An Automatic Caption Filtering and Partial Hiding Approach to Improving the English Listening Comprehension of EFL Students

Ching-Kun Hsu, Gwo-Jen Hwang and Chih-Kai Chang[*]

Department of Technology Application and Human Resource Development, National Taiwan Normal University, No. 162, Sec. 1, Heping E. Rd., Taipei City 10610, Taiwan // Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, No. 43, Sec. 4, Keelung Rd., Taipei, 106, Taiwan // Department of Information and Learning Technology, National University of Tainan, No. 33, Sec. 2, Shu-Lin St. Tainan, 700, Taiwan // ckhsu@ntnu.edu.tw // gjhwang.academic@gmail.com // chihkai@mail.nutn.edu.tw
[*]Corresponding author

ABSTRACT

Fostering the listening comprehension of English as Foreign Language (EFL) learners has been recognized as an important and challenging issue. Videos have been used as one of the English listening learning resources; however, without effective learning supports, EFL students are likely to encounter difficulties in comprehending the video content, leading to frustration and reducing learning interest. In this study, a learning system with an automatic caption filtering and partial hiding mechanism was developed to improve the English listening comprehension of EFL students. An experiment was conducted to evaluate the effects of the proposed approach on students' learning achievements and perceptions. The participants were 76 freshmen from two classes of a non-language-related department of a university. Each class contained 38 students. The two classes of students were situated in the proposed learning context and the conventional technology-enhanced learning context with a counterbalanced experimental design. The experimental results verified that the proposed caption filtering approach effectively improved the listening comprehension of the students in comparison with the conventional approach which provides full captions and an e-dictionary. Moreover, from the data collected by the eye movement tracking device, it was found that more than 76% of the students relied on the captions during the learning activity, reflecting the importance of developing effective caption filtering mechanisms for supporting EFL learners.

Keywords

Computer-assisted language learning, Caption filtering, English as foreign language, English listening comprehension, Eye movement tracking device

Introduction

Learning English as a foreign language has been heavily emphasized in Asia owing to its predominance in international communications. Among listening, speaking, reading and writing, listening has been recognized as the most essential aspect of language development (Chung, 1999; Liu, Chen, & Chang, 2009). Due to increasing needs, it has become an important and challenging issue to develop innovative and effective approaches to improving EFL students' English competence. In recent years, the advancement and popularity of computer, communication and multimedia technologies has provided an effective way to facilitate English listening comprehension training (Vanderplank, 2010). Various web-based or computer-assisted learning systems have been developed for conducting English listening activities (Chapelle, 2009; Liu & Chen, 2007; Liu, Liu, & Hwang, 2011).

Among various English learning sources, videos are the most popular form for training listening comprehension. Researchers have indicated that simultaneous audial and visual input can benefit foreign language learners (Seo, 2002); therefore, many studies related to EFL learning mainly employ videos as learning materials rather than audios or texts (Chapple & Curtis, 2000; Vanderplank, 2010; Williams & Thorne, 2000). Chapple and Curtis's (2000) have further reported that videos not only help learners improve their listening comprehensions, but also promote their confidence in speaking English, in particular, for those who are not native English speakers, such as EFL students.

Furthermore, to assist EFL students in comprehending the learning content, most English learning systems provide subtitles or captions on videos. Subtitles are the on-screen text in the students' native language combined with a second language soundtrack in the video, while captions are the on-screen text in the same language as the soundtrack (Markham, Peter, & McCarthy, 2001; Pujolà, 2002). In this study, the term subtitles refers to on-screen

Chinese text combined with an English soundtrack, while captions refers to on-screen English text combined with an English soundtrack. There are many advantages of providing captions when using videos as foreign language listening training materials (Yang, Huang, Tsai, Chung, & Wu, 2009). For example, Garza (1991) indicated that the use of captions could bridge the gap between the students' competence in reading and listening; Chung's (1996) study reported that videos with captions helped students associate the spoken and written forms of words more easily and quickly than videos without captions. In the meantime, many studies have shown that subtitles can enhance the reading, vocabulary, and listening competences of the students who learn with videos (Danan, 1992; Markham et al., 2001; Hayati & Mohmedi, 2011; Danan, 2010).

On the other hand, researchers have reported that relying heavily on subtitles when watching audio-visual materials is not conducive to improving listening proficiency (Latifi, Mobalegh, & Mohammadi, 2011). Some researchers have further indicated that videos without subtitles or captions are more beneficial as they induce students to pay attention to various pronunciation features, such as reduced forms, assimilation, elision, and re-syllabification (Hulstijn, 2003; Field, 2003; Vandergrift, 2007). In sum, although captions and subtitles could be helpful to learners in comprehending learning materials, it is important to provide learning supports that meet individuals' knowledge level at the right time to avoid possible negative effects. Miller (1956) has noted that people's capacity for processing information is limited, implying the need of providing caption-filtering mechanisms in developing learning system for English listening comprehension by taking into account what the learners need based on their knowledge levels. On the other hand, Mayer and Moreno (2003) have indicated the importance of presenting relevant learning materials (e.g., texts, videos, images) at closer space and time to improve learners' comprehensions; therefore, it is interesting to investigate the helpfulness of filtered captions by analyzing students' fixation status during the video watching process using eye tracking devices.

In this study, an automatic caption filtering and partial hiding mechanism is proposed to cope with this problem. A learning system is developed based on the proposed mechanism to assist EFL students in improving their English listening comprehension. Moreover, an experiment has been conducted to evaluate the performance of the proposed approach with several measuring tools, including achievement tests, questionnaire surveys and an eye-tracking machine.

