Konin Language Studies
Faculty of Philology, State University of Applied Sciences in Konin, Poland
KSJ 5 (4). 2017. 403-424
http://ksj.pwsz.konin.edu.pl
doi: 10.30438/ksj.2017.5.4.2

The modality effect of input repetition on vocabulary consolidation

Wenhua Hsu
I-Show University, Taiwan
whh@isu.edu.tw

Abstract

This study investigated repeated exposure to English through audiovisual support. The researcher-teacher used video for her *English reading* class and tested its effects on word recall and retention. To help students remember forty newly-introduced words from four news stories, two weeks later, the four news videos were broadcast in four audiovisual modes to four groups of students alternately: (1) captioned, (2) non-captioned, (3) silent captioned, and (4) screen-off. Results showed that the four groups of students recalled 17.65-18.81 words in the second encounter with forty target words through video in different modes. In another week's time, they forgot a total of 5.22-6.58 words from the 17.65-18.81 words under no subsequent repetition condition. Concerning the audiovisual effects on vocabulary learning, audio track only prompted the participants to recall the greatest number of target words than the other three modalities, while sound-off captioned video was the most effective in terms of the least attrition of word knowledge over time. Drawing on the cognitive theory of multimedia learning, this study aims to raise awareness of the modality effect when using video as a repetition medium for vocabulary consolidation.

Keywords: cognitive theory of multimedia learning; dual coding; modality effect

1. Introduction

Repetition is one of the most basic learning techniques. For many language teachers, repetition is highly valued, because without repetition, students are

likely to keep learning and forgetting. However, under the constraint of course time, relearning sessions are often omitted, although it may be argued that repeating contents should be the responsibility of learners themselves.

In the survey on the retention of Spanish vocabulary over eight years, Bahrick and Phelps (1987) demonstrated the importance of repetition in achieving permastore retention. They explained that at the optimum interval of time, learners can retrieve some cognitive traces of previously learned material so that subsequent rehearsal has some effects on their memory. Horst, Parsons and Bryan (2011) explored the repetition effects on young children in word learning by reading the same storybooks to them during the shared storybook reading session. Results showed that the children learned more new words when they were read the same stories repeatedly than when they were read different stories that had the same number of target words. She posited that hearing the same stories repeatedly may have helped the children to predict what was to come next, showing a contextual cueing effect (Chun, 2000; Chun & Jiang, 1998). In her later research, Horst (2013) inferred that through repeated exposures to the same plots, characters and scenes, the children were able to form a robust representation of a new word, because contextual repetition helped to free up their attentional resources, thus enabling them to better attend to new words.

Inspired by the same storybook reading, the researcher wishes to help students learn target words by repeating lessons with audiovisual support rather than with the teacher talk again. Nowadays, tablet computers and smart phones make learning material portable and hence enable students to study anytime and anywhere when they have a few moments, such as during a lunch break or waiting in line. In view of the prevalence of multimedia-enabled mobile devices with Internet access, which offer another channel for instant access to the target language (TL), this research aimed to enhance relearning through video or audio and to examine the modality effect on vocabulary recall as well as its potential for aiding EFL learners in retaining vocabulary. The research questions guiding this study were:

1. Is there any significant difference in word recall and retention after the second encounter with the same input but in different audiovisual modalities?
2. What is the students' perception toward relearning with audiovisual support?

2. Literature review

2.1. The cognitive theory of multimedia learning

Multimedia learning has been highly advocated since the 1990s, because multimedia application can create diverse modalities of input to cater to different learning styles. The cognitive theory of multimedia learning (CTML) was developed by Mayer (2001, 2009, 2014) and other psychologists, who endeavored to address the issue of how multimedia material can be adapted to learners' working memory limitations. Drawing upon the studies on different multimedia conditions that may result in better learning, Mayer (2009) enumerated twelve principles for the presentation of multimedia information to maximize learning effectiveness.

The central point of the CTML is that "people learn more deeply from words and pictures than from words alone" (Mayer, 2009, p. 47), which is referred to as the multimedia principle. In support of the multimedia principle, Mayer and Anderson (1991, 1992) undertook a series of experiments, in which some participants viewed narrated animations showing how a bicycle pump and automobile brakes worked, while other participants simply listened to verbal explanation. Results show that those who heard the verbal explanation with animations performed better on problem-solving tests than those who heard the narration only, which corroborates the multimedia principle.

In experimental psychology, *modality* means the presentation mode of study material and the term *modality effect* refers to how learners perform in learning and memory depends on the presentation mode of study material. When better learning occurs in a mixed-mode presentation of information (i.e., partly visual and partly auditory) rather than in a single mode, the modality effect is usually explained from the cognitive load perspective. Sweller (1988, 1994) as well as Moreno and Mayer (1999) theorized that when information consists of pictures and visual text, the visual working memory load increases, since the two types of input are processed in the same system. In contrast, when information is presented both verbally and visually, the total working memory capacity is increased, because auditory and visual information is processed in their respective systems. Based on this assumption, Mayer (2009) put forward the modality principle that people learn better from graphics and narration than from graphics and visual text.

The cognitive effects concerning multimedia learning are primarily based upon three assumptions: dual channel, limited capacity and active processing (cf. Mayer, 2003; Mayer & Moreno, 1998). The dual-channel assumption is derived from Baddeley's (1986) working memory model and Paivio's (1971, 1986) dual coding theory. The limited capacity assumption is based on Sweller's (1988,

1994) cognitive load theory, which presumes that each channel has a limited capacity. The active processing assumption states that humans can only process a finite amount of information in a channel at a time and make sense of incoming information by actively creating mental representations (Mayer, 2009).

