Effects of Automatic Hidden Caption Classification on a Content-based Computer-Assisted Language Learning System for Foreign Language Listening

Ching-Kun HSU[a*] & Chih-Kai CHANG[a]

[a]*Department of Information and Leaning Technology , National University of Tainan, Taiwan*
*d09608002@stumail.nutn.edu.tw

Abstract: This study develops a content-based computer-assisted language learning tool for training English listening through multimedia. The system automatically provides instant vocabulary translation and classification to users whenever they pause while watching videos. Experiments were conducted to compare the novel system with the traditional solution using a translation machine to support subtitle comprehension. The first goal of the novel system is to support listening comprehension, and the second goal is to confirm its usability. These two goals were successfully achieved in the study. Eye tracking was used to observe variations in gazing and reliance on subtitles. The results of vision analysis show that the degrees of reliance on subtitles remain unchanged.

Keywords: instant translation, automatic vocabulary classification, subtitles, listening comprehension

1. Introduction

Language is the instrument for communication. Listening plays an important role regardless of the use of first or second foreign language. Previous research has investigated the importance of listening, speaking, reading, and writing in English in 146 freshmen and found that more than 51% of participants considered listening to be the most essential aspect of language [8]. In Taiwan, learning English is of vital importance because it is the major international communication tool. However, many English teachers had focused on teaching grammar and enhancing the vocabulary acquisition of learners in the past few years [4], so that English listening comprehension has become the weakest skill of the Taiwanese. Training in English listening has currently become more convenient and easier to access because of the progress of technologies; for example, computers and the Internet had brought many new media and diverse formats for listening training. Digital multimedia, such as MPEG Audio Layer III (mp3) and Digital Video Disk (DVD), can expose EFL learners to English programs, songs, and videos. Many Web sites provide online English learning, such as nonstopenglish (www.nonstopenglish.com), englishforums (http://www.englishforums.com), englishclub (http://www.englishclub.com), and tefl (www.tfel.net). There are also online film databases, such as imdb (http://www.imdb.com/) and Drew's scripts-o-rama (http://www.script-o-rama.com). These digital formats often offer various subtitle choices, repeatable viewing at will, and easy storage of files.
Many studies have confirmed the contributions of subtitles or captions in learning foreign languages through movies, films, or videos. In recent decades, research has indicated that videos with subtitles are helpful in learning reading, vocabulary, and listening in a foreign language [1]. The scholars stated that videos with subtitles are conducive to listening

comprehension, especially because the learners are encouraged to select the movies or videos they are interested in to practice their listening skills [7]. Although comprehension of foreign culture and internationalization from the videos are difficult, the subtitles can help solve the comprehension problems. Still, learners cannot rely solely on reading subtitles all the time to train their listening proficiency. Other studies have indicated that the lack of subtitles in a video induces learners to pay attention to various pronunciation appearances, such as reduced forms, assimilation, elision, and resyllabification [11], such that translation subtitles should not always be provided for listening practice. They have also suggested that subtitles should be time-toggled on and off to assess their usefulness [3, 12].

This study aims to investigate whether the automatic caption classification and vocabulary interpretation functions provided by the content-based CALL system can support the listening comprehension of English learners. It utilized multimedia, such as videos, movies, and films, in English for learners to make adaptations to oral dialogue and to train their daily listening comprehension skills. The issue of providing time-appropriate support during video play was incorporated into the pause button, which was used by learners to obtain auxiliaries once they encounter listening comprehension problems. Thus, this study introduced the innovation approach of the automatic caption difficulty classification hidden or appearance as well as interpretations for listening training and comprehension.

2. Literature Review

Many certifications, such as TOEFL, TOEIC, and GEPT, have included listening examinations, so students learning English as a foreign language must develop listening comprehension proficiency. Previous studies have indicated that captioned videos for foreign language learning bring the native speakers' voices of foreign language into the learning surroundings. Further, these have become more and more common due to their use as a pedagogical tool in online courses, especially that they are easily accessed and produced [12].