Literature review

Captions and listening comprehension

Researchers have indicated that simultaneous audial and visual input can benefit foreign language learners (Seo, 2002); therefore, many studies related to language learning mainly employ videos as the learning materials (Chapple & Curtis, 2000; Vanderplank, 2010; Williams & Thorne, 2000). On the other hand, educators have noted that foreign language learners usually require vocabulary support when learning with video materials (Chung & Huang, 1998), implying the need to provide captions to help students understand the materials. Guillory (1998) showed that video-embedded keywords or full-text captions could benefit students more in terms of comprehending the learning content than non-captioned videos.

Several studies have also reported consistent findings related to the effectiveness of using captions in language learning (Chai & Erlam, 2008). For example, Winke, Gass, and Sydorenko (2010) indicated that captions are able to increase learners' attention and improve their learning achievements by linking with their previous knowledge.

The effect of captions or subtitles on video-based learning is highly related to the Information Processing Theory (IPT) (Miller, 1956), which refers to how humans process information via their short-term memory and store the processed results in their long-term memory. Short-term memory is limited, implying the importance of providing effective prompts to help students pay attention to important and critical information (Card, Moran, & Newell, 1983). Wang and Liu (2011) indicated that, with proper captions, students could improve their English learning interest and efficiency, showing the potential of filtered captions in language learning.

Eye-tracking and the noticing hypothesis

Attention and noticing have been recognized as important factors that affect students' learning performance in second language acquisition (Izumi, 2002; Li & Iribe, 2010; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Schmidt, 2001). The advancement of eye tracking technology provides an effective way to investigate the noticing behaviors of students by recording and analyzing their eye focusing tracks on learning materials (Scheiter & van Gog, 2009; van Gog & Scheiter, 2010). Researchers have indicated that eye focusing tracks are related to cognitive processing (Rayner, 1998; West, Carlson & Cohen, 2007); moreover, analyzing eye fixation data is an effective means of determining attentive selections (e.g., Deubel & Schneider, 1996; Henderson & Hollingworth, 1998; Hoffman & Subramaniam, 1995; Kowler et al., 1995; Irwin & Gordon, 1998).

In the past decade, eye tracking equipment has been used in various studies for analyzing the learning attention and cognitive processes of students (Jacob & Karn, 2003; Li & Iribe, 2010; Ozcelik, Arslan-Ari, & Cagiltay, 2010). For example, Ozcelik, Arslan-Ari and Cagiltay (2010) analyzed eye movement tracking data and found that when students watched and listened to learning materials, proper prompting messages that guided them to pay attention to useful or important learning information helped improve their learning effectiveness and efficiency. The study of Loboda and Brusilovsky (2010) further showed the effectiveness of adaptive visualization, that is, the provision of a clear visual metaphor for concepts or comprehension by selectively showing supportive information for important content while hiding other information. The findings from these previous studies have not only confirmed the eye-mind assumption of reading proposed by Just and Carpenter (1980), but have also shown the usefulness of using eye movement equipment in investigating the learning behaviors and cognitive processes of students (Eberhard et al., 1995; Griffin, 2001; Griffin & Bock, 2000; Tanenhaus et al., 1995).

Development of a learning system with automatic caption filtering

In this study, a learning system with an automatic caption filtering mechanism is proposed, as shown in Figure 1. The system was developed with Visual Basic Studio 2010 programming language and Microsoft Access 2010 database. The video playing functions provided by the learning system included "play" and "pause." When the learners press the "pause" button, the vocabulary included in the previous classes they had attended was hidden, while that which was new to them was displayed based on their learning profiles. This software is installed in a laptop for broadcasting movies. The captions of the movies are filtered before showing in the caption area of this application software when the students pressed the button of "Pause" function, and hiding the easier words from the original captions.
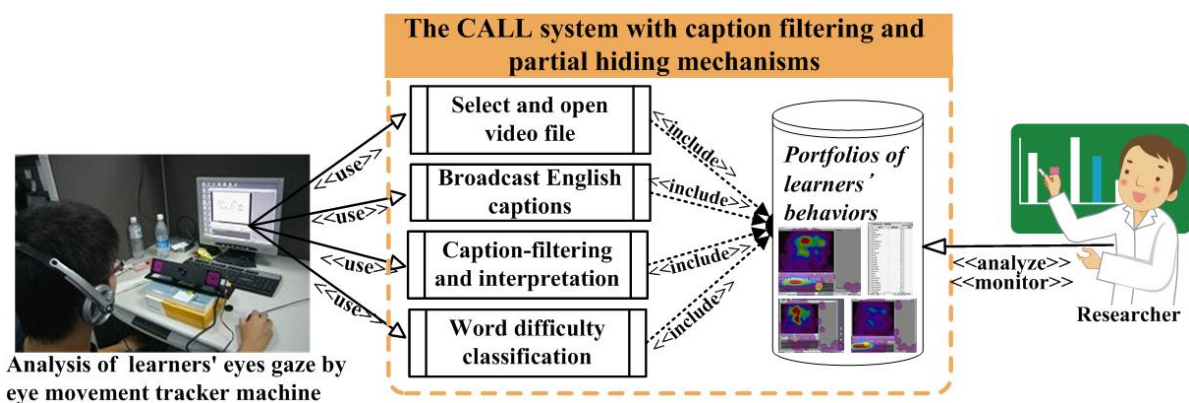