According to Baddeley's (1986) working memory model, there are two modality-specific slave systems involved in the processing of information. The first is for processing visual and spatial information while the second is for processing acoustic information. Paivio (1971, 1986) gave equal weight to verbal and visual information processing. The dual coding assumption postulates that verbal and visual information is processed along the auditory and visual channels respectively with working memory creating distinct representations for information processed in each channel (Paivio, 1971). Mayer (2001) highlighted that both the auditory and visual channels can be used optimally instead of overloading one channel. This principle has often been used as one of the multimedia design guidelines.

In the field of cognitive psychology, cognitive load refers to the total amount of mental effort being used in the working memory. Grounded on Sweller's (1988, 1994) cognitive load theory, the limited capacity assumption in the CTML posits that there is a limit to the amount of information that can be processed at a time by working memory. Each channel has a limited capacity for holding and manipulating knowledge (Baddeley, 1986). When too many visual parts are displayed at a time, the visual channel may become overloaded. Likewise, when spoken words and other streams of sound are broadcast concurrently, the auditory channel may become overburdened. Mayer (2009) cautioned that overloading working memory does not result in more learning. Instead, learning is impaired when the working memory capacity is exceeded (De Jong, 2010).

Redundant input may also cause cognitive overload. In his experiments, Mayer (2009) found that the participants learned better from graphics and narration than from graphics, narration and written text. He pointed out the phenomenon that when spoken text and identical visual text stress working memory and do not lead to additional knowledge gains, redundancy would be a problem. For instance, playing an animation with concurrent narration and captions is equivalent to transmitting the same words in two forms (spoken and written) simultaneously. Learners may experience cognitive overload in the visual information-processing channel because the added captions may compete with animated images for cognitive resources in the visual channel. Learners may split their attention (Sweller, 2005) or use attention selectively (Wickens, 2007) between two visual modalities to infer meanings, because their visual working memory capacity is limited.

2.2. Past studies on audiovisual modalities

For language learning, video content can be transmitted visually, aurally, or both, thereby allowing learners to have multiple channels to the target language. To help listening comprehension, video materials are often augmented with the first language subtitles or TL captions. The question regarding how learners balance the simultaneous intake of text, sound, and image still remains unanswered.

In research on which component of video (i.e., text, sound, and image) is paid the most attention to, Chai and Erlam (2008) reported that when viewing captioned video, learners tend to prioritize reading captions over listening, which may prevent them from processing auditory contextual clues. Sydorenko (2010) also found similar results. She asked the learners to rate text, sound, and image according to the amount of attention they paid to them. They replied that they paid most of their attention to captions although they thought that images were equally helpful. The result is in accord with Jensema, Danturthi, and Burch's (2000) eye-tracking study. They found that learners tend to spend more time on captions (circa 84% of the time) as opposed to viewing video using their peripheral vision. Also using an eye-tracking method to investigate learners' attention paid to captions, Duchowski (2002) as well as Winke, Gass, and Sydorenko (2013) discovered that when the video content is familiar, learners' eye fixation on screen text gets shorter because they do no need captions as much to extract meaning.

The issue of whether captioned video is better than non-captioned video for language learning is still inconclusive. Some researchers presume that captions enable learners to confirm what was heard and to remember words more accurately (Chai & Erlam, 2008; Danan, 2004). Other researchers found that captions aid form-meaning mapping by helping learners visualize what they hear (Bird & Williams, 2002; Winke, Gass, & Syndorenko, 2010). Still other researchers are doubtful about the effectiveness of captions due to an excessive cognitive load (Mayer & Moreno, 1998; Pujola, 2002).

In their quasi-experiment on word learning, Bird and Williams (2002) introduced new words to advanced English learners in three modes: (1) text and sound, (2) text only, and (3) sound only. Results demonstrated that even without subsequent textual support, learners who had viewed text with sound could still identify the words presented aurally. Bird and Williams (2002) concluded that the bimodal presentation of new words (text and sound) resulted in better aural word recognition. Similarly, Sydorenko (2010) conducted a survey on three modalities by playing video with captions, without captions and with captions but without sound, and examined their effects on vocabulary gain. Partially in agreement with Bird and Williams' (2002) findings, her data indicated that the learners receiving captioned video performed better in visual word recognition than

aural word recognition. Conversely, those without receiving captions scored higher in the recognition of aural words than visual words. The results also showed that among the three video modes, the learners learned the most new words when videos were played with captions. Accordingly, Sydorenko (2010) concluded that captioned video tends to aid the learning of word meaning and the recognition of visual word, while non-captioned video tends to facilitate spoken word recognition.

Regarding the effectiveness of visual text, Mayer and Moreno (1998) treated captioning support with reserve. They carried out two experiments by playing animation that showed the formation of lightning and the operation of automobile brakes. The groups of students receiving concurrent narration outperformed those receiving concurrent captions in depicting the process and solving the problems on the subsequent recall tasks. The results provided some evidence that textual information should better be spoken than written when pictures are presented at the same time. Pujola (2002) also cast some doubt on the captioning effect. In his research on learning strategies using help facilities in a web-based multimedia program, he detected that although some learners with the help of captions made progress in listening comprehension, they still depended on screen text instead of paying much attention to spoken words. Another survey on whether a captioned video was helpful was conducted by Taylor (2005) with college learners of Spanish. He found that captions were distracting for many Spanish beginners and made it difficult for them to pay attention to text, sound and image all at a time.

Different from previous studies that focus on learning with multimedia, this research highlighted relearning and aimed to investigate input repetition enhanced with audiovisual support and its effects on vocabulary consolidation. It is hoped that the results may contribute to the literature of multimedia learning in this regard.

