Development of a Language Learning System with an Adaptive Caption Filtering Mechanism on Handheld Devices

Ching-Kun HSU[a], Gwo-Jen HWANG[b*], Chih-Kai CHANG[a]
[a]*Department of Information and Learning Technology, National University of Tainan, Taiwan*
[b]*Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan*
*gjhwang.academic@gmail.com

Abstract: This study developed a language learning system on handheld devices based on different caption-filtering degrees of videos, including 220 sight words, the highest frequently used 1000 words, and the highest frequently used 2200 words, for adaptive assistance in English comprehension and vocabulary acquisition of eleventh graders. Different students were treated different bilingual caption modes based on their pre-test in the experimental group while all the students were treated identical caption modes in the control group. During the three experimental times, the students' English listening comprehension and vocabulary acquisition status was evaluated after each time of experiment. The experimental results found that the students in the experimental group prefer partial captions they used and perceived enjoyments. In terms of listening comprehension and vocabulary test, the experimental group outperformed the control group when the students in the control group were supported by the captions filtered and hidden the highest frequently used 1000 and 2000 words. The experimental group had similar learning effectiveness with the control group when the students in the control group were supported by the captions filtered and only hidden the 220 sight words. The adaptive caption-filtering approach proposed in the study not only enabled the interface adaption for playing videos on handheld devices, but also achieved and confirmed that different students need different quantity of information such as caption support for better listening comprehension and vocabulary acquisitions.

Keywords: Caption filtering, language learning, listening comprehension, vocabulary acquisition, handheld devices

1. Background and Objective

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]. Markham, Peter, and McCarthy [2] stated that videos with subtitles are conducive to listening comprehension, as the students are encouraged to select the movies or videos they are interested in to practice their listening skills. Vision provides both context and non-verbal input. Videos simultaneously provide viewing and hearing for foreign language students, which remedy the lack of language knowledge and comprehension resulting from learning by listening alone [3]. Multimedia can provide the listener with various conditions, including voice, vision, and texts in the form of subtitles or captions offered in different languages. Foreign language films with subtitles have also

been shown to help listening comprehension [2]. Slowing down or speeding up broadcasting tempo is not necessary for listening comprehension [4][5]. Previous studies have shown that computer technology is useful for language learning; and the use of multimedia is conducive for vocabulary learning [6]. Another study carried out English vocabulary learning by using mobile technologies for one year to enhance incidental vocabulary learning in an English as a Medium of Instruction, and its result found that the students applied PDAs to improve referential, situated, constructive, reflective, explorative, and conversing uses of vocabulary learning [7]. Therefore, the researchers indicated that mobile devices can be used in more flexible, novel and extended ways for English as a Foreign Language (EFL) vocabulary teaching and learning in higher education, which considers needs and contexts of students [7]. A study especially focused on the attitudes of language students toward using mobile phones for training listening purpose by browsing wireless application protocol (WAP) sites; accordingly, its result stated that the WAP site was effective for learning listening skills [8].

When learners are exposed to a new message from video, they sense first by paying attentions. In IPT[9], memory is classified into three categories: sensory, short-term, and long-term memory [10]. The external stimulus will be received from the sensory organs such as ears, eyes and so on. After sensory register, individuals select and filter some message which they received in their information processing. The message will then be stored in the short-term memory. After encoding or rehearsal the message in the short-term memory, the forms of message will be transformed in to semantic or mental image forms, and combined with their prior knowledge. People often recall by some clues to retrieve the relevant data in the long-term memory, and then send the obtained data back to short-term memory, such as the memory of tone, scene concept, body language and so on. After information processing in the short-term memory, the new information which is the results of their inferences and judgments is stored back to the long-term memory and their behavioral response will be output from short-term memory.

The short-term memory is limited, but the long-term is infinite. Due to the limitation of short-term memory or working memory, the prompt assistant for EFL listening is necessary, resulting the students in paying attention to the message which is received from sensory register. The capacity of short-term memory is limited and ranged from 5 to 9 words (or 7 words on average) [11]. Given this limitation, most of the recollections in this memory are also immediately forgotten. To conform to the restrictions for short-term memory in IPT, the largest number of words set for 4-inch, 7-inch, and 10-inch mobile devices are 5, 7, and 9, respectively. This study further evaluated whether the results of caption filtering satisfy the thresholds. To better fitting in with the limitation of small screen of smart phone and PDA, and to fulfill the level of students' listening proficiency, an adaptive caption-filtering system was developed in terms of three difficulty levels and used in English courses.

