Wong, L.-H. et al. (Eds.) (2013). Proceedings of the 21st International Conference on Computers in Education. Indonesia: Asia-Pacific Society for Computers in Education

The Effects of Different Presentation Modes of Multimedia Annotations on Sentential Listening Comprehension

Chih-Cheng LIN* & Ying-Chieh WU
Department of English, National Taiwan Normal Univsrsity, Taiwan
cclin@ntnu.edu.tw

Abstract: Multimedia annotations of both definitions and visual aids have been reported to facilitate vocabulary learning based on learners' performances on word retention. Vocabulary auditory input, however, was overlooked in many studies; listening comprehension, likewise, was seldom assessed as learning outcomes. The purpose of the present study was to explore the effects of vocabulary auditory input and those of learning style preference on vocabulary learning, in general, and on sentential listening comprehension, in particular. English beginners, 423 in total, were recruited from various junior high schools in Taiwan. Five nouns and five verbs, selected as the target words, were embedded in a reading text and annotated by one of the four methods: text-only, text-picture, text-sound, and text-picture-sound. One month before the treatment all participants were required to take a pretest of the target words; and, their learning style preferences, including verbal, visual and auditory, were determined by a questionnaire. In the treatment session every participant was randomly assigned to one of the four annotation groups aforementioned and read the text. Immediately after the reading they took a vocabulary recognition test and a listening comprehension test; the two tests were administered again two weeks later without prior notice. The data was submitted to two-way repeated measures ANOVA, with annotation type and learning style as between-subject factors, time of measurement as within-subject factor, and scores of the two tests as dependable variable. The results showed that in the recognition tests, none reached significance level but time of measurement, with the immediate higher than the delayed. In the listening tests, only the main effect of annotation type reached significance level, with text-picture-sound group and text-sound group both outperformed text group; no differences were found between the immediate and the delayed posttests. While various annotations had equivalent effects on vocabulary learning, annotations with audio input contributed to the construction of phonological knowledge of new words, facilitating their listening comprehension in sentences. More importantly, the effects of audio input sustained for two weeks. The learning style preference of our English beginners, whether verbal, visual or auditory, played no role in vocabulary recognition and listening comprehension.

Keywords: Multimedia annotations, auditory input, listening comprehension, learning style preference

1. Introduction

Multimedia annotations for vocabulary learning and teaching have become a research interest for almost two decades (Chun & Plass, 1996; Plass, Chun, Mayer & Leutner, 1998). The major findings of the previous studies supported dual coding theory (Pavio, 1986) and the generative theory of multimedia learning (Mayer, 1997). That is, learning vocabulary with textual definitions and pictures produced the best learning outcomes (Shahrokni, 2009; Yoshii, 2006; Yoshii & Flaitz, 2002) whereas the effects of learning new words with video clips are still inconclusive (Al-Seghayer, 2001; Chun & Plass, 1996; Lin & Tseng, 2012). One of the crucial aspects of vocabulary learning, auditory input, has been well studied in relation to captions (Aldera & Mohsen, 2013; Hsu et al., 2013, Syndorenko, 2010; to name the latest). However, auditory input has received comparatively little attention in multimedia annotations (Chun & Plass, 1996; Yeh & Wang, 2003) even though its role in vocabulary learning and acquisition

is inseparable (Nation, 2001); subsequently, the assessment of vocabulary learning in terms of listening comprehension has been scant. The imbalance found in this trend of research on auditory input for vocabulary learning and on listening comprehension for vocabulary acquisition triggered the present investigation. Other than the presentation modes of multimedia annotations, the learning style preference of a learner is also considered a crucial factor. When learning with multimedia materials, learning outcomes are likely dependent upon learners' multimedia preferences. Verbalizers prefer learning with text and visualizers with images (Plass, Chun, Mayer & Leutner, 1998; Chen, Hsieh & Kinshuk, 2008); but, those who favor auditory input are not yet examined in empirical studies. The present study was mainly intended to investigate the effects of auditory input of new words on vocabulary learning and on comprehending sentences containing the new words; it also examined whether language learners' multimedia preferences affected their learning vocabulary and their comprehension.