3. Research method

3.1. *English reading* course

English reading is a required course for non-English-majoring freshmen at a university in Taiwan. A total of 53 students of similar age participated in this study. They came from one intact, pre-intermediate *English reading* class based on their English scores of the nationwide college entrance exam. Although there were individual variations in language ability, the participants were homogeneous in terms of English learning backgrounds under Taiwan's educational system.

In addition to a designated textbook for use in class, *English reading* teachers are encouraged to use supplementary material. At the end of the semester, the students in the pre-intermediate classes are expected to have a vocabulary of the most frequent 3,000 word families at least. To help students achieve this goal, we use simplified English news articles as a supplement to the textbook. They come from the website *News in levels* (http://www.newsinlevels.com/), which provides one- to two-minute news video segments with narration and transcripts at three levels, ranging from the most frequent 1,000 to 3,000 word families. The reasons for using News in Levels are that the website is free; each news story is real life, and the vocabulary at Level 3 is moderately beyond our students' English abilities.

3.2. Research design

Four video clips were randomly selected from the *News in levels – Level 3* (see Appendix for one transcript). At this level, news stories are written within a controlled vocabulary at the 3,000-word-family level. The video transcripts were entered into the *RANGE* program (Heatley, Nation, & Coxhead, n.d.) for analysis. *RANGE* is installed with the frequency-ranked twenty-five 1,000 word families from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) (Nation, 2012) and can be used to measure the vocabulary level of a text. Table 1 provides some details about the four videos. As the figures show, the four videos seemed to be equal in terms of the duration, the number of words as well as the percentage of words within the first 2,000 and 3,000 word families.

Table 1 Video profile

News video	Duration	Word tokens	Word Types	Word Family	% of words within the 2K word families	% of words within the 3K word families
Video 1: World's deadliest walkway set to reopen	88 seconds	197	125	104	80.89%	90.02%
Video 2: France bans ultra-thin models	92 seconds	219	131	116	81.65%	91.83%
Video 3: Capital punishment in Utah	88 seconds	196	128	106	80.31%	90.42%
Video 4: Japan playground closed over nuclear fears	86 seconds	206	124	105	81.23%	91.14%

From each news story, ten words were selected for teaching (see Appendix for words in bold), totaling forty words as target words. They are the words beyond the 2,000-word family level, which are likely to be unfamiliar to our students (e.g., *alternative*, *concoction*, *condemnation*, *distress*, and *execution*).

They are not the loanwords from Chinese (e.g., *tofu*), since this type of vocabulary allows our students to make associations with similar Chinese pronunciation and may therefore be effortless to learn. Table 2 gives a snapshot of the research design.

Table 2 Research design

English Reading session	One week later	53 students	Lesson repetition via video one week after pretest				Right after video	One week after video
			Video 1	Video 2	Video 3	Video 4		
Four news stories served as supplementary material.	Pretest of target words	Group1 (N=13)	Mode A →	Mode B →	Mode C →	Mode D	Immediate postest for vocabulary recall	One-week delayed postest for vocabulary retention + Questionnaire
		Group2 (N=13)	Mode B →	Mode C →	Mode D →	Mode A		
		Group3 (N=13)	Mode C →	Mode D →	Mode A →	Mode B		
		Group4 (N=14)	Mode D →	Mode A →	Mode B →	Mode C		

Note: A: captioned; B: non-captioned; C: silent captioned; D: screen-off

In the *English reading* session, we used four news stories as supplementary material and taught the content. To prevent the students from becoming conscious of a possible quiz, intensive practice of target words was deliberately ignored. Still, the meaning of each target word and examples for its usage were written on the whiteboard. The time devoted to each word did not exceed three minutes. The students were not told about any vocabulary test given in the following two weeks.

To assess receptive knowledge of the target words, an unannounced test was administered one week after news story reading. The one-week interval was intended to identify which target words after initial learning had not been kept in mind. On the vocabulary test, forty target words were intermixed with sixty other words that did not occur in the four news stories. They served as distracters to prevent alerting the students to the target words. However, only the target words were scored. The students were asked to write down Chinese meanings for each word. A full point was awarded for fully correct answers, a half point for partially correct answers and no point for a totally wrong answer or no answer. Take one target word, such as *dwindle*, for example. One point was given for the answer 減少 or 縮, a half point for 減輕, and zero for 閃爍. Meanwhile, one of the researcher's colleagues, who taught the same subject, was requested to help mark the test papers. For an inter-rater reliability check, Cohen's Kappa was calculated using SPSS, and the *k* value (= .96 > .80) indicated a substantial level of agreement between the two raters. The vocabulary test scores before video served as a baseline and termed as pre-video test.

Another week later, the four news videos were alternately played under four audiovisual conditions to prompt the students to recall the newly-introduced words from the four news stories. The four broadcasting modes were (1) captioned, (2) non-captioned, (3) captioned with sound off, and (4) screen-off (audio track only). The four modes were selected because they involved three audiovisual components (i.e., text, sound and image) as well as two information-processing channels (i.e., auditory and visual) (see Table 3). Thus, when comparing the word recall prompted by the four audiovisual modes, consideration of their constituents may help to explain the results.