Figure 1. Architectural diagram of the automatic caption-filtering system

Figure 2 shows the learning scenario of using the developed learning system (left) and a sample figure of eye focus distributions (right). The bluish color denotes the area to which the user pays less attention, while the reddish color represents where the user gazes more.
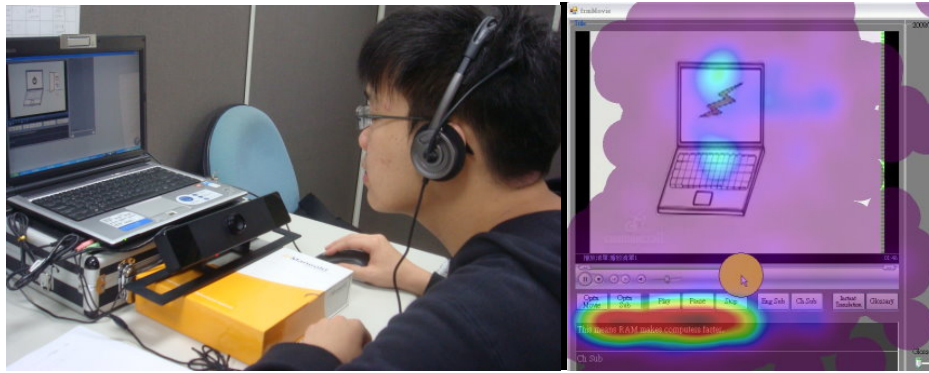
Figure 2. Experimental photos and a sample of the hot area focused mainly on the captions

Research questions

To evaluate the effectiveness of the automatic caption filtering and partial hiding mechanism proposed in this study, an experiment was conducted in the National University of Tainan, Taiwan. The students' learning performance and learning behaviors were analyzed using a Computer Usability Satisfaction Questionnaire (CUSQ), two tests, and an eye movement tracking device to investigate the following research questions.
1. What do the students pay attention to during the learning process?
2. Does the proposed approach enhance the listening comprehension of EFL students in comparison with the conventional computer-assisted language learning approach?
3. What do the students feel about the proposed approach and the conventional computer-assisted language learning approach in terms of the usability, ease of use, and satisfaction with the captions provided by the two approaches?

Experimental design

An experiment was conducted to evaluate the effectiveness of the caption filtering mechanism. In this experiment, the video soundtrack and the corresponding captions were in English; that is, the participants were prevented from watching subtitles in their first language during the video play; in the meantime English captions were displayed at the bottom of the screen.

Participants

In this study, the participants were two classes of freshmen from a non-language-related department of a university in Taiwan. Each class consisted of 38 students. The participants' average age was eighteen. All of them had learned the 2,000 most frequently used words as defined by the Ministry of Education of Taiwan, and had passed the primary level of a national English proficiency test.

Learning materials

Two videos of the same language level and length were selected by two experts who had more than five years' experience of teaching English. To avoid interference between the two videos due to familiarity after watching one of them, the two videos represented different themes aimed at introducing computer science. One introduced PC hardware components, while the other was related to web searching functions. Both videos used colloquial style English to convey the basic concepts as real life characteristics. The two videos were used in the learning activities based on the original curriculum design of the sample course. The aim of the course was to introduce the structure, functionalities and applications of computers and networks. That is, the learning activity as well as the learning content reflected the teaching reality of that course. For example, the participants of this study had acquired the 2,000

most frequently used words defined by the Ministry of Education in Taiwan; therefore, in Figure 3 which is the snapshots of the learning system, those words were hidden, so that they could pay more attention to the target vocabulary presented in the videos. When the students pressed the button of "Pause," all the translations of the other harder words are shown in the instant translation area. In the following, we shall call the 2,000 most frequently used words "familiar words," while referring to the others as "unfamiliar words."
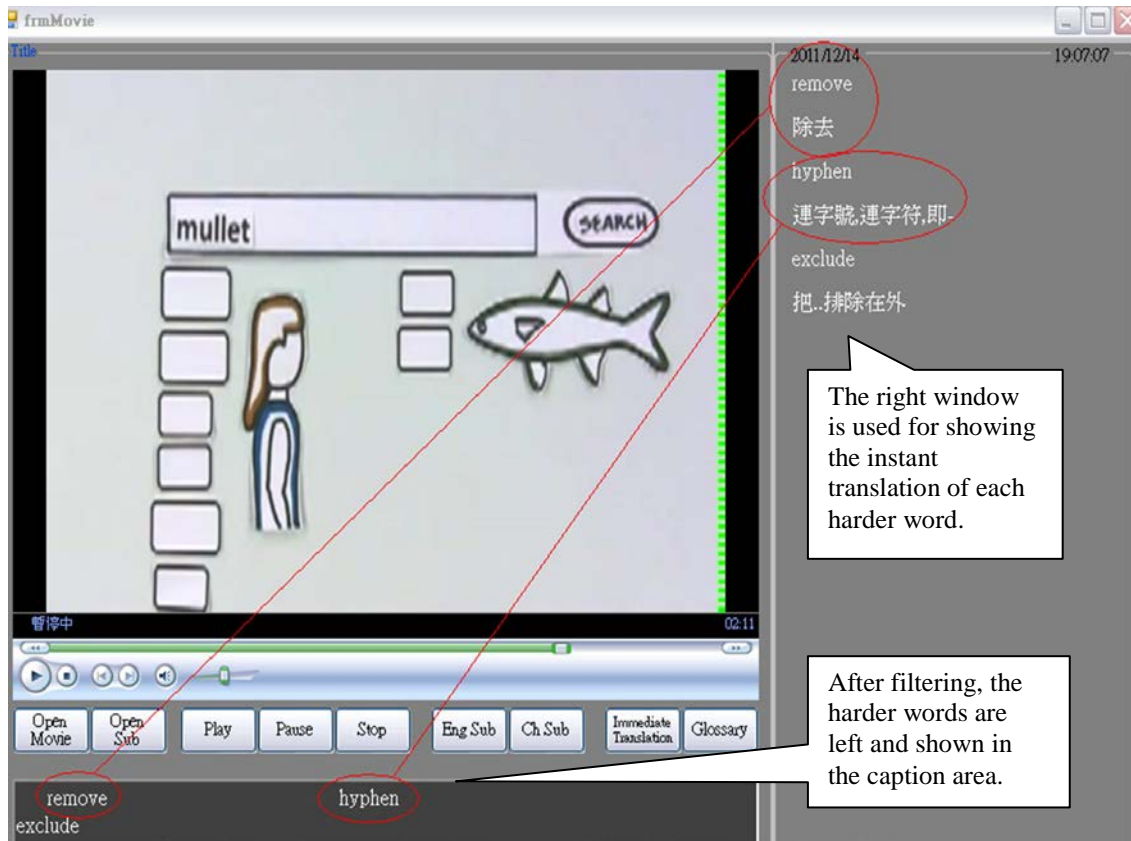

