Mohd Ayub A. F. et al. (Eds.) (2011). Workshop Proceedings of the 19th International Conference on Computers in Education. ChiangMai, Thailand: Asia-Pacific Society for Computers in Education.

Caption Readability Improvement of Language Learning Videos on Mobile Devices

Ching-Kun HSU & Chih-Kai CHANG*
Department of Information and Leaning Technology,
National University of Tainan, Taiwan
*chihkai@mail.nutn.edu.tw

Abstract: To promote the readability of multimedia and captions shown in mobile devices, we developed a system for processing listening materials to reduce and hide easy words which are the top 2200 highest frequent usage words, for better English proficiency and learners who have mobile phones. On the basis of information processing theory, this study set different thresholds for the most number of words displayed each time for diverse mobile devices. We further evaluated whether the results of caption filtering satisfies the thresholds. Results show that more than 71% of all the captions are hidden. Moreover, the proposed system achieves the expected values of the requirements for different device screen sizes, and readability improvement is remarkable.

Keywords: caption, mobile device, readability, information processing theory.

1. Introduction

This study developed caption filtering for promoting the readability of multimedia or video in mobile devices. In Taiwan, most high school students had learned more than 2000 frequently used English vocabulary. However, their English listen comprehension level cannot match their vocabulary amount. For the purpose of this study, we hide the 2200 most frequently used words from the captions shown in students' mobile phones equipped with small screens to evaluate improvement in readability. This study is not only related to human–computer interactions but also based on information processing theory (IPT). In IPT, memory is classified into three categories: sensory, short-term, and long-term memory [1,2]. The simulation of information processing is shown in Figure 1. When external stimulus is received from the sensory organs, such as vision, hearing, olfactory, tactile sense, and so on, the initial memory is fed from the sensory registers. The majority of temporary memory is instantly forgotten [3]. Meanwhile, only a small proportion of this memory, especially attendant memory, is transmitted to short-term memory, also called working memory [4].

However, the capacity of short-term memory is limited. Its volume ranges from 5 to 9 bits; its length is often regarded as 7 bits on average [5]. Given this limitation, most of the recollections in this memory are also immediately forgotten. Compared with prior knowledge, only a small fraction of short-term memories are elaborated in detail, analyzed, and combined for conversion into long-term memories. 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.

Short-term memory stores information longer than does sensory memory, but the former, which records "at-the-moment" effects, declines easily unless the input information

is repeated and coded for integration into long-term memory. Otherwise, short-term memory cannot be retained. Scholars proposed that the language information in short-term memory is coded in the form of hearing, whereas that in long-term memory is coded in the form of language meaning [6]. Accordingly, to enhance the listening proficiency and degree of comprehension for foreign language, we hid the caption of easy words to compel learners to pay attention to aural input. In addition, the more difficult words are shown in the captions to drive learner comprehension of the full content being listened to.

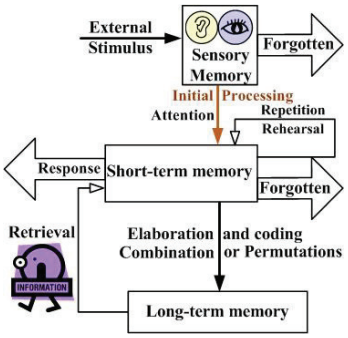

Figure 1. Information processing simulation [1]

The development of mobile devices has brought convenience to learners, providing access to information anytime, anywhere [7]. Researchers reported a tendency toward mobile learning from 2009 to 2014, which shows that mobile learning is associated with mobile devices and that the function of portable information is to carry out the transmission and storage of knowledge [8]. The application of language learning into mobile devices has also been widely studied. For example, Ogata et al. (2008) used mobile devices to help foreign students learn English outside classrooms through the performance of tasks. The result shows that learners adopted a positive learning attitude and exhibited increased interactions with teachers. Kukulska-Hulme (2006) indicated that using mobile devices for language learning results in good learning effectiveness and easy access [9]. Scholars also indicated that mobile assisted language learning (MALL) is helpful for countries with lower education levels [10].

Mobile devices are characterized by diverse sizes and interfaces; these include netbooks, PDAs, smartphones, and so on. Such diverse devices have different screen sizes, as well as distinct operating systems, network explorations, and functions. Studies showed that students can learn more quickly, remember more steadily, and achieve more objectives when they are interested in learning materials [11]. Multimedia and videos are such types of learning materials for foreign language learners. Research indicated that information and computer technology benefits language learning, and multimedia is contributory to vocabulary learning [12]. A player is an essential equipment for language learning [10]. Many different kinds of mobile devices are currently being developed. However, the elucidation of whether different styles or brands of mobile devices have corresponding roles for mobile learning is as important as device development. Examples include media players with readability options, operating systems and application software with compatibility, and so on.