Table 3 Four modalities

Mode	Components	Information-processing channels
Captioned	Sound, text, image	Auditory & visual
Non-captioned	Sound, image	Auditory & visual
Silent captioned	Text, image	Visual
Screen-off	Sound	Auditory

To maintain the class intact, the video phase was carried out in a multimedia laboratory. Every participant had a cubicle desk equipped with a headset and a computer connecting to the Internet. Based on the students' English scores on the college entrance exam, the pre-intermediate English Reading class was further divided into four mixed-proficiency groups in a balanced fashion. The four groups of students received video simultaneously each time but in different modes over the four news stories (see Table 2 for the alternation of video modes between and within groups). During the video phase, any form of dictionary use was forbidden. Then an immediate postest for measuring word recall was administered. One week later, the same vocabulary test (but with the test items being arranged in a different order) was given unannounced. The one-week delayed postest was used to assess the retention of the target words.

It may be queried whether, in class activities, it is common to use such a methodology with reading first and video playing in the next class. The goal of using video was to contextualize the newly-taught words. After reading, the video provided contextual cueing to promote word recall. In practice, relearning by watching video can be an after-school assignment. For the sake of experiments, lesson repetition through video was given in the course time. It is also worth mentioning that since this research focused on audiovisual support to enhance lesson repetition, the treatment without video or audio was not factored in.

Upon completion of the vocabulary test, the questionnaires with five open-ended questions concerning perceptions towards audiovisual support were distributed to the students. The questionnaire was used to ensure triangulation of data and to offer some insight into the entire learning process.

4. Results and discussion

4.1. Pre-video test

A one-way between-groups ANOVA was conducted to compare the pre-video test scores. Results show that there was no significant difference among the four groups in the pre-video test [$F(3, 49) = 0.082$, $p = .969$, $\eta_p^2 = .005$]. The four groups of students altogether remembered an average of 3.46 to 4.31 words one week after the initial learning of forty target words (see Table 4).

Table 4 Descriptive statistics of pre-video test

		Pre-video test scores	
Groups	*N*	*M*	*SD*
Group 1	13	3.77	0.55
Group 2	13	4.08	0.54
Group 3	14	4.31	0.68
Group 4	13	3.46	0.50

Note: *N* = Number of students; *M* = mean; *SD* = standard deviation. Full marks = 40 with 1 point per word.

4.2. Immediate postest

Table 5 provides a summary of the immediate postest scores of target words for four groups alternately receiving different video modes across four news stories. Receiving different forms of audiovisual support to enhance vocabulary repetition, the four groups of students recollected a total of 17.65 to 18.81 target words on average (see the rightmost column in Table 5) in comparison with the recall of only 3.46 to 4.31 words in the pre-video test (see Table 4).

Table 5 Descriptive statistics of immediate postest

Group	News story 1			News story 2			News story 3			News story 4			Total	
	Mode	*M*	*SD*	Mode	*M*	*SD*	Mode	*M*	*SD*	Mode	*M*	*SD*	*M*	*SD*
Group 1	A	3.69	0.60	B	4.46	0.48	C	4.35	0.72	D	5.58	0.76	18.08	2.56
Group 2	B	4.54	0.83	C	3.81	0.63	D	6.15	0.85	A	4.31	0.38	18.81	2.69
Group 3	C	3.92	0.86	D	5.46	0.66	A	3.88	0.84	B	4.77	0.53	18.03	2.89
Group 4	D	5.43	0.62	A	3.93	0.58	B	4.29	0.43	C	4.00	0.62	17.65	2.25

Note: A: captioned; B: non-captioned; C: silent captioned; D: screen-off; *M*: Mean; *SD*: standard deviation. Full marks = 40 with 1 point per word and 10 words per news story

Since each group had access to four video modes, a direct comparison within the group was made. The modality effect was found in the Mode D (screen-off/audio track only), which prompted the greatest recall of words

among the four modes (in Group 1, *M* under Mode D = 5.58 > *M* = 3.69, 4.46, 4.35 under Modes A, B, C in turn; in Group 2, *M* under Mode D = 6.15 > *M* = 4.31, 4.54, 3.81 under Modes A, B, C in turn; in Group 3, *M* under Mode D = 5.46 >M = 3.88, 4.797, 3.92 under Modes A, B, C; in Group 4, *M* under Mode D = 5.43 > *M* = 3.93, 4.29, 4.00 under Modes A, B, C).

The pure audio effect was also found in the between-groups comparison. The better results of Mode D in the between-groups comparison were consistent with those in the within-group comparison. When a group was in the turn of receiving soundtrack only, they remembered more words than the other three groups receiving the other video modes. In news story 1, Group 4 with Mode D remembered 5.43 words, while Groups 1, 2 and 3 with Modes A, B and C had recall of 3.69, 4.54 and 3.92 words, respectively. In news story 2, Group 3 with Mode D recalled the most words as the data 5.46 versus 4.46, 3.81 and 3.93 has shown. In news story 3, Group 2 receiving screen-off video performed far better than Group 3 receiving captioned video (having recall of 6.15 words out of 10 target words with Mode D versus 3.88 words with Mode A). In news story 4, it was the turn of Group 1 to receive Mode D to prompt word recall and Group 1 recollected the largest number of target words among the four groups (5.58 versus 4.31, 4.77 and 4 words). Overall, across the groups, screen-off/audio track only (Mode D) facilitated recall of 5.43 to 6.15 words out of ten target words from each news story while captioned video (Mode A) fostered 3.69 to 4.31 words, non-captioned video (Mode B) 4.29 to 4.77 words and silent captioned video (Mode C) 3.81 to 4.35 words per news story.