1.1 System Development

To prevent students from distractions stemming from simultaneous attention to images and captions, and rectify limited comprehension of listening training materials, the study improved interface adaption of caption on handheld devices by caption filtering, as well as supporting listening comprehension. More importantly, caption filtering must not become learning obstacles but should have positive effects on listening once interface adaption for small screen is improved. This is the primary goal of the study. Toward the achievement of the aforementioned goal, more difficult or new words were shown in the caption and

the Chinese translation of the English captions was also showin on handheld devices to aid listening comprehension, shown as the left of Figure 1.

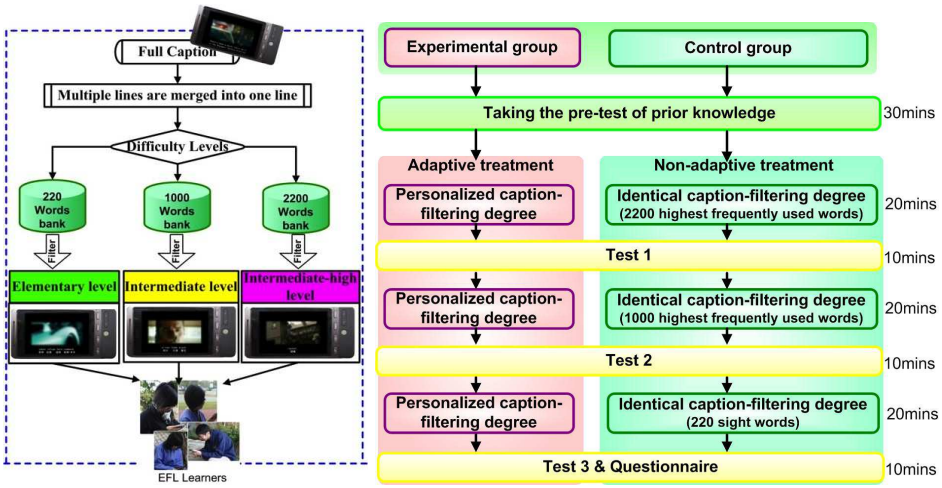


Figure 1. The learning system and experimental procedure

Python programming language was used to develop the caption filtering program. The program is called the Caption Filtering-based Language Listening System (CFLLS) on handheld device. The operation of the system comprises two parts. The first is the merging of captions that are originally displayed in more than one line in a video clip. The second step is the filtering of the words to hide the most frequently used words and to show the more difficult or seldom seen words. Moreover, two databases are employed. One is the word bank which contained three sub-bank storing the 2200 most frequently used words, 100 highest frequently used words, 220 sight words respectively. In other words, there were totally three caption-filtering levels which are 220 sight words, the highest frequently used 1000 words, the most frequently used 2200 words in the system. The other database stores the captions after using the merge mode and the filter mode based on the caption-filtering level user selected from the three word banks. The original caption files can be stored as “srt” files, consisting of time sequence, time interval, and captions. After the highest-level caption-filtering process, the remaining captions are subjected to statistical computation to determine whether they satisfy the expectation for the adaptation to the screen sizes of different handheld devices. The results of the merging and filtering processes are also stored in the database. The information on the original captions, such as the number of words in each sentence and total number of words in the video, is saved in the database. The information on the merged captions, such as the times at which merging in a video was conducted is stored in the database whenever the merge mode is run. The information that remains after running the filtering mode, such as the captions filtered, remaining captions, or the number of remaining words in each sentence is also stored.

2. Method

2.1 Participants

The subjects included two classes whose age was eighteen on average in a country learning English as Foreign Language. A total of 73 eleventh-grader students participated in the study. One class served as the experimental group and the other was the control group. The experimental group, comprising of 39 students, was guided by the adaptive

captions which was defined as a personalized caption-filtering degree that met the requirements and proficiency of individual students. The control group with 34 students was guided by the caption-filtering degree which was identical in the whole class. All of the students have learned English for eight years on average.

2.2 Research design

2.3

The research tools in this study consisted of a pre-test, three times of post-tests, and the questionnaire used for investigating the students' perceived satisfaction, enjoyment and their preferences of caption modes. The listening pre-test is the primary and intermediate degree of General English Proficiency Test consisted of three sets of questions. Each set of question contained fifteen multiple-choice items. There were forty-five multiple-choice items in total, with the full score of 120. The students who gained more than 80 marks were regarded as achieving the intermediate degree of English listening proficiency. The experimental group students falling into this group would then watched the movie involving 2200 most frequently used words. The experimental group students who received 60-80 marks watched the movie involving 1000 most frequently used words. Finally, the experimental group students who obtained less than 60 marks watched the movie involving only 220 sight words. After watching each video clip, the students in the experimental and control groups took the same post-tests comprising of five questions for listening comprehension tests and five questions for vocabulary tests. The study compared the number of questions which each student correctly answered.

