texts from the video transcriptions that match our edited videos and aligned it with the speech.

For important sentence captions, we extracted sentences that we thought as important from the full text captions. To differ it from full text captions, we limited the number of sentences in important sentence captions to be about 40-50% from total number of sentences in the full text captions.

Important phrase captions were created from important sentence captions. From each sentence in the important sentence captions, we extracted 1 or 2 important phrases while maintaining the number of words in important phrase caption is about 40-50% from number of words in important sentence captions.

As for important keyword captions, it was extracted from important phrase captions. From each important phrase, we extracted about 1 or 2 important keywords. Summary for each caption type can be seen in Table 1 while example for each caption type is shown in Figure 1.

Table 1 Statistics of various captions

(a) Full Text

| Lecture # | # of sentences | # of words | Duration (minutes) |
|-----------|----------------|------------|--------------------|
| 1 | 76 | 905 | 5:15 |
| 2 | 52 | 834 | 4:32 |
| 3 | 59 | 1054 | 5:22 |
| 4 | 57 | 975 | 5:31 |
| 5 | 31 | 428 | 5:22 |

(b) Important Sentence

| Lecture # | # of sentences | # of words |
|-----------|----------------|------------|
| 1 | 31 | 323 |
| 2 | 23 | 282 |
| 3 | 25 | 326 |
| 4 | 26 | 339 |
| 5 | 17 | 191 |

(c) Important Phrase

| Lecture # | # of phrases | # of words |
|-----------|--------------|------------|
| 1 | 31 | 156 |
| 2 | 23 | 114 |
| 3 | 28 | 133 |
| 4 | 28 | 152 |
| 5 | 19 | 81 |

(d) Important Keyword

| Lecture # | # of keywords | # of words |
|-----------|---------------|------------|
| 1 | 27 | 72 |
| 2 | 24 | 66 |
| 3 | 23 | 66 |
| 4 | 25 | 75 |
| 5 | 19 | 73 |

Using above captions, we embedded the captions into the lecture videos. The final outcome was five videos with five different types of captions in English for each lecture.

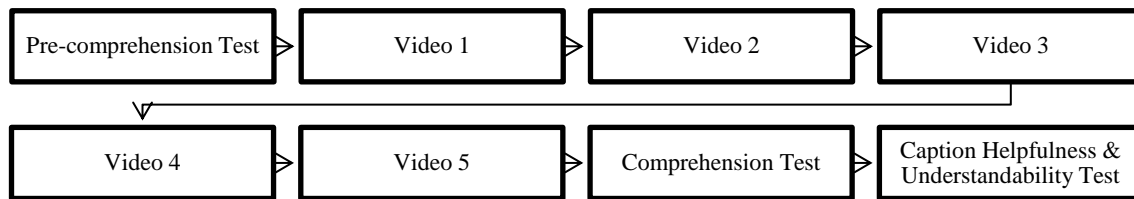


Figure 2 Study procedure

(b) Japanese captions

For the Japanese captions, we created the same five types of captions as the English version. Each caption type was a direct translation by a professional translator from its corresponding English caption.

3.3 Test Set

We created a test set consisted of three different types of test to find out whether captioning video really helps learners to understand the content of the video and to measure which type of captions gives the most benefit to the learners: (1) pre-comprehension objective test, (2) comprehension objective test, and (3) caption helpfulness and understandability subjective test.

Multiple choice questions were designed and created to provide quantitative data by which to compare the pre-comprehension test score, comprehension test score, and all captions score. Eight questions were prepared for the pre-comprehension test and ten questions were prepared for the comprehension test. All questions in pre-comprehension test are aligned with the topics that will be discussed in the videos. Each question in both pre-comprehension test and comprehension test has five choices as possible answer.

Scale questions from one to five were created in caption helpfulness and understandability test to measure the level of helpfulness and understandability for each type of caption. An open question for each caption asking what the learners thought about each type of caption was also put in this test.

All questions and possible answers in all tests were originally written in English. For English captions experiment, in order to accommodate Japanese students who may have difficulties in reading English questions, we provided Japanese translation below the English questions but all possible answers were left un-translated in purpose. For Japanese captions experiment, all questions and possible answers are translated into Japanese language.