Figure 3. Hiding the easier words and instantly translating the harder words after pausing

Measuring tools

For each video, a test was developed by three experienced English teachers, and was pronounced by two native English speakers who had more than five years of English teaching experience. Each of the tests included two parts: that is, an auditory test and a listening comprehension test. In the auditory test, the questions and choices for testing the video content were presented via audios; in the listening comprehension test, the questions and choices for testing the video content were presented in texts. Both the test parts contained four multiple-choice items with a perfect score of 100. To examine the reliability of the tests, the study conducted a preliminary examination on 14 students in the same department of that university. In terms of the two auditory tests, the Cronbach's alpha value of reliability statistics is .928, and the correlation coefficient value of Spearman's rho is .866 from the equivalent-forms method of pretesting. As for the two listening comprehension tests, the Cronbach's alpha value of reliability statistics is .980, and the correlation coefficient value of Spearman's rho is .962. From the statistical results, it is inferred that the two tests can be viewed as equivalent.

The CUSQ was developed by Lewis (1995). It consists of three dimensions (i.e., "ease of use," "usability" and "satisfaction") with a total of 19 questions using a seven-point Likert scale. The Cronbach's alpha values of the three dimensions were 0.75, 0.71 and 0.79, respectively.

Experimental procedure

The treatment of this study is an experimental counterbalanced design, which is one of the formal educational research methods (Gall, Gall & Borg, 2007), and is also used in language learning studies (e.g., Freiermuth, 2001; Snellings, van Gelderen & de Glopper, 2004). The choice of this approach is to reduce the influence of ordering when comparing the effectiveness of two treatments, while allowing the students in both groups to experience the two approaches. This design meets the request from the teacher that all of participants have the same opportunity to receive the new learning approach. As shown in Table 1, the participants in both classes experience both the experimental group treatment and the comparison group treatment. That is, 38 students in class A learned with the experimental group treatment first and then experienced the comparison group treatment thereafter. On the other hand, the other 38 students in class B learned with the comparison group treatment first and then experienced the experimental group treatment later.

Table 1. Counterbalanced experimental design

Order	N	First	Latter
Class A	38	X_1O_1	X_2O_2
Class B	38	X_2O_2	X_1O_1

Note. X_1: Comparison group treatment; X_2: Experimental group treatment; O_1: Post-test of the auditory test and listening comprehension test; O_2: Post-test of the auditory test and listening comprehension test.

During the learning process, the students in the comparison group learned with the conventional technology-enhanced learning approach; that is, they were instructed to push the pause button of the broadcast software whenever they could not comprehend what they had heard, allowing them to consult an e-dictionary embedded in the system for finding the Chinese translation of the unfamiliar words. When the students paused the videos and pointed at an unfamiliar word in the caption, possible Chinese translations of the word were displayed.

On the other hand, the students in the experimental group learned with the proposed system, where the familiar vocabulary was automatically filtered, while the unfamiliar vocabulary (e.g., the new vocabulary to be learned in the present unit) and the corresponding Chinese translations were displayed when the video player was paused.
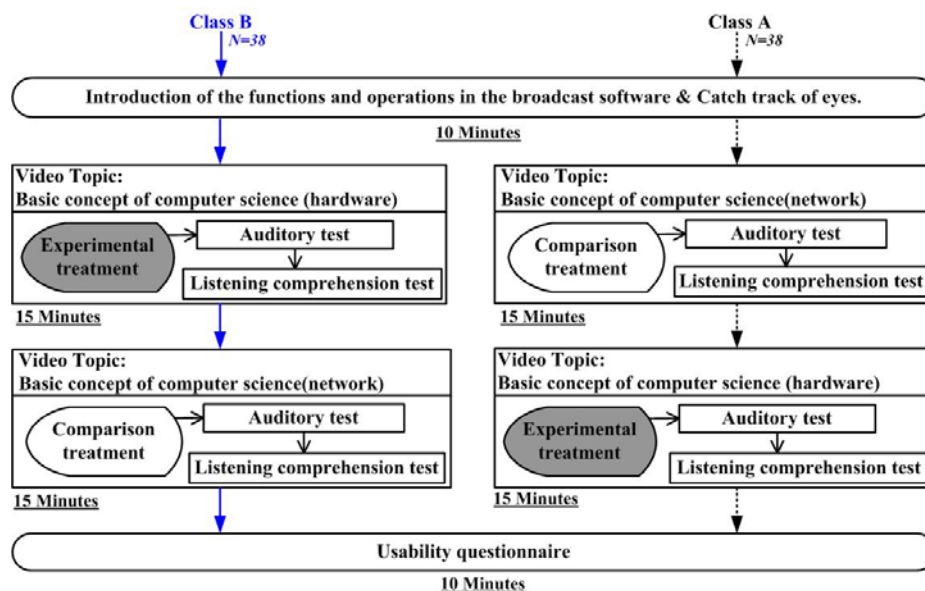