The popularity of information and computer technology enables the constant accessibility of video sharing platforms, such as YouTube and websites for online movies, among others. Although accessing videos is convenient and simple, the small screens of mobile devices pose viewing and readability problems. A previous study indicated that classifying vocabulary to hide easy words by caption filtering and showing difficult words with instant translation in the players of personal computers facilitate good learning effectiveness for English listening training [13]. As for mobile learning, the same mechanism but different settings and research questions can be employed to solve the

problem of readability in mobile devices. That is, for learners who are not novices at learning English, the present study solves the display and readability restrictions encountered in small screens by hiding the easy words, such as the 2200 most frequently used words, in the captions. More important, using video to learn not only results in more exciting learning avenues, but also affords learner the freedom to choose according to their learning interests. The use of instructional DVDs in a classroom is convenient typical example. Selecting various languages to be shown in the captions is a simple approach. Previous research compared the effectiveness of captions and scripts and found that most learners used these as a learning auxiliary when they cannot comprehend video content [14]. Therefore, the auxiliary and accessibility of the captions are obvious. In terms of the broadcast functions of players, literature showed that speeding up or slowing down the videos is unhelpful or unnecessary in foreign language listening proficiency training [15].

Accordingly, the aim of this study is not to modulate the functions of the players in mobile devices, but to adjust the quantity of captions displayed in small screens to improve readability. This approach is adopted because of the increasing popularity of mobile phones, which many senior high school or college students possess. We first filtered the most frequently used words, which must be learned before graduation from high school in Taiwan. The top 2200 highest frequently used words were hidden from the caption so that learners are compelled to focus on the aural input of the words. By contrast, the more difficult words were still shown in the caption as aid in the listening comprehension of learners.

To design freeware that adaptively processes the listening materials, we initially set up the top 2200 highest frequently used words in a word bank to filter such primary words for learners graduating from high school. Previous research verified that filtering easy words for advanced learners has a positive learning effect [13]. We expected to evaluate the readability of the filtered captions on different screen sizes. To address the display and readability problem presented by long captions in the listening materials shown in mobile devices, we randomly selected 30 video samples to assess the degrees of readability improvement in different screen sizes after filtering. The research objectives for improving usability and readability in differently sized mobile devices are as follows:

1. For 4-inch mobile devices, to achieve an expected ratio of captions (which are no more than 5 words) of 98% after system processing of the listening material;
2. For 7-inch mobile devices, to achieve an expected ratio of captions (no more than seven words) of 99% after system processing of the listening material.
3. For 10-inch mobile devices, to achieve an expected ratio of captions (with no more than nine words) of 99.99% after system processing of the listening material.

2. Methods

2.1 System design

The font size of the captions is no smaller than 1/25th of the full screen; that is, the captions should correspond to the proportion of the frame, making them easy to read. As for the display of the caption in the player embedded in mobile devices with different screen sizes, however, few studies have explored approaches to improving the readability of the display and fitting the number of words on a given screen size. Consequently, the presence of many multimedia instructional materials restricted by small screen sizes has caused difficulties for learners to engage in mobile learning. Moreover, the large quantity of messages in a video clip may result in overloading of the learners' short-term memory.

To solve the above-mentioned problems, prevent learners from distractions stemming from simultaneous attention to images and captions, and rectify limited

comprehension of listening training materials, the present study focused on improving the readability of video captions shown on mobile devices. We achieved these by caption filtering, as well as supporting listening comprehension. Most important, caption filtering must not present learning obstacles but should have positive effects on listening once readability on mobile devices 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 on mobile devices to aid listening comprehension.

Python programming language was used to develop the caption filtering program. The program is called the caption filtering system of listening materials for adaptive MALL. The system framework is shown in Figure 2. The procedure 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 show the more difficult or seldomly seen words. Moreover, two databases are employed. One is the work bank, which stores the 2200 most frequently used words from the corpus. The other database stores the captions after merging and those after filtering. The original caption or subtitle files can be stored as srt files, consisting of time sequence, time interval, and captions. After the caption filtering process, the remaining captions are subjected to statistical computation to determine whether they satisfy the expectation for the adaptation to different mobile device sizes. 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. The filtered captions can be embedded into the video for broadcasting on mobile device, enhancement of the readability of multimedia shown in the small screen, and assistance in the listening comprehension of learners.