Due to different news stories, the confounding variable needed to be examined. To test whether there was an interaction between modes and news stories, a series of ANCOVAs was performed. The independent variable involved video modes and the dependent variable was the immediate postest scores against the pre-video test scores as the covariate. In Table 6, all $p > .05$ for the covariate pretest scores indicates that there was no interaction between the pretest and the modes over the four news stories [in News Story 1, $F(3, 49) = 3.23$, $p = .079$, $\eta_p^2 = .063$; in News Story 2, $F(3, 49) = 0.789$, $p = .379$, $\eta_p^2 = .016$; in News Story 3, $F(3, 49) = 0.883$, $p = .352$, $\eta_p^2 = .018$; in News Story 4, $F(3, 49) = 0.234$, $p = .631$, $\eta_p^2 = .005$]. In contrast, the main effect of the mode variable was significant [in News Story 1, $F(3,49) = 16.121$, $p < .001$, $\eta_p^2 = .502$; in News Story 2, $F(3,49) = 21.245$, $p < .001$, $\eta_p^2 = .570$; in News Story 3, $F(3,49) = 25.451$, $p < .001$, $\eta_p^2 = .614$; in News Story 4, $F(3,49) = 17.783$, $p < .001$, $\eta_p^2 = .526$]. There were consistently significant audiovisual effects on immediate word recall after controlling the confounding variable news stories. The strength of relationship between the video modes and the immediate postest scores, as measured by

partial eta squared, was strong, with the mode variable explaining 50.2%, 57%, 61.4% and 52.6% of the variance of the immediate postest scores in news stories 1 to 4 respectively, holding constant the pretest covariate.

Table 6 ANCOVA results for the immediate postest against pretest.

			News Story 1			*News Story 2*
Source	*F*	Sig.	η_p^2	*F*	Sig.	η_p^2
Pretest covariate	3.23	.079	.063	0.789	.379	.016
Video modes	16.121	.000*	.502	21.245	.000*	.570
			News Story 3			*News Story 4*
Source	*F*	Sig.	η_p^2	*F*	Sig.	η_p^2
Pretest covariate	0.883	.352	.018	0.234	.631	.005
Video modes	25.451	.000*	.614	17.783	.000*	.526

Note: * < .05

Table 7 Post hoc test results in immediate recall

News Story 1			*News Story 2*		
Condition	*MD*	Sig.	Condition	*MD*	Sig.
A-B	-0.585	.083	A-B	-0.551	.421
A-C	-0.308	.777	A-C	0.143	.973
A-D	-1.797	.000*	A-D	-1.511	.000*
B-C	0.577	.289	B-C	0.654	.224
B-D	-0.912	.027*	B-D	-1.000	.022*
C-D	-1.489	.000*	C-D	-1.654	.000*
News Story 3			*News Story 4*		
Condition	*MD*	Sig.	Condition	*MD*	Sig.
A-B	-0.423	.579	A-B	-0.423	.515
A-C	-0.462	.552	A-C	0.360	.631
A-D	-2.308	.000*	A-D	-1.231	.001*
B-C	-0.038	.999	B-C	0.783	.053
B-D	-1.885	.000*	B-D	-0.808	.049*
C-D	-1.846	.000*	C-D	-1.591	.000*

Note: *MD*: mean difference; A: captioned; B: non-captioned; C: silent captioned; D: screen-off.
* $p < .05$

In Table 7, the post hoc analyses using the Scheffé criterion for significance indicate that the mean differences of immediate vocabulary recall between captioned and screen-off video (Modes A-D), between non-captioned and screen-off video (B-D), and between silent captioned and screen-off video (C-D) were consistently significant across the four news stories (all $p < .05$). Ignoring any negative sign, we can see that the mean differences between the three pairs of mode comparison (A-D, B-D, C-D) were larger than those between the other three pairs of mode comparison (A-B, A-C, B-C) across the four news stories. All the negative signs in the mean differences between Modes A-D, B-D and C-D

show that lesson repetition via pure soundtrack (Mode D) resulted in greater recall of words than the other three modes A, B and C.

Among the three pairwise comparisons A-B, A-C and B-C after the exclusion of Mode D, the mean differences between captioned and non-captioned video (A-B) across the four news stories were all negative, reflecting that captioned video (Mode A) was less effective than non-captioned video (Mode B) in prompting word recall. As to the B-C comparison, the three positive values over the four news stories (in News Story 1, *MD* =0 .577; in News Story 2, *MD* = 0.654; in News Story 4, *MD* = 0.783) indicate that the group receiving non-captioned video (Mode B) generally outperformed the group receiving silent captioned video (Mode C) in word recall.

To sum up, concerning the audiovisual effect on immediate word recall, screen-off video/audio track only (Mode D) as a repetition medium ranked top and non-captioned video (Mode B) placed second, followed by silent captioned video (Mode C) and captioned video (Mode A) at the bottom. The results have some implications for relearning with audiovisual support. The greatest word recall as a result of pure audio stimuli contradicts the multimedia principle, which holds that words and pictures are more conductive to learning than words alone. This may be explained by the fact that learning and relearning are different and therefore the multimedia principle may not apply in the relearning condition. In the later questionnaire section, the student views on screen-off video may reflect why better recall occurred in auditory support only. The second best immediate recall performance prompted by non-captioned video supports the dual coding theory of working memory (Baddeley, 1986; Mayer & Moreno, 1998), which maintains that memory load is reduced by auditory and visual working memory that work in tandem to promote information processing. The least vocabulary recall after captioned video bore some evidence of split-attention (Sweller, 2005) and redundancy effects (Clark & Mayer, 2016; Hoffman, 2006). The learners under this mode may have experienced cognitive overload in the visual channel (i.e., image and on-screen text) and could not focus. The redundancy effect can also be detected in the comparison between captioned and non-captioned video. The redundancy principle of multimedia learning states that people learn better from graphics and narration than from graphics, narration and on-screen text. In the relearning session, concurrent on-screen text may be redundant, which is shown in the result that non-captioned video consistently prompted the students to recall more words than captioned video across the four news stories.