Figure 4. The experimental procedure in the counterbalanced design

While watching the videos, all of the students were observed by an eye movement tracking device. The eye-tracking system used was MangoldVision, developed by Mangold International GmbH. The portable MangoldVision Eyetracker comes in a case the size of a small briefcase (25x40x10 cm).
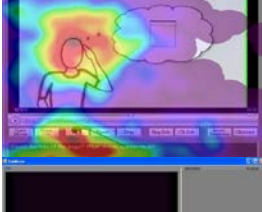
Figure 4 shows the duration of each treatment stage and the distribution of the participants. The participants took an auditory test (15 minutes), and then took a listening comprehension test (15 minutes) after watching the video with each treatment. The scores were collected for the paired-samples t-test because all of the students participated in the counterbalanced design of the experiment to prevent the impact of sequence between the two treatments. After the counterbalanced treatments, the students filled in the usability questionnaire to give feedback regarding their opinions. Finally, the results were analyzed for future reference.

Results and discussion

Students' eye fixation analysis

From the analysis of the eye focusing tracks, it was found that most of the students in both groups focused on the captions during the video play. The eye-focusing positions were categorized into four types based on the data collected from the eye tracking device, as shown in Table 2. It was found that more than 76% of the participants (N = 59) paid much attention to the captions when watching the videos; that is, they tended to rely on the captions to comprehend the learning content, indicating the importance of developing more effective caption mechanisms for EFL students.

Table 2. Distribution of the visual focus of the participants

Stare position	Example figures	Class A			Class B		
		X1	X2	Total	X2	X1	Total
Mainly on the captions		39.47% (N = 30)	38.16% (N = 29)	77.63% (N = 59)	38.16% (N = 29)	38.16% (N = 29)	76.32% (N = 58)
Mainly on the video		5.26% (N = 4)	3.95% (N = 3)	9.21% (N = 7)	3.95% (N = 3)	9.21% (N = 7)	13.16% (N = 10)
Both captions and video		5.26% (N = 4)	7.89% (N = 6)	13.16% (N = 10)	7.89% (N = 6)	1.32% (N = 1)	9.21% (N = 7)
Distracted from captions and video		0% (N = 0)	0% (N = 0)	0% (N = 0)	0% (N = 0)	1.32% (N = 1)	1.32% (N = 1)

Note. % is of total number (i.e., 76); X1: Comparison group treatment; X2: Experimental group treatment.

Results of the auditory tests and listening comprehension tests

Table 3 shows the statistical results of the scores of the two groups. It was found that, for the students in both groups, the auditory test scores after experiencing the experimental group treatment were always higher than those after experiencing the comparison group treatment. The difference between the means of the two test scores was more than 7 points. Moreover, for both groups, the listening comprehension test scores after receiving the experimental

group treatment were higher than those after receiving the comparison group treatment. The difference between the means of the two test scores was more than 5 points.

The t-test result further showed that the two treatments differed significantly in the listening outcomes ($t = 2.66^*$; $p = .01$), but not in the comprehension degrees ($t = 1.51$; $p = .13$). From the t-test result, it is concluded that the first research hypothesis is validated; that is, the learning system with the caption filtering approach is more effective than the conventional approach in terms of facilitating English listening comprehension ability.

Table 3. Paired-samples t-test of the two treatments in the auditory tests and listening comprehension tests

Post-test	Treatments	Samples	N	Mean	SD	t
Auditory Tests	X1	Class A	38	49.34	29.34	
		Class B	38	45.39	24.56	
		ALL	**76**	47.37	26.95	
	X2	Class B	38	57.89	27.33	
		Class A	38	55.26	30.84	
		ALL	**76**	56.58	28.97	
	Paired sample t test (X2_ALL-X1_ALL)		76		30.23	**2.66**[*]
Listening comprehension Tests	X1	Class A	38	57.24	25.27	
		Class B	38	58.55	24.16	
		ALL	**76**	57.89	24.57	
	X2	Class B	38	62.50	23.79	
		Class A	38	63.82	25.13	
		ALL	**76**	63.16	24.32	
	Paired sample t test (Sum of X2 - Sum of X1)		76		30.36	1.51

Note. $^*p < .05$. X1: Comparison group treatment; X2: Experimental group treatment.

To understand more deeply the relationship between the learning outcomes and eye-gazing, the study selected those students whose performance was in the top 27% of the listening and listening comprehension test outcomes as the high-achievement group, and those in the bottom 27% as the low-achievement group (Ebel, 1972). The fixation distributions of the students in different achievement groups were compared.