3.4 Scoring

Scoring method for pre-comprehension and comprehension test was different from scoring method for caption helpfulness and understandability test. For the pre-comprehension and comprehension test, 10 points were given for each correct answer and 0 point for wrong answer. For caption helpfulness and understandability test, the score depends on participants' answer about each caption's level of helpfulness and understandability, from one to five.

3.5 Participant Selection

Participant was divided into two different groups. First group is international students and second group is Japanese students. Several criteria must be met by each participant: (1) they have to be majored in Computer Science in Toyohashi University of Technology, (2) for international students, they have to come from non-English native speaking country, and (3) for Japanese students, they must have a TOEIC score not less than 450. A total of 21 subjects – 10 international students and 11 Japanese students, participated in the testing sessions.

3.6 Procedure

All testing sessions were conducted in our Spoken Language Processing laboratory. All sessions were administered by the first author. Each testing session lasted approximately one hour. A total of about 10 testing sessions were held between March and June 2013.

The participants first took pre-comprehension test. There is no time limit for this test. They then watched a series of five videos with five different types of captions in randomized order. While watching the videos, the participants are not allowed to take notes, pause, rewind, and fast forward the video. Each participant viewed all five videos consecutively without any break. After watching all videos, all participants are required to take comprehension test and captions helpfulness and understandability test. There is also no time limit for both tests. A flowchart of this procedure is shown in Figure 2.

4. Results

Figure 3 presents the comparison among pre-comprehension test score, comprehension test score, and TOEFL score for each participant who viewed the English captioned videos. In the original data, all Japanese students' English proficiency is measured by TOEIC score and all international students are measured by TOEFL score. In order to make it comparable, we converted Japanese students' TOEIC score into TOEFL score by using tables in [14]. The data showed that 19 participants have a higher score in their comprehension test if compared to the pre-comprehension test. Only 2 Japanese students had lower score in their comprehension test compared to pre-comprehension test.

Correlations between pre-comprehension test score and comprehension test score with TOEFL score are measured by correlation coefficient as shown in Table 2. For international students, it almost did not show any correlation because the TOEFL score between all international students didn't have a