4.3. One-week delayed postest

Table 8 provides a summary of one-week delayed postest scores of target words. One week after the video session, the students remembered a total of 12.12 to

12.81 words out of 40 target words (see Table 8) as opposed to the immediate recall of 17.65 to 18.81 words in the second contact with the lesson through video (Table 5). Results showed that there was a decline in vocabulary retention. However, the one-week delayed postest scores were still higher than pre-video test scores (see Table 4 for the recall of only 3.46 to 4.31 words under no review), revealing the importance of relearning. Despite the lapse of time, the students with the second exposure to target words still performed better in word retention than when there was no subsequent review, as 12.12 to 12.81 words versus 3.46 to 4.31words have shown.

Table 8 Descriptive statistics of one-week delayed postest

News Story 1			*News Story 2*		
Condition	*MD*	Sig.	Condition	*MD*	Sig.
A-B	-0.585	.083	A-B	-0.551	.421
A-C	-0.308	.777	A-C	0.143	.973
A-D	-1.797	.000*	A-D	-1.511	.000*
B-C	0.577	.289	B-C	0.654	.224
B-D	-0.912	.027*	B-D	-1.000	.022*
C-D	-1.489	.000*	C-D	-1.654	.000*
News Story 3			*News Story 4*		
Condition	*MD*	Sig.	Condition	*MD*	Sig.
A-B	-0.423	.579	A-B	-0.423	.515
A-C	-0.462	.552	A-C	0.360	.631
A-D	-2.308	.000*	A-D	-1.231	.001*
B-C	-0.038	.999	B-C	0.783	.053
B-D	-1.885	.000*	B-D	-0.808	.049*
C-D	-1.846	.000*	C-D	-1.591	.000*

Note: A: captioned; B: non-captioned; C: silent captioned; D: screen-off; *M*: Mean; *SD*: standard deviation. Full marks = 40 with 1 point per word and 10 words per news story.

To examine whether there were significant differences among the four groups in the one-week delayed postest scores, a series of post hoc tests using the Scheffé's method were performed. Table 9 shows that the four groups of students, receiving the same lesson repetition but in different channels of exposure retained target words significantly differently (all $p < .05$; in News Story 1, $F(3, 49) = 3.517$, $p = .022$; in News Story 2, $F(3, 49) = 5.192$, $p = .003$; in News Story 3, $F(3, 49) = 6.202$, $p = .001$; in News story 4, $F(3, 49) = 4.164$, $p = .011$).

In Table 9, the mean difference of delayed postest scores between captioned and silent captioned video (Modes A-C) was significant in news stories 2 and 4 ($p = .004$ and $p = .011$), while the mean difference between non-captioned and silent captioned video (B-C) was significant in news stories 1 and 2 ($p = .037$ and $p = .043$). With one week passing by, the students repeating lessons under Mode A or under Mode B retained significantly fewer words than those under

Mode C (for Modes A-C, *MD* = -0.876 and *MD* = -0.734 in News Stories 2 and 4; for Modes B-C, *MD* = -0.615 and *MD* = -0.597 in News Stories 1 and 2). The mean difference of delayed postest scores between captioned and screen-off video (A-D) as well as that between non-captioned and screen-off video (B-D) was significant in News Story 3 (for Modes A-D, *MD* = 0.846, *p* = .039; for Modes B-D, *MD* = 1.159, *p* = .002). The students repeating News Story 3 under Mode A or under Mode B retained significantly more words than those under Mode D after one week. In light of these data, the vocabulary decline under the pure auditory mode (screen-off) was the largest among the four modes, while the attrition of word knowledge under the pure visual mode (silent captioned) was the smallest.

Table 9 ANOVA results in vocabulary decline and post hoc tests

News Story 1				*News Story 2*			
ANOVA	Condition	*MD*	Sig.	ANOVA	Condition	*MD*	Sig.
F (3, 49)= 3.517 p= .022*	A-B	0.154	.902	F (3, 49)= 5.192 p= .003*	A-B	-0.299	.627
	A-C	-0.464	.175		A-C	-0.876	.004*
	A-D	0.022	1.00		A-D	-0.453	.272
	B-C	-0.615	.037*		B-C	-0.597	.043*
	B-D	-0.132	.932		B-D	-0.154	.930
	C-D	0.484	.132		C-D	0.423	.347
News Story 3				*News Story 4*			
ANOVA	Condition	*MD*	Sig.	ANOVA	Condition	*MD*	Sig.
F (3, 49)= 6.202 p= .001*	A-B	-0.313	.735	F (3, 49)= 4.164 p= .011*	A-B	-0.308	.561
	A-C	0.115	.982		A-C	-0.734	.011*
	A-D	0.846	.039*		A-D	-0.308	.561
	B-C	0.429	.500		B-C	-0.426	.261
	B-D	1.159	.002*		B-D	0.000	1.00
	C-D	0.731	.095		C-D	0.426	.261

Note: MD: mean difference; A: captioned; B: non-captioned; C: silent captioned; D: screen-off.
* $p < .05$

Table 10 Descriptive statistics of vocabulary decline one week after video (the mean difference between one-week delayed postest and immediate postest)