Table 4 shows the statistical data of four categories of fixation distributions (i.e., paying attention to captions, videos, both, or none) for the low- and high-achievement students. From the Pearson Chi-Square analysis, it was found that different achievement students had similar fixation distributions. For example, in terms of the comparison group treatment students' behaviors during the auditory test, 75% of the high-achievement students and 87.5% of the low-achievement students mainly focused their eyes on the captions during the learning activity. From the results of the Chi square tests, the dependent degrees on the captions are similar (X2 = 0.65; $p > 0.05$), indicating that most students, no matter whether they were high-achievement or low-achievement, needed the assistance of the captions to comprehend the learning materials.

Table 4. The fixation distribution of the high- and low-achievement students

Tests		High-achievement students				Low-achievement students				Chi-Square
		Caption	Video	Both	None	Caption	Video	Both	None	
Auditory tests	C	**75.00%**	15.00%	10.00%	0%	**87.50%**	8.33%	4.17%	0%	0.65
	E	**68.18%**	22.73%	9.09%	0%	**76.19%**	4.76%	19.05%	0%	3.47
Listening comprehension tests	C	**83.33%**	11.11%	5.56%	0%	**88.89%**	**0%**	11.11%	0%	2.36
	E	**65.22%**	21.74%	13.04%	0%	**70.83%**	12.50%	12.50%	4.17%	1.54

Note. C: Comparison group treatment; E: Experimental group treatment.

In comparison with the comparison group treatment, it was found that, on average, the students who learned with the experimental group treatment, no matter whether they were high- or low-achievement, spent less time on reading the captions. For example, in the listening comprehension tests, only 65.22% of the high-achievement students who

learned with the experimental group treatment spent most of the time reading the captions, while 83.33% of those who learned with the comparison group treatment kept focusing on the captions. Moreover, only 70.83% of the low-achievement students who learned with the experimental group treatment spent most of the time reading the captions, while 88.89% of those who learned with the comparison group treatment comparison group treatment kept focusing on the captions. This finding implies that filtered captions were able to ease the load of the students during the learning process.

This study further analyzed the learning performance of the students who looked at the captions, to see if their learning behaviors were helpful in improving their listening or comprehending performance. It was found that most of the students receiving the experimental group treatment revealed better performance than those who learned with the comparison treatment. That is, the students had significantly better performance when reading the selected words in the captions than when reading the full captions in terms of their listening comprehension. Consequently, it can be inferred that the automatic caption-filtering system was helpful to the students.

Results of the usability evaluation

Table 5 shows the statistical results of the students' ratings on the questionnaire items of the "ease of use," "usability" and "satisfaction" dimensions.

Table 5. Adaptive caption-filtering system usability questionnaire (Revised from CUSQ)

Perceived Ease-Of-Use	mean	SD
1. It was simple to use this system.	5.24	1.33
2. I can effectively complete my work using this system.	4.63	1.21
3. I am able to complete my work quickly using this system.	4.87	1.02
4. It was easy to learn to use this system.	5.78	1.07
5. Whenever I make a mistake using the system, I recover easily and quickly.	4.28	0.99
6. It is easy to find the information I need.	4.80	1.23
7. The information provided by the system is easy to understand.	5.50	1.05
Average	**5.01**	**1.23**
Perceived Usability		
8. I am able to efficiently complete my work using this system.	4.74	1.02
9. I believe I became productive quickly using this system.	4.71	1.09
10. The system gives error messages that clearly tell me how to fix problems.	4.17	0.97
11. The information (such as online help, on-screen messages, and other documentation) provided with this system is clear.	4.59	1.07
12. The information is effective in helping me complete the tasks and scenarios.	5.20	0.95
13. The organization of information on the system screens is clear.	5.09	1.07
14. This system has all the functions and capabilities I expect it to have.	4.84	0.95
Average	**4.76**	**1.06**
Perceived Satisfaction		
15. Overall, I am satisfied with how easy it is to use this system.	4.92	1.10
16. I feel comfortable using this system.	4.76	1.04
17. The interface of this system is pleasant.	4.99	1.08
18. I like using the interface of this system.	4.83	1.00
19. Overall, I am satisfied with this system.	5.34	0.95
Average	**4.97**	**1.05**

In terms of the "ease of use" dimension, the average rating was quite positive (mean = 5.01; SD = 1.23). Among the seven items in this dimension, one item, "It was easy to learn to use this system," reached 5.78 on average, indicating that the students perceived that the English listening competence training system with the adaptive caption-filtering mechanism was easy to use.

As for the dimension of usability, the average rating was 4.76. In this dimension, the average rating of the item "The information is effective in helping me complete the tasks and scenarios" is 5.2, implying that most students confirmed the usefulness of the proposed approach.

With regard to the dimension of perceived satisfaction, the average score was 4.97; moreover, the average rating of the item "Overall, I am satisfied with this system" was 5.34, revealing that most of the students were highly satisfied with the learning approach.