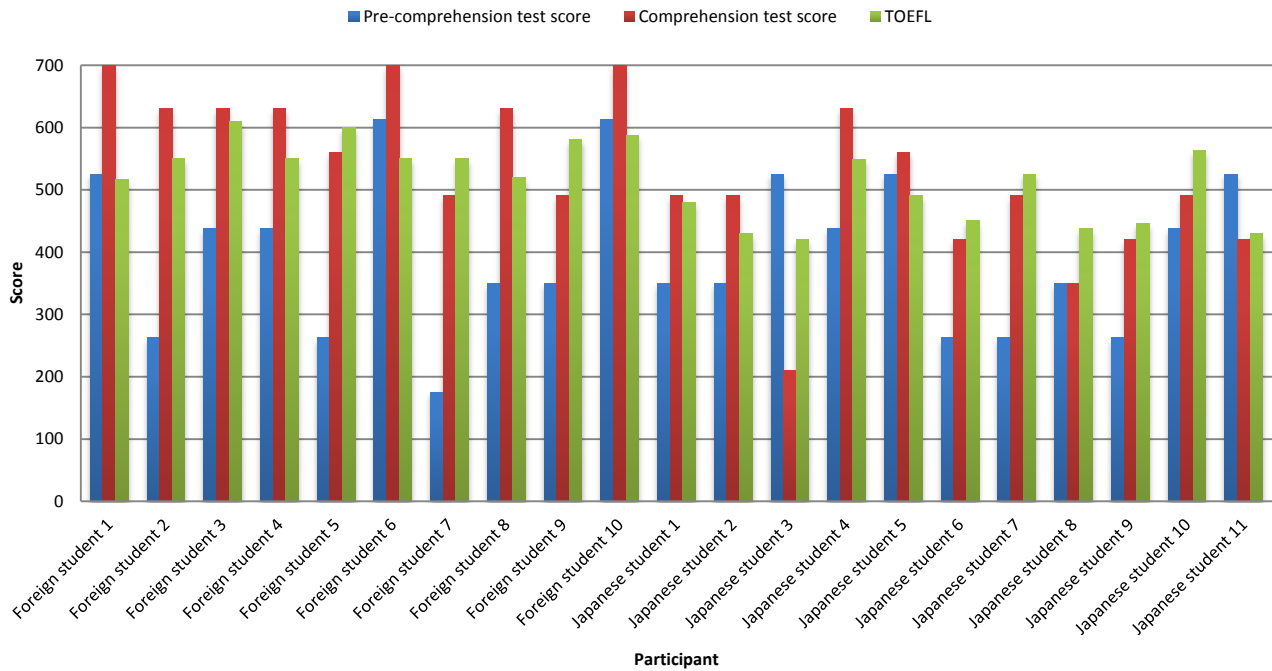


Figure 3 Pre-comprehension test vs. comprehension test vs. TOEFL score for English captioned videos

big difference. For Japanese students, the correlation coefficient is high and closer to +1 which means Japanese students with higher TOEFL score tend to have higher score in their pre-comprehension test and comprehension test. This is as expected that a sufficient level of English proficiency is needed to understand the content of lectures taught in English.

Figure 4 presents the comparison among pre-comprehension test score, comprehension test score, and TOEIC score for each participant who viewed the Japanese captioned videos. The data showed that all participants have a higher score in their comprehension test if compared to the pre-comprehension test. The correlation coefficient between pre-comprehension test score and comprehension test score with TOEIC score for the Japanese captioned videos are shown in Table 3.

Table 3 Correlation coefficient for Japanese captioned videos

| Pre-comprehension test vs. TOEFL | Comprehension test vs. TOEFL |
|----------------------------------|------------------------------|
| 0.257 | 0.576 |

Table 2 Correlation coefficient for English captioned videos

(a) Pre-comprehension test vs. TOEFL

| Pre-comprehension test vs. TOEFL (All) | Pre-comprehension test vs. TOEFL (Foreign students) | Pre-comprehension test vs. TOEFL (Japanese students) |
|--|---|--|
| 0.029 | -0.027 | 0.692 |

(b) Comprehension test vs. TOEFL

| Comprehension test vs. TOEFL (All) | Comprehension test vs. TOEFL (Foreign students) | Comprehension test vs. TOEFL (Japanese students) |
|------------------------------------|---|--|
| 0.698 | -0.257 | 0.692 |

Figure 5 shows the average score for caption helpfulness and understandability test. The result shows that the full text caption was the most helpful and the most understandable type of caption according to the participants, followed by important sentence captions, important phrase captions, important keyword captions and the least helpful and the least understandable was no caption. On the other hand, not all participants learned towards full text captions on their comments in the open question about what did thought about each type of caption. Some participants' comments are as follows:

- “Other captions than full text captions can be misleading.” (foreign student, 550 TOEFL)
- “Full text captions are the most helpful.” (foreign student, 610 TOEFL)
- “I prefer full text captions especially if the topic is hard so I can focus on the captions.” (foreign student, 600 TOEFL)
- “The speed of full text captions is too fast! I think important sentence captions are more suitable for me.” (Japanese student, 595 TOEIC)
- “Important sentence captions are easier to understand than full text captions.” (Japanese student, 690 TOEIC)
- “Important phrase is good but still lacking information.” (Japanese student, 550 TOEIC)