2. Vocabulary Learning and Multimedia Annotations

Learning vocabulary has been a primary task for language learners. Language learners have been suggested to do extensive reading to expand vocabulary size incidentally (Krashen, 1989). However, they may run the risk of wrong guessing and inference of word meanings from the context in the reading process (Shahrokni, 2009; Yoshii & Flaitz, 2002). Researchers thus become interested in exploring the effectiveness of glossing in assisting vocabulary learning in reading activities. Watanabe (1997) suggested that marginal glosses facilitated incidental vocabulary learning; Kost, Foss and Lenzini (1999) stated that looking up printed text-plus-picture annotations resulted in better vocabulary acquisition. With the development of computer technology, there are more studies investigating the effects of different modes of multimedia annotations on word learning and revealing the beneficial effects of dual-modal annotations (Shahrokni, 2009; Yoshii, 2006; Yoshii & Flaitz, 2002).

The facilitative effect of presenting multimedia annotations in dual modes is supported by dual coding theory (Pavio, 1986) and generative theory of multimedia learning (Mayer, 1997). According to Pavio (1986), learning is enhanced when verbal and nonverbal messages are processed in working memory simultaneously. The generative theory of multimedia learning (Mayer, 1997) posited that better retention is made possible if learners select, organize, and integrate incoming messages presented in verbal and visual modes. In accordance with the theories, provision of input in dual modes in multimedia annotations were suggested and encouraged for a desirable learning outcome.

2.1 Learning new words with definitions, pictures and films

Since theories have suggested that presentation of both verbal and visual messages enhances learning, there are abundant studies on exploring the effectiveness of supplying varied visual input such as pictures and videos in addition to textual input in multimedia annotations. The series of studies on multimedia annotations mainly began with the investigation of the effects of annotations incorporating translations and still pictures on incidental word learning in reading tasks because pictures can be easily presented in a paper-based or computer-mediated learning environment. Kost et al. (1999) examined the effects of reading printed glosses showing textual and pictorial materials on vocabulary learning. The participants were randomly assigned to read a German text under three glossing conditions, including textual glossing, pictorial, and textual-plus-pictorial glossing. The results, corroborating dual-coding theory (Pavio, 1986), showed that the learners accessing both textual and pictorial input recalled word meanings better. They retained the meanings of the new words after a lapse of time. The researchers thus concluded that building verbal and visual representations of new words resulted in enhanced mapping of the new information in mental model. Additional pictorial input provided more cues for retrieval of word meaning, thus enhancing learning and retaining of new words.

By replicating the study done by Kost et al. (1999), Yoshii and Flaitz (2002) explored the effectiveness of multimedia annotations in facilitating learning verbs. The college students were at the beginning level and were randomly assigned to read an English text by consulting English definitions, simple-line drawings, or both materials in glosses. In line with Kost et al. (1999), it was found that presenting dual modes of glosses facilitated word learning. The students exposed to only the pictorial input had better performance on recalling word meanings than those exposed to only textual input. Yet, the superiority disappeared in the definition-supply test with a strict scoring system. Yoshii and Flaitz then contended that the pictorial input may not fully convey word meanings, which in contrast were likely to be achieved effectively by textual input.

After the study on verbs, Shahrokni (2009) particularly examined whether the use of multimedia annotations enhanced the learning of concrete nouns. The college students at the beginning level were recruited to read an online passage and were randomly assigned to an annotation group, namely text only, picture only, and combination group. Definitions in the target language and real-life pictures were regarded as verbal and visual input respectively. The results, extending the early findings, revealed that dual-modal multimedia annotations were conducive to learning nouns. It concluded that the students reading both verbal and visual messages retained more words, regardless test type; specifically, they did well on both word and picture recognition tests.

Questioning the validity of vocabulary assessments, Yanguas (2009) argued that word tests alone did not serve as satisfying measures because they did not reveal learners' learning behavior. He applied think-aloud protocols in addition to word tests to examine the effectiveness of multimedia annotations. The participants were randomly assigned to text-only, picture-only, and combined group. Unlike many previous studies, the results indicated that the combined group did not significantly outperform the text-only and the picture-only group. In other words, the finding did not support generative theory of multimedia learning (Mayer, 1997). The effectiveness of multimedia annotations in this qualitative study was not evident because learners made more efforts to comprehend the reading text instead of attending to the learning of specific words.