Figure 5 summarizes the results of the 76 questionnaires. According to this figure, 65% of the students indicated that the proposed system was usable and easy to use; more than 62% of the participants were satisfied with the learning approach. Moreover, no student selected the 7th category (i.e., strongly disagree) in answering the questionnaire items. To sum up, the caption filtering mechanism enabled the students to concentrate on the words that were difficult for them, which made the learning process more effective; furthermore, the provision of instant translations of the target words helped the students comprehend the new vocabulary in a more efficient way. As a consequence, the students gave quite positive feedback in terms of ease of use, usability and satisfaction.
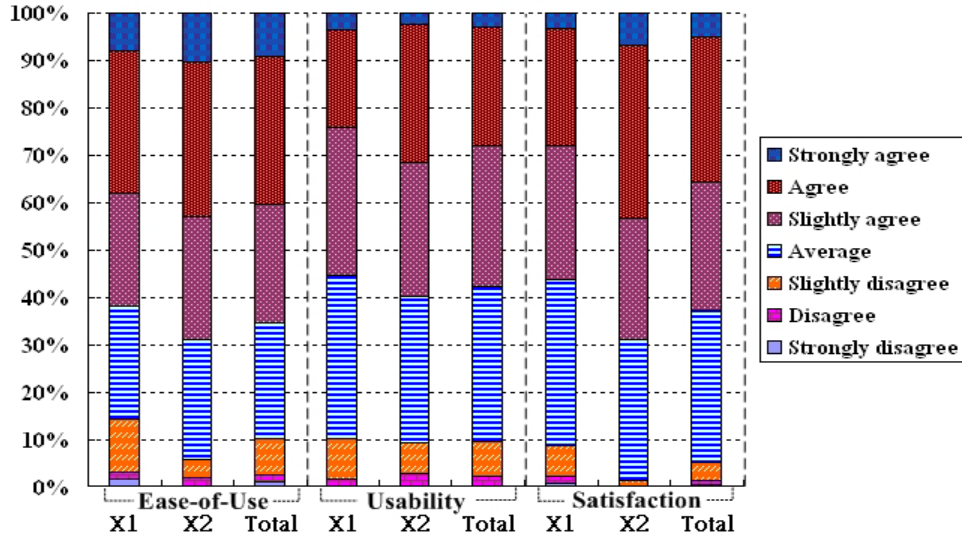

Figure 5. Results of the usability questionnaire study

Conclusions and implications

When learning a second or foreign language, students often encounter difficulties in watching or listening to those learning materials which involve a rapid rate of presentation and unknown vocabulary; moreover, redundant information (e.g., subtitles) can also confuse learners (Smidt & Hegelheimer, 2004; Kellerman, 1990). In this paper, an English listening comprehension training system with an automatic caption filtering approach is presented. To evaluate the performance of the proposed approach, the study compared the effects of the proposed learning system with the caption filtering mechanism and the conventional technology-enhanced learning approach with an e-dictionary on the auditory test and listening comprehension test. Moreover, an eye movement tracking device was used to collect the students' eye-fixation data during the learning process.

For the first research question, the eye-tracking data analysis showed that 78% of the students in class A, and 76% of the students in class B focus their look at the captions, revealing the importance of providing effective captions in English listening training programs. This finding conforms to some previous studies reporting that participants' eyes tended to fixate more on the text than on the illustrations; moreover, text seems to provide more helpful information than pictures (Liu & Chuang, 2010). In the meantime, the eye-tracking data also showed that, when learning with the videos with filtered captions, the students were able to focus more on the target words (new vocabulary) in comparison with the full captions, which could be the reason why the experimental group treatment benefited the students more in terms of their listening comprehension. This finding conforms to the study of Danan (2004), who indicated that captions visualize the information of the foreign language that students hear in videos, and showing only unfamiliar words can usually provide sufficient information to assist students with their listening comprehension. Several studies have also reported that hiding familiar words in captions enables students to pay

more attention to the target words, which is of great benefit to the development of their listening ability (Hulstijn, 2003; Field, 2003; Mitterer & McQueen, 2009; Stewart & Pertusa, 2004).

For the second question, it was found that the students who learned with the proposed caption-filtering approach showed significant better learning achievements in the auditory tests and listening comprehension tests than those who learned with the conventional technology-enhanced learning approach.

Moreover, for the third research question, the questionnaire results showed that most of the students highly accepted the proposed approach in terms of "ease of use," "usability" and "satisfaction." The average ratings of the three dimensions were 5.01, 4.76, and 4.97, respectively.

In the future, we plan to conduct several continuous studies. First, a large-scale experiment will be conducted using videos of various lengths and with more vocabulary. Second, we plan to employ the proposed approach in other English training programs in high schools. Third, in addition to applying the eye-tracking machine to detect and record the learning behaviors of students, we plan to record and analyze the time and frequency with which students press the "pause" button, such that a more effective caption mechanism might be proposed based on the analysis results. Fourth, we plan to expand the learning system by providing more facilities to help the students comprehend the learning content, as well as enhancing their foreign language listening proficiency, such as some Mindtools, context-aware facilities, or discussion forums developed in some previous studies (Chu, Hwang, Tsai, & Chen, 2009; Hwang, Chu, Lin, & Tsai, 2011; Hwang, Tsai, Chu, Kinshuk, & Chen, 2012).

Acknowledgements

References

Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. *New Zealand Studies in Applied Linguistics, 14*(2), 25-44.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human Computer Interaction Hillsdale*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chapelle, C. A. (2009). The relationship between second language acquisition theory and computer-assisted language learning. *The Modern Language Journal, 93*, 731-753.

Chapple, L., & Curtis, A. (2000). Content-based instruction in Hong Kong: Student responses to film. *System, 28*(3), 419–433.

Chu, H. C., Hwang, G. J., Tsai, C. C., & Chen, N. S. (2009). An innovative approach for promoting information exchanges and sharing in a web 2.0-based learning environment. *Interactive Learning Environments, 17*(4), 311-323.

Chung, J. M. (1996). The effects of using advance organizers and captions to introduce video in the foreign language classroom. *TESL Canada Journal, 14*(1), 61-65.

Chung, J. (1999). The effects of using video texts supported with advance organizers and captions on Chinese college students' listening comprehension: An empirical study. *Foreign Language Annals, 32*(3), 295-308.