Vision can provide both context and non-verbal input. Videos provide simultaneous seeing and hearing for foreign language learners, which remedy the lack of sufficient language knowledge and comprehension resulting from learning by listening alone [9]. Multimedia can provide the listener with various conditions, including voice, vision, and texts in the form of subtitles or captions offered in different languages. Previous studies have shown that computer technology is useful for language learning, and the use of multimedia is beneficial for vocabulary learning [5]. Foreign language films with subtitles have also been shown to help listening comprehension [7]. Slowing down or speeding up broadcasting tempo is not necessary for listening comprehension [2, 11].

2.1 Literature Summary and Research Questions

Past studies have pointed out that captioned videos had better listening comprehension and learning outcomes than non-captioned ones [10]. However, there are few works on the development and exploration of an assisted language learning tool similar to the system developed in this study. The experimental group will be exposed to automatic classification of degrees of vocabulary difficulty in order to determine the necessity of subtitles. Thus, the captions of easier words will be hidden, while those of difficult words will be automatically shown with translations. The experimental group will be compared with the comparison group using captioned videos that have been verified to produce better listening comprehension than non-captioned videos. Eye tracking will be used to observe gaze variations while the learners watching the video in the experimental and comparison groups.

Finally, all participants will be assessed using the Computer System Usability Questionnaire (CSUQ) [6]. The following are the research questions that will be explored:

1. Does the experimental group achieve similar comprehension of video content and learning of vocabulary as the comparison group?
2. Do eye movements with classified hidden captions have differences or similarities with the eye movements with captions shown throughout?
3. Do the users perceive any difference in usability between the experimental group and the comparison group?

Based on prior research, captioned videos result in better listening comprehension than non-captioned ones. Therefore, the first hypothesis in this study is that the experimental group exhibits comprehension and vocabulary gains similar to the comparison group, so the independent paired t-test between the experimental and comparison groups shows no significant differences. The ordering effects of the two groups have been excluded from the counterbalanced experiment design. The experimental group is also expected to demonstrate high usability.

3. Method

The participants in this study were learners of English as a foreign language, and the sound track and captions of the video were both shown in English. Both control and experimental groups were unable to see subtitles in their first language during the video play. The comparison group was exposed to normal video play in a computer. Students in the comparison group were instructed to push the pause button of the broadcast software whenever they were not able to comprehend what they heard, allowing them to consult a dictionary and find the Chinese translation of the unfamiliar words, just like the traditional and common practice. All the English captions were shown at the bottom of the screen throughout the video duration. For convenient and quick dictionary consultation during video play, instant translation was embedded in the broadcast software. When the video is paused, and the user pointed to an unfamiliar word in the caption, its Chinese translation will be shown in a small window on the right.

The experimental group was exposed to the novel system, where the easier captions were automatically hidden, and the Chinese translations of the more difficult words were displayed in a small window on the right when the video is paused. The system automatically classified the words based on their degree of difficulty. Each participant was included in both groups and so was able to experience both control and experimental treatment. To exclude the ordering effects of the two different instruments, the experimental research method had a counterbalanced design (Table 1).

Table 1. Counterbalanced Design

Duplication	Instruments	
	Experimental group method	Comparison group method
Earlier	31 participants	20 participants
Latter	20 participants	31 participants

There were 31 students who used the experimental method first and then used the control method afterwards. On the other hand, 21 used the control method first and then used the experimental method later.