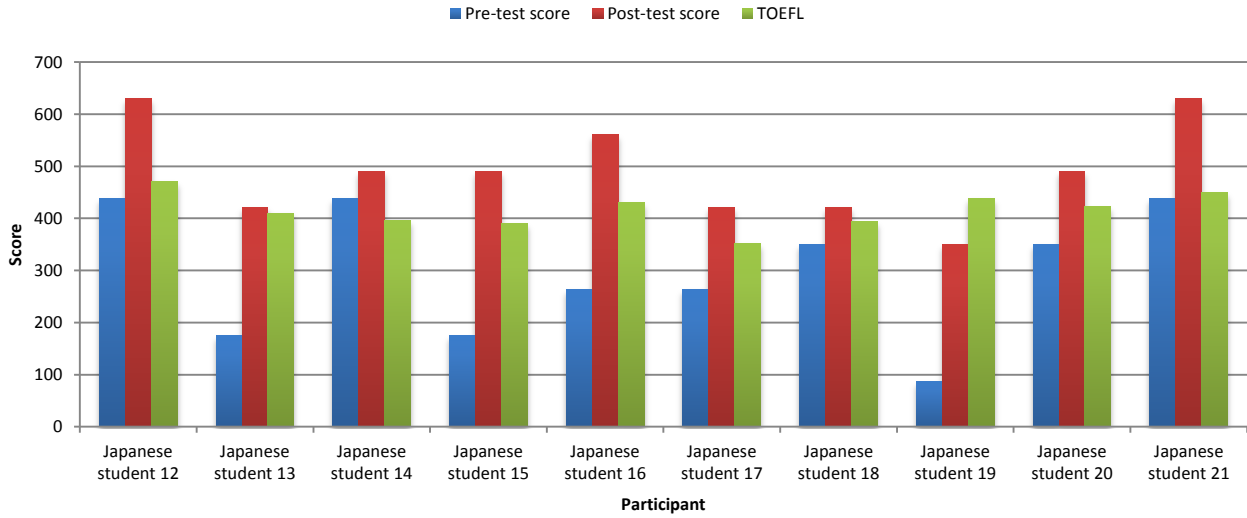


Figure 4 Pre-comprehension test vs. comprehension test vs. TOEFL score for Japanese captioned videos

the comprehension test and we calculated the average of each caption type.

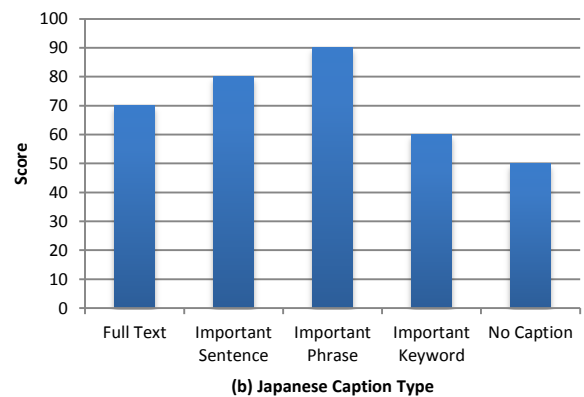
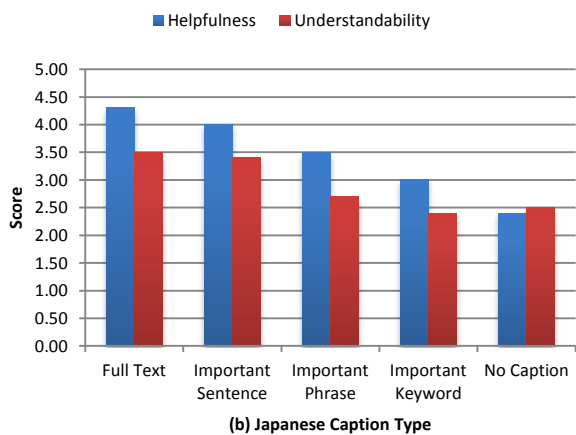
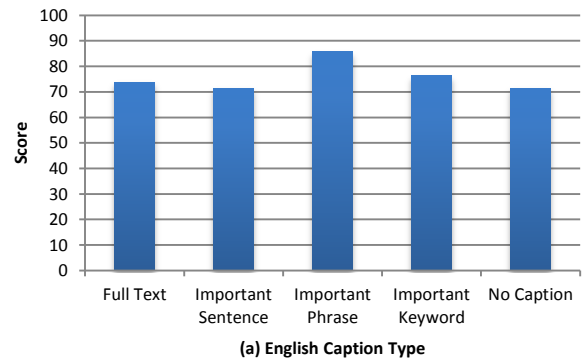
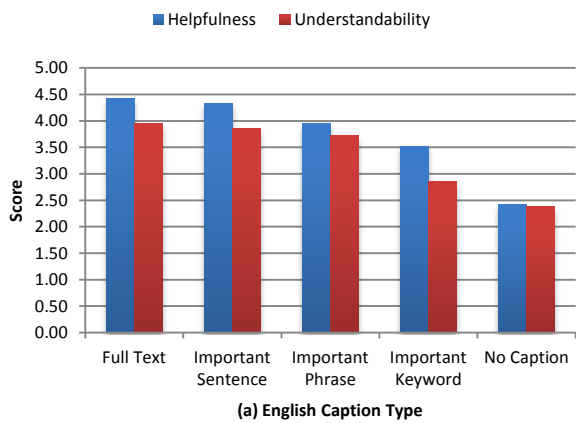