Chung, J. M., & Huang, S. C. (1998). The effects of three aural advance organizers for video viewing in a foreign language classroom. *System, 26*(4), 553-565.

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

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

Danan, M. (2010). Dubbing projects for the language learner: A framework for integrating audiovisual translation into task-based instruction. *Computer assisted language learning, 23*(5), 441-456.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*(12), 1827-1837.

Ebel, R. L. (1972). *Essentials of Educational Measurement* (2nd ed). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*(6), 409-436.

Field, J. (2003). Promoting perception: Lexical segmentation in second language listening. *ELT Journal*, 57, 325–334.

Freiermuth, M. R. (2001). Native speakers or non-native speakers: Who has the floor? Online and face-to-face interaction in culturally mixed small groups. *Computer assisted language learning, 14*(2), 169-199.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational Research: An Introduction* (8th ed.). New York, NY: Pearson education.

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals, 24*(3), 239-258.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*(1), B1-B14.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science, 11*(4), 274–279.

Guillory, H. G. (1998). The effects of key word captions to authentic French video in foreign language instruction. *CALICO Journal, 15*(1-3), 89-108.

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology, 42*(1), 181-192.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (Vol. 11, pp. 269-293). Amsterdam, the Netherlands: Elsevier Science Ltd.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Attention, Perception, & Psychophysics, 57*(6), 787-795.

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.

Hwang, G. J., Chu, H. C., Lin, Y. S., & Tsai, C. C. (2011). A knowledge acquisition approach to developing Mindtools for organizing and sharing differentiating knowledge in a ubiquitous learning environment. *Computers & Education, 57*(1)*, *1368-1377.

Hwang, G. J., Tsai, C. C., Chu, H. C., Kinshuk, & Chen, C. Y. (2012). A context-aware ubiquitous learning approach to conducting scientific inquiry activities in a science park. *Australasian Journal of Educational Technology, 28*(5), 931-947.

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention and trans-saccadic memory. In W. X. Schneider & S. Maasen (Eds.), *Mechanisms of Visual Attention: A Cognition Neuroscience Perspective* (Vol. 5, pp. 127-155). East Sussex: Psychology Press.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis. *Studies in Second Language Acquisition, 24*(04), 541-577.

Jacob, R. J. K., & Karn, K. S. (2003). Eye-tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. R. Hyona & H. D. Radach (Eds.), *The mind's eyes: Cognitive and applied aspects of eye movements.* Oxford, UK: Elsevier Science.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-355.

Kellerman, S. (1990). Lip service: The contribution of the visual modality to speech perception and its relevance to the teaching and testing of foreign language listening comprehension. *Applied Linguistics, 11*(3), 272–280.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*(13), 1897-1916.

Latifi, M., Mobalegh, A., & Mohammadi, E. (2011). Movie subtitles and the improvement of listening comprehension ability: Does it help? *The Journal of Language Teaching and Learning (JLTL), 1*(2), 18-29.

Li, K., & Iribe, Y. (2010). Promoting learning attention with gaze tracking integrated e-learning contents. *Knowledge-Based and Intelligent Information and Engineering Systems*, 173-179.

Liu, H. C., & Chuang, H. H. (2010). An examination of cognitive processing of multimedia information based on viewers' eye movements. *Interactive Learning Environments*, 1-15.

Liu, G. Z., & Chen, S. W. (2007). A taxonomy of Internet-based technologies integrated in language curricula. *British Journal of Educational Technology*, 38(5), 934-938.

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.

Liu, G. Z., Liu, Z. H., & Hwang, G. J. (2011). Developing multi-dimensional evaluation criteria for English learning websites with university students and professors. *Computers & Education, 56*(1), 65-79.

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.

Loboda, T. D., & Brusilovsky, P. (2010). User-adaptive explanatory program visualization: evaluation and insights from eye movements. *User Modeling and User-Adapted Interaction*, 1-36.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181-210). Philadelphia, PA, USA: John Benjamins Publishing Company.

Markham, P. L., 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.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist, 38*(1), 43-52.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The psychological review, 63*, 81-97.

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

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*(1), 110-117.

Pujolà, J. T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, *14*(2), 235-262.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372 -422.

Scheiter, K., & Van Gog, T. (2009). Using eye tracking in applied research to study and stimulate the processing of information from multi representational sources. *Applied Cognitive Psychology, 23*(9), 1209-1214.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). New York, NY: Cambridge University Press.

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

Smidt, E., & Hegelheimer, V. (2004). Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use. *Computer assisted language learning, 17*(5), 517-556.

Snellings, P., Van Gelderen, A., & De Glopper, K. (2004). The effect of enhanced lexical retrieval on second language writing: A classroom experiment. *Applied Psycholinguistics, 25*(2), 175-200.

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1634.

Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95-99.

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

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching, 43* (1), 1–37.

Wang, K., & Liu, H. (2011). Language acquisition with the help of captions. *Studies in Literature and Language, 3*(3), 41-45.

West, R., Carlson, L., & Cohen, A. L. (2007). Eye movements and prospective memory: What the eyes can tell us about prospective memory. *International journal of psychophysiology, 64*(3), 269-277.

Williams, H., & Thorne, D. (2000). The value of teletext subtitling as a medium for language learning. *System, 28*(2), 217–228.

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.

Yang, J. C., Huang, Y. T., Tsai, C. C., Chung, C. I., & Wu, Y. C. (2009). An automatic multimedia content summarization system for video recommendation. *Educational Technology & Society, 12* (1), 49–61.