3.1 System framework

The broadcast content of the video files provided both visual and aural stimulus. The database stored vocabularies, translations, difficulty levels, and usage frequencies. When the player software is paused, the words within the difficulty level that the learner set are hidden, while the words beyond the difficulty level are automatically translated. The player software was designed with functions specific for English listening training (Figure 1). The research tools include headphones, video files, eye-tracking machine (Mangold Vision), and a personal computer installed with the system.
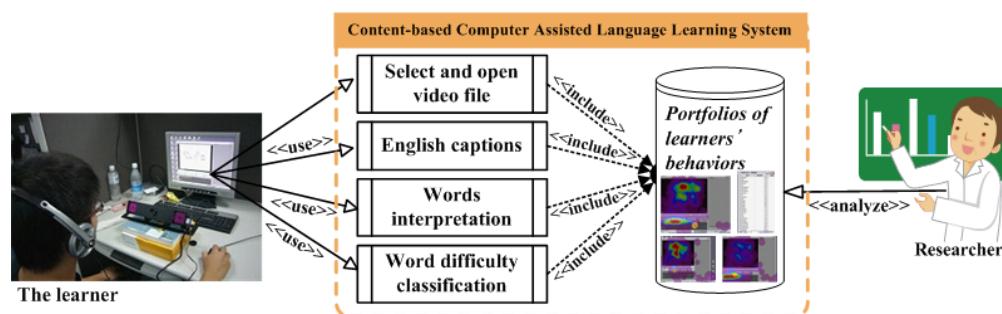

Figure 1. Architectural diagram of the system

Because past research has been pointed out that it is not necessary to slow down or speed up broadcasting tempo in terms of listening comprehension, the study will not taken the buttons of slowing down or speeding up into consideration. The listeners have to practice getting used with the common oral speed. The basic functions in the video player program are the buttons of the play, pause, and replay. The study only used the buttons of play and pause during the experiments lest the function of replay interferes with the results of the study. The study aimed at investigating the effectiveness of the captions-filtering when the player is paused, so that the experimental group can provide users with automatic classification of the caption words by an easy and quick action.

3.2 Participants

A total of 51 freshmen from the same department participated in the study. The English learning experience of the participants was controlled, with seven years' experience on the average for each participant. Some experienced the control method first, while some were exposed to the experimental method first.

3.3 Materials

3.3.1 Videos

Two videos of the same consistency were selected: one for experimental group and the other for the comparison group. Both short videos were related to computer science knowledge, had the same degree of English difficulty, and were of approximately similar lengths.
To avoid interference between the two videos due to familiarity after watching one of them, the two videos were of different themes introducing computer science. One video is an introduction to personal computer hardware, and the other is on network information searching. The participants were not allowed to open and select videos by themselves.

3.3.2 Listening and comprehension tests

The pronunciation of two native speakers of English was recorded in advance for the listening tests. Both of them are professional English teachers for more than five years. The recorded contents of the listening tests were related to video content. The comprehension tests were related to the professional knowledge taught and the vocabulary used, and was designed by a computer science teacher after watching the video.

3.3.3 Questionnaire

There are 19 questions in the CSUQ (Lewis, 1995) with seven-point Likert scales. Three dimensions can be investigated by the questionnaire, including perceived ease of use, usability, and satisfaction.

3.4 Procedure

After introducing the functions and operation of the broadcast program to the users, they were exposed to both the experimental and comparison groups. The video was selected and opened by the researcher beforehand. After being exposed to one group, the participants were made to take the listening and comprehension tests related to video content. The duration of every treatment stage and the distribution of the participants are shown in Figure 2.
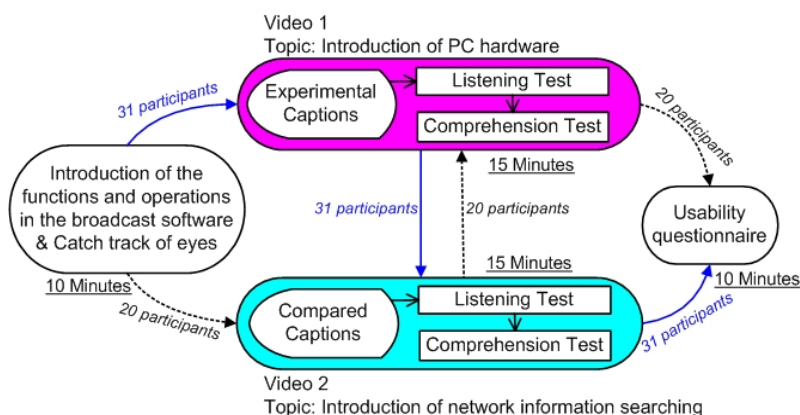

Figure 2. The flowchart o the experimental procedure.