Group	News Story 1			News Story 2			News Story 3			News Story 4			Total	
	Mode	*MD*	*SD*	*Mode*	*MD*	*SD*	*Mode*	*MD*	*SD*	*Mode*	*MD*	*SD*	*MD*	*SD*
Group 1	A	-1.69	0.43	B	-1.31	0.43	C	-1.50	0.61	D	-1.46	0.48	5.96	1.95
Group 2	B	-1.85	0.63	C	-0.73	0.53	D	-2.23	0.67	A	-1.77	0.60	6.58	2.43
Group 3	C	-1.23	0.48	D	-1.15	0.75	A	-1.38	0.92	B	-1.46	0.59	5.22	2.74
Group 4	D	-1.71	0.51	A	-1.61	0.59	B	-1.07	0.65	C	-1.04	0.50	5.43	2.25

Note: *MD*: mean difference; *SD*: standard deviation; A: captioned; B: non-captioned; C: silent captioned; D: screen-off

Table 10 displays some details about the reduction in word retention. As one week went by after the second encounter with the 40 target words through

video, the students forgot a total of 5.22 to 6.58 words out of the short-term memory of 17.65 to 18.81 words (see Table 5) and retained knowledge of 12.12 to 12.81 words (Table 8). As regards the attrition of word knowledge under each mode across four news stories, the reduced vocabulary under captioned video (Mode A) totaled 6.45 words [-1.69 + -1.61 + -1.38 + -1.77= -6.45] and the vocabulary loss under non-captioned (Mode B), silent captioned (Mode C) and screen-off video (Mode D) was 5.69, 4.5 and 6.55 words respectively. Concerning the audiovisual effect on vocabulary retention, pure soundtrack (screen-off) as a repetition medium was the least effective, since its vocabulary loss was the biggest, followed by captioned video. Because of the least attrition of word knowledge, pure visual stimuli (silent captioned) appeared to be of greatest help in retaining vocabulary over time. The result of one-week delayed postest contradicted that of immediate postest, which shows that audio was the most helpful in stimulating the immediate recall of newly-introduced words.

The outcome that lesson repetition enhanced with silent captioned video (pure visual stimuli) led to better vocabulary retention than the other three modes may be explained by modality congruency, since the vocabulary test was in written form. Rummer, Schweppe, and Martin (2009) came up with the modality congruency effect, which posits better oral recall of auditory input and better written recall of visual input. When the presentation mode and the recall mode are in congruent relation, there should be an advantage for written recall of visually-presented materials over written recall of auditorily-presented materials and vice versa. In this research, the higher written test scores under the silent captioned mode (visual text) than under the screen-off mode (verbal text) seemed to meet this hypothesis.

In terms of eclectic benefits, non-captioned video gave rise to the second minimal decline in word retention and the second greatest immediate recall of target words. As discussed earlier, dual-coding theory (Paivio, 1971) may account for such results. The cognitive load in either the auditory channel or the visual channel is reduced when the auditory working memory and the visual working memory work jointly.

As with the immediate postest results, the average score under the captioned mode was the lowest among the four modes in the one-week delayed postest. Too many visual components (text and image) as well as concurrent verbal and visual texts may have brought about cognitive distraction (Sweller, 2005) and the redundancy effect (Clark & Mayer, 2016; Hoffman, 2006), respectively, thus leading to the worst performance both in the immediate word recall test and in the one-week delayed test.

4.4. Student views

After the word retention test, the questionnaires were distributed to the class. To elicit more responses, the students were allowed to answer the questions in Chinese. Their comments were categorized according to the gist of their statements. Questions 1 to 4 were asked separately, ''How did you feel about (1) captioned, (2) non-captioned, (3) silent captioned, and (4) screen-off video?'' The answers to these four questions showed a level of agreement with Question 5, ''Which video mode do you prefer?''

For Question 5, 30 out of 53 students preferred non-captioned video while fourteen students preferred to view video with captions. Nine students indicated their preference for soundtrack only; however, none of the students liked sound-off captioned video. When asked how they felt about captioned video, more than half of the students mentioned that they did not pay much attention to captions when provided. Some of them said that when a string of text was on screen, they subconsciously diverted their eyes to other viewable areas of the screen. The students' little need for captions may be attributed to the fact that the content had already been taught and they knew the general idea. The questionnaire result coincides with the work by Winke, Gass, and Sydorenko (2013), who found that learners' eye fixation on captions was shorter when video content was familiar.

Among the 30 students who preferred non-captioned video, three indicated that without inattention to captions, they could attend better to the native English narrator's pronunciation through the earphone. This implies that their attention seemed to be selective. When the students repeated lessons, they had some leeway to select their focus of attention based on individual needs. After comprehending word meanings at the initial learning, they tended to attend to spoken form in the second exposure to the same lesson. As Chai and Erlam (2008) as well as Sydorenko (2010) have advocated, using captioned or non-captioned video depends on learners' needs and proficiency levels.

The nine students who showed their preference for audio pointed out that upon hearing some keywords from the news narration, they recollected the general meaning that had been taught in the previous lesson. However, silent captioned video made a handful of students become uneasy. They said that when the sound was off, they had to strain their eyes to read the captions rather than viewing the video. A similar result can be found in Sydorenko's (2010) investigation into the effect of input modality. Most of her students paid more attention to screen text than image.

The students' aversion to silent captioned video may be explained by the limited capacity assumption and dual coding theory. The added on-screen text

may compete with video image for the limited capacity in the visual information-processing channel and therefore the visual channel becomes overburdened, resulting in a split-attention effect. In contrast, the cognitive load is reduced during non-captioned video, because aural words are processed in the auditory channel, which frees up the working memory capacity. One observation supports this inference. It was found that while viewing silent captioned video, quite a few students had a serious facial expression and did not look lighthearted. They were also less responsive than those viewing captioned or non-captioned video.