3. Results

3.1 4.1 Different degree caption-filtering results of experimental video clips

There are three video clips. The students were given 20 minutes to complete watching movie (whose actual running time is 15 minutes each).The distribution of words left in each sentence after different caption-filtering degrees is shown in Figure 2. For confirming the interface adaption of the video clips to meet the screen size of the handheld devices, the background information of the three video clips in terms of captions before and after caption-filtering process are calculated.

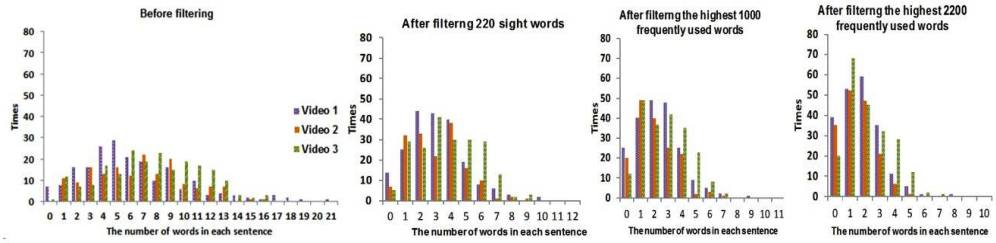
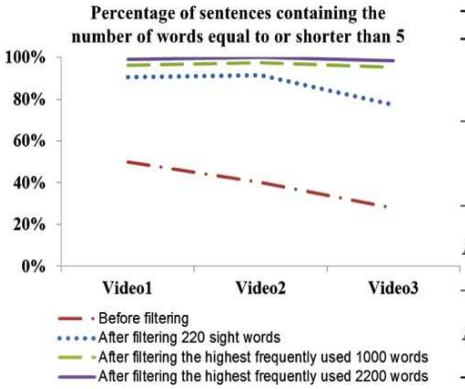


Figure 2. The distribution of words left in each sentence after different caption-filtering degrees

Figure 3 shows the percentage of sentences with five words or less each within the video clips before and after different caption-filtering degree processes. It was found that there are more than half of sentences containing more than five words in the video clips before caption-filtering. After the captions were filtered by the most frequently used 2200 words, almost one hundred percent of words left in each sentence are equal to or shorter than five words. Around eighty to ninety percent of words left in each sentence are equal

to or shorter than five words after the captions are filtered by the 220 sight words so as to stuff the small screen with the shortened captions in a clear font size. The left of Figure 3 shows all the results of the percentage of words left in captions after different filtering process in terms of different thresholds which are 5, 7, 9 respectively. It was found that almost all of the sentences fulfilled the limitation of the small screen size after filtering 2200 words.



Caption process	Threshold	<= 5words	<=7 words	<=9 words
Before filtering	Video1	50%	69.61%	82.35%
	Video2	40.12%	61.11%	81.48%
	Video3	27.88%	48.56%	66.83%
After filtering 220 sight words	Video1	90.69%	97.55%	99.02%
	Video2	91.36%	98.15%	100%
	Video3	77.40%	97.60%	100%
After filtering the highest frequently used 1000 words	Video1	96.08%	99.51%	100%
	Video2	97.53%	100%	100%
	Video3	95.19%	100%	100%
After filtering the highest frequently used 2200 words	Video1	99.02%	99.51%	100%
	Video2	100%	100%	100%
	Video3	98.56%	100%	100%

Figure 3. Percentage of sentences containing the words equal to or shorter than the threshold number

3.2 Results of pre-test and post-tests

A pre-test was conducted before the experiment. Results showed that the mean values and standard deviations of the pre-test scores were 69.85 and 9.42 for the control group, and 68.33 and 8.94 for the experimental group. The t-test result ($t=-0.71$, $p > .05$) shows that there was no significant difference between the two groups. Therefore, it is evident that the two groups of students had equivalent prior knowledge before the learning activity.