3.5 Data Collection and Analysis

The participants took the listening and comprehension test right after watching each video. The scores were collected for paired-samples t test. The lack of significant differences between the two groups indicates that the automatic caption classification system trains the listening of the learners better than without captions while producing the same comprehension degree as the captioned videos. The movement of the eyes and the frequencies of the stare positions were classified and analyzed from the data collected by the eye tracking machine. After the experimental treatment, the students filled in the usability questionnaire, and the results were analyzed.

4. Results

4.1 Research question 1: Does the experimental group achieve similar comprehension of video content and learning of vocabulary as the comparison group?

The mean scores between the experimental and comparison groups were measured after conducting the counterbalanced instruments. The mean scores of the listening test in the experimental group were improved by more than five marks, while those of the comprehension test were upgraded by more than four marks. Table 2 presents the descriptive statistics.

Table 2. Results of the listening and comprehension tests

Two tests after the counterbalanced instruments	Order	N	Listening test Mean	Listening test Sd	Comprehension test mean	Comprehension test Sd
Experimental group	Earlier	31	55.65	31.43	63.71	25.69
	Latter	20	55.00	30.99	61.25	24.97
Column average	**Integration**	**51**	**55.39**	**30.95**	**62.75**	**25.19**
Comparison group	Latter	31	52.42	30.52	57.26	26.77
	Earlier	20	46.25	24.70	60.00	24.87
Column average	**Integration**	**51**	**50.00**	**28.28**	**58.33**	**25.82**

Paired-samples t-test of the two treatments revealed no significant differences in the listening outcomes (t=1.35) and comprehension degrees (t=1.07) (Table 3). Thus, the first research hypothesis is accepted. The CALL system that partially hides captions automatically has the same efficacy as captioned videos. Moreover, the novel system appears to be more effective than the provision of captions throughout video play.
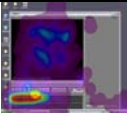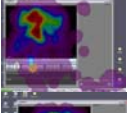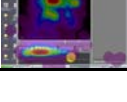
Table 3. Paired-Samples T-Test of the control and experimental groups

Experimental – comparison group	N	Paired Differences Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Listening test	51	5.39	28.42	3.98	1.35	50	0.182
Comprehension test	51	4.41	29.46	4.12	1.07	50	0.290

4.2 Research question 2: Do eye movements with classified hidden captions have differences or similarities with the eye movements with captions shown throughout?

Analysis of the hot areas detected by the eye tracking machine shows that there is no remarkable difference between gaze focus in the control and experimental groups. Thus, the system does not evidently change the habits of the participants when watching videos in a foreign language. The stare positions were categorized into three types, and their distributions are shown in Table 4. More than 70% of the participants have the habit of depending on captions when they watch movies in a foreign language.

Table 4. Distribution of visual focus of the participants

Stare position	Example figures	Comparison group			Experimental group		
		Latter	Earlier	Total	Earlier	Latter	Total
Mainly on the captions		23	15	38	24	16	40
Mainly on the video		4	2	6	3	3	6
Both captions and video		4	3	7	4	1	5
Total numbers of students		31	20	51	31	20	51

Among the six participants in the comparison group staring mainly on the video and not on the captions in the experimental group, four achieved full marks in both tests. Thus, higher-level students are not affected by the lack of captions, while general-level students do not notice the captions when they have fewer captions to depend on.

4.3 Research question 3: Do the users perceive any difference in usability between the experimental group and the comparison group?