Figure 5 Test score (subjective evaluation)

Figure 6 Captions score (objective evaluation)

Figure 6 shows the average score for each type of caption. This score was calculated by grouping comprehension test questions based on caption type. Each caption has 2 questions in

From Figures 5 and 6, we can expect that speech translation for only important sentences or important phrases is enough as understandable captions.

5. Conclusion and Future Work

In this paper, we described a study of the effect of captioning lecture videos for learning in foreign language. We showed that captioning lecture videos will benefit the students for learning in foreign language. From the caption helpfulness and understandability test results, we showed that subjectively, the full text captions and important sentence captions are the most helpful and understandable caption types but objectively, by looking at the captions score, the important phrase captions is the most useful caption types.

Our planned future work is to create automatic Japanese captioning system for English lectures. Firstly, we will create the ASR system for English lectures that we will use to transcribe all lectures from MIT OpenCourseWare. Then we will create the captions, especially important sentence and important phrase captions, by using the transcribed lecture while translating it into Japanese language.

Reference

- 1) S. Kogure, H. Nishizaki, M. Tsuchiya, K. Yamamoto, S. Togashi, and S. Nakagawa: Speech Recognition Performance of CGLC: Corpus of Japanese Lecture Contents, Proc. INTERSPEECH, pp.1554-1561 (2008).
- 2) A. Ando, T. Imai, A. Kobayashi, H. Isono and K. Nakabayashi: Real-Time Transcription System for Simultaneous Subtitling of Japanese Broadcast News Programs, IEEE Trans. Broadcasting, Vol.46, No.3, pp.189-196 (2000).
- 3) A. Ortega, J.E. Garcia, A. Miguel and E. Lleida: Real-Time Live Broadcast News Subtitling System for Spanish, Proc. INTERSPEECH, pp.2095-2098 (2009).
- 4) I. Trancoso, et al.: Recognition of Classroom Lectures in European Portuguese, Proc. INTERSPEECH, pp.281-284 (2006).
- 5) M. Cettolo, F. Brugnara, and M. Federico: Advances in the Automatic Transcription of Lectures, Proc. ICASSP, pp.1-769-772 (2004).
- 6) Y. Zhao, et al.: An Automatic Captioning System for Telemedicine, Proc. ICASSP, pp.957-960 (2006).
- 7) T. Garza: Evaluating the use of Captioned Video Materials in Advanced Foreign Language Learning, Foreign Language Annals 24-3, pp.239-258 (1991).
- 8) P. Winke, S. Gass, and T. Sydorenko: The Effects of Captioning Videos Used for Foreign Language Listening Activities, Language Learning & Technology 14-1, pp.65-86 (2010).
- 9) H. G. Guillory: The Effects of Keyword Captions to Authentic French Video on Learner Comprehension, Calico Journal 15-1/3, pp.89-108 (1998).
- 10) N. T. Zanon: Using Subtitles to Enhance Foreign Language Learning, Porta Linguarum 6, pp.41-52 (2006).
- 11) A. A. Zarei: The Effect of Bimodal, Standard, and Reversed Subtitling on L2 Vocabulary Recognition and Recall, Pazhuhesh-e Zabanha-ye Khareji 49, pp.65-85 (2009).
- 12) M. Kolss, et al.: Simultaneous German-English Lecture Translation, Proc. IWSLT, pp.174-181 (2008).
- 13) S. Stuker, et al.: The KIT Lecture Corpus for Speech Translation, Proc. LREC, pp.3409-3414 (2012).
- 14) TOEFL Equivalency Table
<http://secure.vec.bc.ca/toefl-equivalency-table.cfm>