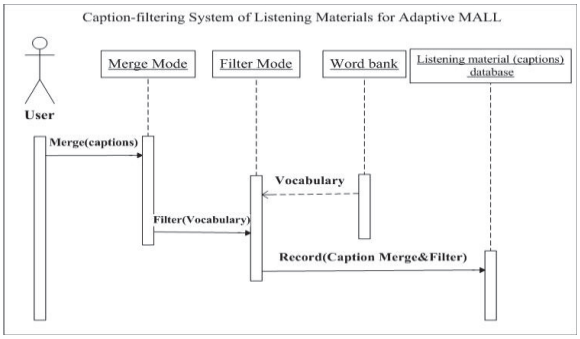

Figure 2. System framework.

2.2 Procedure

The EFL learners who passed the primary level of GEPT certification in Taiwan are considered familiar with the hidden words. Most learners, such as students who graduated from high school, need the support of advanced vocabulary or unfamiliar words. The system leaves advanced or unfamiliar words in the captions to facilitate listening comprehension. In the pilot study, the number of words in the original caption of each video clip was computed. In addition, the percentage of words filtered and hidden from captions and the percentage of sentences that were not adjusted were calculated. The average number of words left in a sentence after filtering was also computed. Finally, whether the captions satisfy the expected value for each size of mobile device was evaluated using the 30 DVD samples.

2.3 Listening materials

Thirty random Discovery DVDs were borrowed from a library in a national university in Taiwan as the samples materials. The system examines the videos to determine the lengths of the video and sentence in the caption, what proportion of captions requires merging into a line or filtering, and how long each filtered sentence in the caption is. The total words, as well as the average number of words in each sentence and in each filtered sentence, were statistically computed. All the related statistical results and the above-mentioned information were calculated. This approach enables further evaluation of whether the modulated results are adaptive to broadcast in mobile devices. The readability of multimedia in mobile devices is promoted as well.

The information on the videos was collected by the system. The results for the first threshold (the five-word standard set for promoting readability in 4-inch mobile devices) show that originally, 78.44% of the captions have longer than five words on average in the 30 videos. After hiding the easiest 2200 words from the captions, only 1.14% of the captions have longer than five words. The results for the second threshold (the seven words set for improving readability in 7-inch mobile devices) show that originally, 55.40% of the captions have longer than seven words on average in the 30 videos. After filtering, only 0.02% of the captions have longer than seven words. The results for the third threshold (the nine-word set for enhancing readability in 10-inch mobile devices) show that originally, 31.39% of the captions have longer than nine words. Furthermore, the filtered sentence approximated that 0% of captions have longer than nine words. Figure 3 shows the example of the captions before and after filtering in the 4-inch mobile phone. The effect on readability after filtering the most frequently used words is remarkable. The visual range also does not exceed the average width of the "hot area" detected by the eye tracking machine. Therefore, the design can be beneficial and adaptive to small screen sizes.

Figure 3. Readability before and after caption filtering in a cell phone (Discovery video)

3. Results

3.1 Analysis of the captions in original listening materials

The mean of the playing times of the 30 videos is 54.53 minutes (SD=13.69). The two longest videos are 80 and 100 minutes. The average number of sentences in the 30 videos are 760.07 (SD=255.16), and the average number of words are 6030.8 (SD=1760.93). The descriptive statistics of each video is shown in Table 1. In the 30 videos, Table 1 shows that the average number of words per sentence is 8. A special video, called the Biggest and Baddest Bugs (the longest video in the 30 samples) contains the most number of sentences and words. The total number of sentences in this video is 1731, and the average number of words per sentence is 6.69, which is the minimum number among the 30 videos. The video introduces insects for around 100 minutes; however, it depicts the sounds of bugs in the captions. Therefore, we suggest that special captions, such as the words that representing the sounds, be processed separately during caption filtering, even though these sounds are not included in the most frequently used words. Table 1 also shows that the most number of words in a sentence on average in the first video is 9.12, which exceeds the limitation of short-term memory on average. This video is called The Ultimate Guide T-Rex, which broadcasts for 50 minutes, and contains a total of 665 sentences. Accordingly, the number of words per sentence in a video is directly related to the introduction of content.