Although in the questionnaire, none of the students liked silent captioned video and two students even complained about eye strain during viewing with sound off, the result of one-week delayed vocabulary test showed a different picture. Among the four broadcasting modes, the sound-off captioned mode helped students to retain the memory of the greatest number of target words over time. Generally speaking, relearning through audiovisual support seemed to be acceptable to the students. They felt that repeating lessons through video reduced boredom and tension as opposed to the teacher talk.

5. Conclusion

5.1. Findings

In this research, there were significant differences in immediate word recall and word retention after the lesson was repeated through video in different modes. The four groups of students recollected a total of 17.65 to 18.81 words during the second encounter with forty target words through video in four broadcasting modes. Concerning the modality effect on immediate word recall, pure audio as a repetition medium achieved the best result, followed by non-captioned and silent captioned video with captioned video at the bottom. However, in another week's time, the students forgot an average of 5.22 to 6.58 words from the immediate recall of 17.65-18.81 words under no subsequent repetition. Under the screen-off mode, the decline in vocabulary recall was the largest while the attrition of word knowledge under the silent captioned mode was the smallest, followed by the non-captioned mode. In other words, sound-off captioned video was the most effective in retaining vocabulary over time. Overall, the data bears some evidence that the presentation mode of lesson repetition has some effects on word recall and retention.

5.2. Pedagogical implications and recommendations

Although the results can only be deemed as indicative rather than conclusive due to the small sample size, the outcomes have some pedagogical implications

for vocabulary learning. In an EFL setting, insufficient exposure to target words may impede learners from retaining them. Repeated exposure is one of the keys to vocabulary acquisition. If repetition is not pursued, learning may be in vain. Before memory fades out, any form of access to the TL after initial contact can enhance recall. With the prevalence of mobile phones with audiovisual setup, mobile lesson repetition may be feasible in helping EFL learners review words in the context beyond traditional classrooms. Secondly, English reading teachers may need to spend some time selecting audiovisual materials in connection with course lessons because contextual repetition facilitates word learning. They also need to take account of their students' proficiency and adopt audiovisual materials which are appropriately challenging in lexis.

Admittedly, this study has been conducted within a focus on words. When watching video, students may take heed to recurrent multiword sequences. The audiovisual impact on multiwords is worth being investigated but beyond the present scope. Last but not least, this study aimed to explore the modality effect when using video as a repetition medium for vocabulary consolidation. The researcher wishes to emphasize that repeated exposure to target words requires no radical new methodology. In consideration of limited working memory, teachers may need to put some thought into how audiovisual support can be utilized in a way that reduces split-attention and cognitive load. The purpose of this paper has been to raise such awareness.

References

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 344-349.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, *23*(4), 509-533.

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*, 25-44.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*(5), 170-178.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28-71.

Clark, R., & Mayer, R. E. (2016). E-learning and the science of instruction (4th ed.). San Francisco: Pfeiffer.

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

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*, 105-134.

Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Methods, Research, Instruments, and Computers, 1*, 1-15.

Heatley, A., Nation, I. S. P., & Coxhead, A. (n.d.). RANGE [Computer software]. http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Hoffman, B. (2006). *The encyclopedia of educational technology.* San Diego, CA: Montezuma Press.

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in Psychology, 4*, 149.

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology, 2*, 17.

Jensema, C. J., Danturthi, R. S., & Burch, R. (2000). Time spent viewing captions on television programs. *American Annals of the Deaf, 145*(5), 464-468.

Mayer, R. E. (2001). *Multimedia learning.* Cambridge: Cambridge University Press.

Mayer, R. E. (2003). Elements of a science of e-learning. *Journal of Educational Computing Research, 29*(3), 297-313.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R. E. (Ed.). (2014). *The Cambridge handbook of multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of the dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484-490.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*, 444-452.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-320.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.

Nation, I. S. P. (2012). The BNC/COCA word family lists 25,000. http://www.victoria.ac.nz/lals/about/staff/paul-nation

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

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

Rummer, R., Schweppe, J., & Martin, R. (2009). A modality congruency effect in verbal false memory. *European Journal of Cognitive Psychology, 21*(4), 473-483.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159-168). New York: Cambridge University Press.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology,* 14(2), 50-73.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals,* 38(3), 422-427.

Wickens, C. D. (2007). Attention to the second language. *IRAL, 45*(3), 177-191.

Winke, P., Gass, S. M., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*, 66-87.

Winke, P., Gass, S. M., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *Modern Language Journal, 97*(1), 254-275.

Appendix

Transcript of one video clip.
Capital punishment in Utah (88 seconds)
https://www.newsinlevels.com/products/capital-punishment-in-utah-level-3/

Lawmakers in Utah have voted to bring back executions by firing squad if lethal injections are not readily available. The news comes as a number of US states struggle to obtain lethal injection drugs amid a nationwide shortage and concerns over their effectiveness.

European manufacturers have refused to sell the concoctions to US prisons and corrections departments over opposition to the death penalty. Many states have been led to consider alternative methods as supplies dwindle.

Texas is said to have only enough drugs on hand to perform two more executions while the head of Utah's prison system has said the state does not currently have any. Supporters of the legislation say three states – Oklahoma, Ohio and Arizona – recently carried out lethal injections that led to inmates' physical distress and drawn-out deaths. They claim death by firing squad is more humane.

Opponents, however, say it's a cruel holdover from the state's Wild West days and will earn it international condemnation. If approved by Governor Gary Herbert, the move would make Utah the only state in the country to permit the practice. It used firing squads for decades before adopting lethal injections in 2004.