The results of 51 questionnaires are displayed in Figure 3. These show that 65% the students found the novel system to be agreeable and easy to use, and only a few did not perceive usability. More than 62% of the participants were satisfied with the system. Positive responses were received based on ease of use, usefulness, and satisfaction.
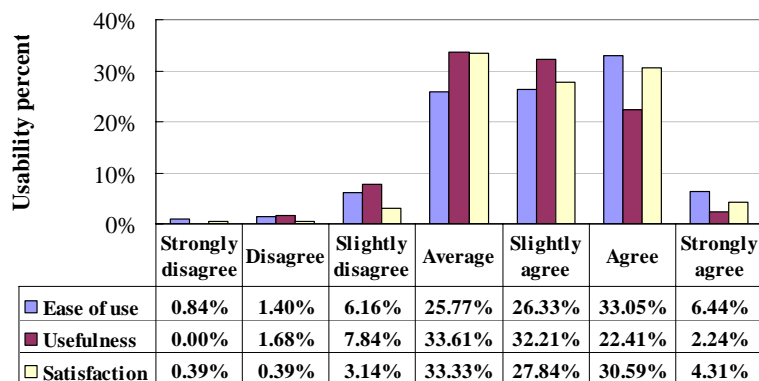

	Strongly disagree	Disagree	Slightly disagree	Average	Slightly agree	Agree	Strongly agree
Ease of use	0.84%	1.40%	6.16%	25.77%	26.33%	33.05%	6.44%
Usefulness	0.00%	1.68%	7.84%	33.61%	32.21%	22.41%	2.24%
Satisfaction	0.39%	0.39%	3.14%	33.33%	27.84%	30.59%	4.31%

Figure 3. Results of the usability questionnaire investigation

5. Conclusion

The study developed a content-based computer-assisted language learning system. Counterbalanced experiments revealed that automatic classification, concealment, and interpretation of words resulted in listening and comprehension degrees comparable with those of the comparison group that used real-time captioning and translation software. Moreover, the novel system appears to be even more effective than the control method. The usability questionnaire showed that more than 60% of the participants found the novel

system to be usable in terms of ease of use, usefulness, and satisfaction. In addition, the eye tracking system detected that the different display modes of captions did not result in remarkable differences in the dependence degrees on captions of the students. The study correspondingly provides two suggestions for future research in this area. First, in light of system development, vocabularies with difficult degrees and usage frequencies should be expanded to include more words and classification levels. In the perspective of teaching experiments, the pre-test can be taken into consideration. Moreover, video displays with various lengths or videos in other languages can be further included in the experiments. Overall, the novel system proposed in this study is expected to be used for adaptive learning in foreign language listening training.

Acknowledgements

References

[1] Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language learning*, *42*(4).497-527.

[2] Derwing, T. & Munro, M. J. (2001). What speaking rates do non-native listeners prefer? *Applied linguistics*, *22*(3), 324–337.

[3] Grgurovi, M., & Hegelheimer, V. (2007). Help options and multimedia listening: students' use of subtitles and the transcript. *Language learning & technology*, *11*(1), 45-66.

[4] Liu, P. L., Chen, C. J., & Chang, Y. J. (2009). Effects of a computer-assisted concept mapping learning strategy on EFL college students' English reading comprehension. *Computers & Education*, *54,* 436-445.

[5] Liu, M., Moore, Z., Graham, L., & Lee, S. (2002). A Look at the Research on Computer-Based Technology Use in Second Language Learning: A Review of the Literature from 1990-2000. *Journal of research on technology in education*, *34*(3), 250-273

[6] Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International journal of human-computer interaction, 7*(1)**,** 57-78.

[7] Markham, P., Peter, L. & McCarthy, T. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign language annals*, *34*(5), 439–445.

[8] Ou，H.C. (1996). The effectiveness of teaching English listening comprehension as a second or foreign language learning. New York: Longman.

[9] Seo, K. (2002). The effect of visuals on listening comprehension: A study of Japanese learners' listening strategies. *International journal of listening*, *15*, 57–81.

[10] Stewart, M. A. & I. Pertusa. (2004). Gains to language learners from viewing target language close-captioned films. *Foreign language annals*, *37*(3), 438–447.

[11] Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching*, *40*(03), 191-210.

[12] Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1). 65-86.