Table 1. Average words per sentence before caption filtering

Rank	Video ID	Video name	Time length of video (Min.)	Avg.words per sentence
1	016	The Ultimate Guide T Rex	50	9.12
2	027	Channel Discovering Ardi	87	8.84
3	023	Understanding Time	51	8.48
⋮	⋮	⋮	⋮	⋮
28	026	Channel Raising the Kursk 2002	51	7.26
29	025	Great Books Galileos Dialogue	50	7.24
30	024	Worlds Biggest and Baddest Bugs	89	6.69
		Mean	**54.53**	**8.00**
		SD	**13.69**	**0.50**

3.2 Results of reducing most frequently used words

The percentage of the words filtered is 71.88% (Table 2). In the 30 videos, the average number of words per sentence decreases from 8 to 5.88 after filtering. The effects of decreasing long captions are remarkable so that showing them clearly on a small-screen device is more easily achieved. That is, most of the captions contain the words that amount to no more than 7 words after filtering. In detail, 99% of the captions contain no more than 7 words, and 98% of the captions contain no more than 5 words. However, 0.94% of the sentences on average remain unchanged after filtering (SD=0.01). Biggest and Baddest Bugs has the longest broadcasting time and the most number of sentences, but also the highest percentage of sentences unchanged. Of these sentences, 7.68% are unchanged. The other videos have approximately 1% unchanged sentences. Table 2 shows that most of the videos have less than 1% unchanged sentences. EFL learners who have graduated from junior high school or passed the primary certification of GEPT in Taiwan can use their mobile devices for English listening training using video instructional materials because the filtering system effectively enhances the readability and usability of multimedia players in small-screen mobile devices.

Table 2. Percentage of sentences filtered.

Rank	Video ID	Video name	Movie length (Min.)	Sentence unfiltered (%)	Words filtered in captions (%)
1	022	Understanding Space Travel	50	0.59%	75.82%
2	006	The Ultimate Guide Elephants	51	0.00%	74.87%
3	023	Understanding Time	51	0.71%	74.58%
⋮	⋮	⋮	⋮	⋮	⋮
28	026	Raising the Kursk 2002	51	0.93%	69.37%
29	027	Discovering Ardi	87	0.73%	68.32%
30	017	The Ultimate Guide Whales	51	0.78%	64.28%
		Mean	**54.53**	**0.94%**	**71.88%**
		SD	**13.69**	**0.01**	**0.02**

3.3 Readability improvement after caption filtering

Before caption filtering, the words in each sentence are normally distributed, showing a distribution of 1 to 15. The statistical bar chart of the distribution is shown in Figure 4. After caption filtering, the number of words in each sentence decreases and is normally distributed so that the peak of the curve in the diagram shifts to the left. Most of the sentences contain no more than 7 words after filtering; thus the objective and expectation of promoting readability are achieved.

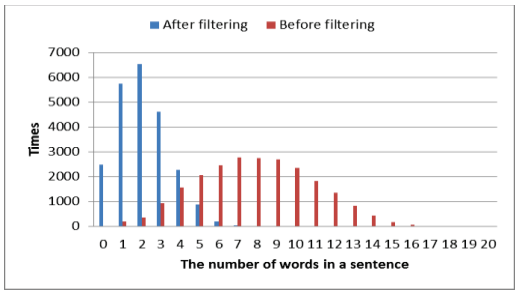

Figure 4. Distribution of words per sentence before and after caption filtering.

The average number of words per sentence in the 30 videos decreases to 2.12; thus the long sentences in the caption are almost shortened. To assess whether the results of caption filtering satisfy the expectations on the 5-, 7-, and 9-word thresholds, we computed the number of sentences that exhibit less than 5, 7, or 9 words before and after caption filtering. The results are shown in Table 3. The degrees of improvement for the different thresholds are all remarkable. For the 4-inch mobile devices, 21.56% of the sentences on average in the 30 videos have no more than 5 words before caption filtering; by contrast, 98.86% of the sentences have no more than 5 words after caption filtering. For the 7-inch mobile devices, 44.46% of the sentences on average have no more than 7 words before caption filtering; 99.98% of the sentences have no more than 7 words after caption filtering. For the 10-inch mobile device, 68.51% of the sentences have no more than 9 words before caption filtering; after caption filtering, 99.99% of the sentences have no more than 9 words. These results indicate that all the expected values for different screen sizes are reached.

Table 3. Readability improvement for mobile devices of different screen sizes.