After the learning activity, the independent t test was used to examine the difference between the two groups and to compare the number of questions which each student answered correctly. When the control group watched the first video clip whose captions were filtered the most frequently used 2200 words, the mean value and standard error of the post-test for the listening comprehension in the control group were 1.88 and 1.15 for the control group. Meanwhile, the experimental group watched the first video clip whose captions were filtered by different degrees such as 220 sight words or the most frequently used 1000 words or 2200 words. The mean value and standard error of the post-test scores for the listening comprehension of the experimental group were 3.03 and 0.92. According to the results ($t=4.83$, $p<.05$), there was a significant difference between the two groups; that is, the students who learned with the support of adaptive caption showed significant better listening comprehension than those who learned with the support of the identical caption-filtering degree. All the tests are shown in Table 1. It was found that hiding 220 sight words from captions for senior high school students in the English as Foreign Language country achieved similar support effectiveness on both listening comprehension and vocabulary acquisition with providing adaptive captions for students in the experimental group.

If a student answers a question correctly, the student will gain 20 marks. Conversion of the number which each student answered correctly into marks with a full score of 100 is shown in the following Figure 4. Students with different capabilities in terms of pre-test results received different quantity of information from captions in the experimental group. There is rare difference between the three tests when the leaners watch video with assistance of personalized captions in the experimental group. However, the system

provides all the students with identical quantity of information from caption no matter how their listening proficiency in the control group. The results showed that more information provided from the captions helped better listening comprehension and vocabulary acquisition.

Table 1. Descriptive data and Independent t test of the post-test results

Filtering of control group	Tests	Group	N	Mean	SD	*t*
2200	Listening	Experimental	39	3.05	0.92	4.83*
		Control	34	1.88	1.15	
	vocabulary	Experimental	39	2.97	0.93	8.45*
		Control	34	1.21	0.84	
1000	Listening	Experimental	39	3.03	0.93	2.08*
		Control	34	2.38	1.58	
	vocabulary	Experimental	39	2.77	0.96	3.07*
		Control	34	2.09	0.93	
220	Listening	Experimental	39	3.31	1.10	-0.66
		Control	34	3.47	0.99	
	vocabulary	Experimental	39	3.26	1.16	-0.33
		Control	34	3.35	1.30	

*$P<.05$

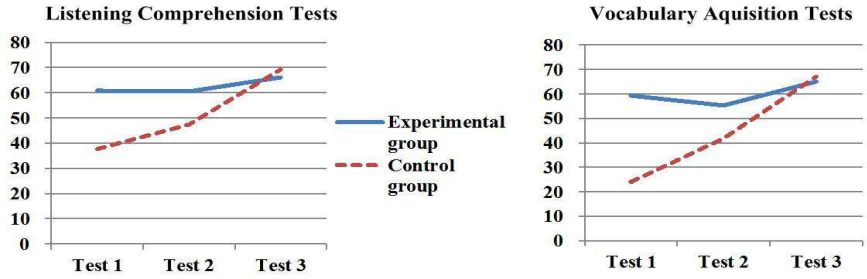


Figure 4. The effectiveness of listening comprehension (left) and vocabulary acquisition (right)

4. Discussions and Conclusions

In the study, an adaptive English listening system by caption-filtering based on the difficulty degree of vocabulary was developed. The effects of the supporting learning system of adaptive caption-filtering (i.e, experimental group) and identical caption-filtering approach (i.e., control group) on the listening comprehension and vocabulary acquisition of students were compared. In addition, after the experiment, the study ranked the students' preference of caption and subtitle modes through a questionnaire administered to both control and experimental groups. It was found that most students had a preference and tendency to reply on both subtitles and captions, which is like the filtered captions and the translation of the left words in the study. The findings conform to a previous study by Guillory [12] considering video embedded keywords having similar effects to full text captions, but which had better effects on comprehension than non-captioned videos. Guillory used keyword captions in French, while this study adopted the captions of those English words which were difficult for the participants. In comparison with the vocabulary filtered, the words shown in the caption are relatively unfamiliar to them. Hence, showing the harder captions and their meanings become the key factor to assist their perception processing based on information processing theory. In a related note, Kelly [13] argues that vocabulary problems are also a significant barrier to listening

comprehension for advanced students. Thus, providing necessary captions combined with their meanings can enhance their listening comprehension.

When captions visualize the information of the foreign language which students hear in the video [1], such unfamiliar words carry sufficient information to assist students in listening comprehension, although some familiar words can only be heard and not seen. On the other hand, hiding the familiar words in the captions asks students call attention to phenomena in connected speech such as reduced forms, assimilation, elision, resyllabification and so on [14][15][16], so that watching closed-captioned films in the target language did benefits for language students [17]. Therefore, the study combined the advantages of full captions and no captions to adaptively scaffold listening training of a foreign language by using different degrees of caption-filtering.