Other than studies investigating the effects of multimedia annotations by presenting drawings or pictures visual input, there were studies focusing on examining videos or animations as visual stimuli. In the study conducted by Chun and Plass (1996), three types of multimedia annotations were designed, including definition only, definition plus picture, definition plus video, and were all annotated in the reading text. The participants were allowed to consult these three types of annotations at will. Interestingly, it was found that the availability of multimedia annotations induced more look-up behavior. The results showed that the consultation of additional pictures instead of videos contributed to better word learning. Chun and Plass contended that it may be because the learners were allowed more time to process the new information when they saw the pictures; thus, better learning outcomes.

Using same multimedia annotations, Al-Seghayer (2001) conducted a similar study. The finding confirmed the generative theory of multimedia learning (Mayer, 1997). He agreed that presenting dual-modes of multimedia annotations was beneficial for vocabulary; but, unlike Chun and Plass (1996), he found that better learning outcomes were achieved by using additional videos not pictures. Dynamic input was suggested to be of merits to facilitate "conceptualizing language," which meant the mapping of meaning to language form, because videos "more readily depict connections or provide a gestalt" (Al-Seghayer, 2001, p.224).

Different from the previous studies, Lin and Tseng (2012) incorporated animations and films in the multimedia annotations. They intended to examine whether presenting text, text plus picture, and text plus animation or video in multimedia annotations lead to different outcomes in the learning of nouns containing difficult concepts. The results revealed the beneficial effects of providing multimedia annotations; in line with Al-Seghayer (2001), viewing animations or videos was found to help to learn difficult words. In contrast, Lin and Tseng argued that presenting additional pictures failed to enhance word learning because of the cognitive overloads imposed on the students.

Apart from research specifically investigating varied supplementary visual input, there are studies examining the effects of presenting triple-modes of multimedia annotations on incidental vocabulary learning. Audio input was particularly provided in the studies. Yeh and Wang (2003) probed into in what way the college students in Taiwan benefited from different multimedia annotations presenting text, text plus pictures, text plus pictures and word sound. The college students were reported to have benefited from consulting annotations in which text and pictures are presented

instead of annotations that displaying text, pictures, and word sound. The authors explained that the students' limited listening ability degraded the positive effect of presenting audio input. Their study also showed that the students' learning style preference did not affect the effectiveness of different types of annotations in enhancing incidental vocabulary learning.

Individual difference has been acknowledged to have an influence on the effectiveness of instructional designs. As a result, some researchers further attended to the role individual difference plays along the line of studies on multimedia annotation. Yoshii and Flaitz (2002) indicated that both learners at beginning and intermediate level may benefit from the consultation of multimedia annotations. Yeh and Wang (2003) found that the effectiveness of varied multimedia annotations was not impacted by the college students' learning style preference. On the other hand, the findings in the study done by Pass et al. (1998) revealed that visualizers were more likely to have better retrieval of word meanings if they were able to trace images. In addition, Chen, Hsien and Kinshuk (2008) found that learners of higher visual yet lower verbal ability especially benefited from multimedia annotations presenting pictures, while learners with higher verbal ability were able to learn well by accessing merely textual information. Although the results was not conclusive, it cannot be denied that individual difference have a role to play in affecting the effectiveness of multimedia annotations in word learning.

2.2 The present study

The concerns of the present study are audio input as one type of the multimedia annotations, sentential listening comprehension as one assessment of the learning outcomes, and learning style preference as one factor of learning effects. As discussed before, the positive effects of presenting dual modes of multimedia annotations on facilitating vocabulary learning have been acknowledged. Few studies examined in what way learners benefit from multimedia annotations incorporating additional audio input to learn vocabulary. Audio input such as word pronunciation was not present in most early studies. In effect, informing word pronunciation is of value because phonological knowledge of a word along with word form and meaning are aspects of word knowledge (Nation, 2001). For beginners, word pronunciation may be important input when they need to familiarize themselves with the rules or systems of a new language. In terms of learning outcomes, most studies designed their assessments by asking learners to choose or provide word meanings. Whether learners differed their levels in acquiring new words, for example, comprehending sentential oral descriptions based on knowledge of newly-acquired words, is not clear. Finally, the mediating effect of individual difference on the effectiveness of various multimedia annotations is not yet conclusive. Yeh and Wang (2003) found that learning style preference did not impact the way English students at tertiary level benefited from multimedia annotations. However, it remains unknown if English beginners' learning style preference affects the effectiveness of different types of multimedia annotations in assisting learning. Therefore, the purpose of the present study is to investigate the effects of different types of multimedia annotations incorporating textual, pictorial, and audio input and those of different learning preferences on facilitating vocabulary learning as well as comprehension of sentential oral descriptions. The research questions are addressed as follows: Which type of multimedia annotations facilitates vocabulary learning? Which type of multimedia annotations facilitates sentential listening comprehension? And, does learning style preference affect vocabulary learning? Does learning style preference affect sentential listening comprehension?