Thresholds	<=5words(for 4-inch)		<=7words(for 7-inch)		<=9words(for 10-inch)	
Filter or not	After	Before	After	Before	After	Before
Mean	**98.86%**	21.56%	**99.98%**	44.46%	**99.99%**	68.51%

As mentioned above, the results agree with the expectations set in this study. For example, we expected more than 98% of the captions to be less than 5 words after caption filtering. The percentage of captions with less than 5 words is shown in Figure 5 to compare before and after caption filtering values. Almost 100% of the captions agree with the expectations. In summary, the system effectively addresses the problem of readability and facilitates the listening training of EFL learners in the mobile devices by hiding the most frequently used words, which are also usually the easier words. The usability of playing videos in three different sizes of mobile interfaces is promoted, and the readability of caption-assisted learning improves.

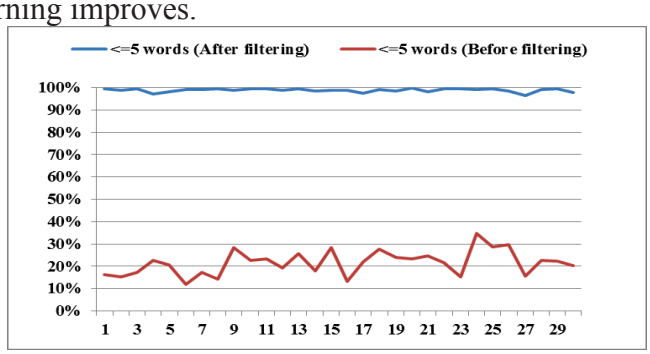

Figure 5. Readability improvement after caption filtering of the listening materials.

4. Conclusions

In this study, the captions fitting more than two lines were merged into one line. We expected all the words to be in a line and within the visual restriction, as well as within the

short-term memory limitation; therefore, we set 5-, 7-, and 9-word thresholds for different screen sizes. In addition, future work should consider using the proposed system or its updated version in analyzing more videos or listening materials in diverse domains, such as English news, soap operas, and so on, so that the effects of the system can be more widely evaluated. Finally, we found that more than 70% of the captions are filtered and hidden in the video samples. Whether the development of harder level options for hiding more words is needed should be considered. Given that the percentage of filtering is quite high, we encourage future researchers to develop easier level options first to hide fewer words to support primary learners and provide adaptive support for EFL listening. In the future, this study will be extended to the release of the system as freeware to enable access for English teachers. The availability of the system can facilitate the use of English listening materials for adaptability to the demands of different proficiency levels.

Acknowledgements

The study is supported by the National Science Council of the Republic of China under contract number NSC 97-2628-S-024-001-MY3 and NSC 100-2631-S-001-001.

References

[1] Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human-computer interaction. *Hillsdale, NJ: LEA*.

[2] Walisten, T.S. (1980). *Cognitve process in choice and decision behavior*. Hillsdale N.Y. Lawrence Erlbaum and Assoc p. 227.

[3] Bos, C.,& Vaughn, S. (2006). *Strategies for Teaching children with learning and behavior problems,* Boston, Massachusetts: Pearson Education.

[4] Kintsch, W.,& van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.

[5] 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.

[6] Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation: Advances in research and theory, 2*, 89-195.

[7] Ogata, H., Wada, M., Gan, L. H., & Yano, Y. (2008). *Supporting a Decision Making for Task Assignments in Language Learning Outside Classroom with Handhelds.* Paper presented at the ICCE 2008, Taipei, Taiwan.

[8] Greer, T. (2010). The US Healthcare Market for Mobile Learning Products and Services: 2009-2014 Forecast and Analysis. from http://www.ambientinsight.com/Resources/Documents/Ambient-Insight-2009-2014-US-Healthcare-Mobile-Learning-Executive-Overview.pdf

[9] Kukulska-Hulme, A. (2006). Mobile language learning now and in the future. In P. Svensson (Ed.), *Fran vision till praktik: Sprakutbildning och Informationsteknik (From vision to practice: language learning and IT)* (pp. 119-134): Sweden: Swedish Net University (Natuniversitetet).

[10] Shohel, M. M. C., & Power, T. (2010). Introducing mobile technology for enhancing teaching and learning to the English language classroom in Bangladesh. Retrieved May 20, 2011, from http://wikieducator.org/images/5/53/M_Mahruf_C_Shohel.pdf

[11] Alexander, P. A., Kulikowich, J. M., & Jetton, T. L. (1994). The role of subject matter knowledge and interest in the processing of linear and non-linear texts. *Review of Educational Research, 64*, 201-252.

[12] 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 Mackey, T. P. & Ho, J. (2008). Exploring the Relationships Between Web Usability and Students' Perceived Learning in Web-Based Multimedia (WBMM) Tutorials. Computers & Education, 50(1), 386–409.

[13] 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. *ICCE2010*, Putrajaya, Malaysia.

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

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