As students in non-English speaking countries generally possess limited English vocabulary, they need greater assistance of captions when watching videos. As for the adaptive CFLLS for English listening on handheld device, the easier words are filtered and harder words are retained and shown in the videos. The study concluded that most of the students in the study positively perceived caption-filtering method. The students in the experimental group had higher perception than the students in the control group. By focusing on the perceptual processing of ears and train the comprehension of listening, the study facilitated different students with different proficiency in learning with different caption-filtering degrees. The least caption-filtering degree in the study is 220 sight words. If the leaners are younger, the teacher can consider other filtering mechanism such as target-word strategies which has also been confirmed to have better effects on vocabulary acquisition for low-achievement elementary school students [18]. The elder students, such as senior high school students or undergraduates, have learned many most frequently used English words. Consequently, the undergraduates were able to benefit from the practice of filtering more words [19]. This study confirmed that the partial hidden caption mechanism can be used in an adaptive way that presents the selected vocabulary with different degrees of difficulty based on the learning level of the students.

In the future, we plan to conduct more experiments by using longer videos, since the present study only applied short videos in 15 minutes. Therefore, it is worth conducting extended studies with a longer test period in the future. It is expected that the extended learning time may make contributions to listening proficiency, especially as there is a trend that mobile devices will become a common learning device in the near future [8][20]. It would be an interesting research issue in the future to record the frequency and the time when each student presses the “pause” button based on the time serial in the system database. Researchers can analyze the captions shown or video clip played at the time the students press the pause button.

Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 99-2511-S-011-011-MY3 and NSC 101-2511-S-011 -005 -MY3.

References

[1] Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Journal des traducteurs, 49*(1), 67-77.

[2] Markham, P. L., Peter, L., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals, 34*(5), 439-445.
[3] Seo, K. (2002). The effect of visuals on listening comprehension: A study of Japanese students' listening strategies. *International journal of listening, 15*, 57–81.
[4] Derwing, T., & Munro, M. J. (2001). What speaking rates do non-native listeners prefer? *Applied linguistics, 22*(3), 324-337.
[5] Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(03), 191-210.
[6] 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.
[7] Song, Y., & Fox, R. (2008). Using PDA for undergraduate student incidental vocabulary testing. *ReCALL, 20*(03), 290-314.
[8] Peng, H. Y., Chuang, P. Y., Hwang, G. J., Chu, H. C., Wu, T. T., & Huang, S. X. (2009). Ubiquitous performance-support system as Mindtool: A case study of instructional decision making and learning assistant. *Educational Technology & Society, 12*(1), 107-120.
[9] Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review, 63*(2), 81.
[10] Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: LEA.
[11] Vaurio, R. G. (2004). The Structure and Function of Semantic Memory in Children with Attention-Deficit/Hyperactivity Disorder. Unpublished doctoral dissertation, University of Texas, Austin.
[12] Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *Calico Journal, 15*(1-3), 89-108.
[13] Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language students. *International Review of Applied Linguistics in Language Teaching, 29*(2), 135-149.
[14] Field, J. (2003). Promoting perception: lexical segmentation in second language listening. *ELT Journal, 57*, 325-334.
[15] Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning, 16*(5), 413-425.
[16] Mitterer, H., & McQueen, J. M. (2009). *Foreign Subtitles Help but Native-Language Subtitles Harm Foreign Speech Perception*. *PLoS ONE, 4*(11), e7785. doi: 10.1371/journal.pone.0007785
[17] Stewart, M. A., & Pertusa, I. (2004). Gains to language students from viewing target language closed-captioned films. *Foreign Language Annals, 37*(3), 438-442.
[18] Hsu, C.-K., Hwang, G.-J., Chang, Y.-T., & Chang, C.-K. (In press). Effects of Video Caption Modes on English Listening Comprehension and Vocabulary Acquisition Using Handheld Devices. *Educational Technology & Society*.
[19] Hsu, C.-K., & Chang, C.-K. (2010). *Effects of Automatic Hidden Caption Classification on a Content-based Computer-Assisted Language Learning System for Foreign Language Listening*. Paper presented at the ICCE2010, Putrajaya, Malaysia.
[20] Norris, C., Hossain, A., & Soloway,, E. (2011). Using Smartphones as Essential Tools for Learning- A Call to Place Schools on the Right Side of the 21st Century. *Educational Technology & Society, 51*(3), 18-25.