3. Method

3.1 Participants

The participants were 423 seventh and eighth graders in five junior high schools in northern and southern Taiwan. During the experiment, they were considered English beginners who needed textual and auditory input when learning new vocabulary.

3.2 Materials

3.2.1 Reading text

The participants were required to read an annotated online text on dancing. The reading text adapted from various sources by the researchers contained 185 words in length and its readability was 3.0 in accordance with Flesch-Kincaid Grade Level.

3.2.2 Target words

Five nouns and five verbs in the reading text were chosen as the target words. Webb (2005) stated that nouns and verbs were the most often seen parts of speech in unmodified reading passages. The ten target words were on the English reference word list at senior high level recommended by College Entrance Examination Center (CEEC) in Taiwan. They were of pedagogical value and were unknown to the participants. Each target word was annotated in the reading text explained immediately below.

3.2.3 Annotations

To explore the effects of different modes of annotations on vocabulary learning and sentential listening comprehension, four types of multimedia annotations were designed. First, text-only annotations (the Text Group) showed textual information including spelling, translation, and an example sentence for each target word. Second, text-plus-picture annotations (the Picture Group) presented the textual information mentioned above and pictures illustrating each example sentence. Third, text-plus-sound annotations (the Sound Group) provided the textual information and pronunciation of each target word and each example sentence. Fourth, text-plus-picture-plus-sound annotations (the Combined Group) provided the textual information, pictures, and pronunciation of the words and example sentences.

3.3 Instruments

3.3.1 Vocabulary pretest

The vocabulary pretest was designed to assess the participants' background knowledge of the target words. The students were required to provide definitions of the words that they knew; otherwise, they checked "I don't know" next to the words.

3.3.2 Perceptual English learning style questionnaire

To determine the participants' learning style preference in this study, a modified version of perceptual English learning style survey was designed based on Perceptual Learning Preferences Survey proposed by Kinsella (1993) and Ko (2002). There were 21 statements in the questionnaire, each of which required the participants to rate on a 4 Likert-scale of frequency from always (4) to never (1).

3.3.3 Vocabulary recognition test

The recognition vocabulary test was in a multiple-choice format. The participants were asked to choose the correct Chinese translation for each target word. For each question, the correct answer was presented along with two other distracters. The questions in the delayed recognition test were the same with those in the immediate test except that the order was changed.

3.3.4 Listening comprehension test

The listening comprehension test on the ten target words was designed to assess the participants' ability to comprehend sentential oral descriptions. Each question contained a sentence made with one target word and read by a native speaker; and, each question was presented with three colorful pictures. The participants were asked to listen to each sentence and to choose the picture that correctly

depicted the sentence they heard. The questions in the immediate and delayed listening comprehension tests were identical except that the order of the questions was changed.

3.4 Procedure

The vocabulary pretest and perceptual learning style preference survey were administered one month before the treatment. The participants were first identified to be learners of verbal, visual and auditory inclination based on their responses on the learning style preference questionnaire. Then the participants reporting identical learning style preference were randomly assigned to one of the four annotation groups, namely, Text, Picture, Sound, and Combined. In the treatment session, the participants were asked to read an online text and to consult multimedia annotations. Right after the treatment, they took an unexpected vocabulary recognition test and listening comprehension test. Two weeks later, they took the two same tests without prior notice.

3.5 Data analysis

The present study looked into the effects of different modes of multimedia annotations on facilitating vocabulary learning and listening comprehension. In addition, it examined whether learners' learning style preference affected the effectiveness of multimedia annotations. The scores on both the vocabulary recognition tests and the listening comprehension tests were analyzed by using two-way repeated measures ANOVA with annotation type (4 levels, Text, Picture, Sound and Combined) and learning style preference (3 levels, Verbal, Visual and Auditory) as two between-subject factors and time of measurement (2 levels, Immediate and Delayed) as within-subject factor.

4. Results and Discussion

4.1 Results

4.1.1 Pretest of the target words

All participants were required to take the pretest that assessed their knowledge of the target words. To make sure that the target words were unknown to the sampled participants, those who scored one point or more in the pretest were excluded from the study. That left 423 participants who received zero point on the pretest, which ensured the homogeneity of the sampled participants.

4.1.2 Vocabulary recognition posttests

The descriptive statistics of the two recognition posttests were shown in Table 1.

Table 1: Descriptive Statistics of Vocabulary Recognition Posttests

Annotation Type	Learning Preference	N	Immediate Posttest		Delayed Posttest	
			M	SD	M	SD
Text	Verbal	25	6.96	2.406	6.84	2.641
	Visual	52	7.69	1.925	7.10	2.724
	Auditory	28	7.18	2.881	6.68	2.816
	Total	105	7.38	2.326	6.92	2.709
Picture	Verbal	26	6.77	2.438	6.65	2.560
	Visual	49	7.20	2.731	6.86	2.739
	Auditory	26	8.08	2.208	7.46	2.687
	Total	101	7.32	2.553	6.96	2.672
Sound	Verbal	26	7.46	2.596	7.58	2.671
	Visual	45	7.56	2.554	7.20	2.881
	Auditory	36	7.94	2.254	7.64	2.789
	Total	107	7.66	2.453	7.44	2.782

Combined	Verbal	26	7.85	2.493	8.12	2.776
	Visual	50	7.68	2.591	7.46	2.525
	Auditory	34	8.18	1.930	7.41	2.476
	Total	110	7.87	2.370	7.60	2.564

In the immediate posttest, the Combined Group attained the highest mean (M = 7.87, SD = 2.370), followed by the Sound Group (M = 7.66, SD = 2.453), the Picture Group (M = 7.32, SD = 2.553), and the Text Group (M = 7.38, SD = 2.326). In the delayed posttest, the results showed that the Combined Group performed the best (M = 7.60, SD = 2.564), followed by the Sound Group (M = 7.44, SD = 2.782), the Picture Group (M = 6.96, SD = 2.672), and the Text Group (M = 6.65, SD = 2.560).

A two-way repeated measures ANOVA revealed no interaction between annotation type and learning style preference ($F(6,411) = 0.705$, *n.s.*), no effect of annotation type ($F(3,411) = 1.997$, *n.s.*) and no effect of learning style preference ($F(2,411) = 0.526$, *n.s.*). The differences between the immediate and the delayed posttests were significant ($F(1,411) = 7.431$, $p < .05$), with the scores of the immediate posttest higher than those of the delayed one, suggesting a forgetting slope over two weeks.

4.1.3 Listening comprehension posttests

Table 2 showed the descriptive statistics of immediate and delayed listening comprehension tests; and, same rankings were found as those in the vocabulary recognition posttests. In the immediate posttest, the Combined Group attained the highest mean (M = 7.06, SD = 2.177), followed by the Sound Group (M = 6.77, SD = 2.226), the Picture Group (M = 6.62, SD = 2.144), and the Text Group (M = 6.12, SD = 2.269). In the delayed posttest, the results showed that the Combined Group achieved the best listening comprehension among the four groups (M = 7.10, SD = 2.475), followed by the Sound Group (M = 6.79, SD = 2.706) and the Picture Group (M = 6.65, SD = 2.535); and, the Text Group came the last (M = 6.07, SD = 2.486).

Table 2: Descriptive Statistics of Listening Comprehension Posttests

Annotation Type	**Learning Preference**	**N**	**Immediate Posttest**		**Delayed Posttest**	
			M	**SD**	**M**	**SD**
Text	Verbal	25	6.00	2.062	6.16	2.495
	Visual	52	6.42	2.387	6.35	2.504
	Auditory	28	5.68	2.212	5.46	2.426
	Total	105	6.12	2.269	6.07	2.486
Picture	Verbal	26	6.23	2.550	6.04	2.676
	Visual	49	6.59	2.081	6.78	2.486
	Auditory	26	7.08	1.787	7.04	2.474
	Total	101	6.62	2.144	6.65	2.535
Sound	Verbal	26	6.58	2.469	6.96	2.569
	Visual	45	6.73	2.359	6.78	2.575
	Auditory	36	6.94	1.897	6.67	3.014
	Total	107	6.77	2.226	6.79	2.706
Combined	Verbal	26	7.19	2.079	7.58	2.335
	Visual	50	6.62	2.294	6.84	2.518
	Auditory	34	7.62	1.985	7.12	2.532
	Total	110	7.06	2.177	7.10	2.475

A two-way repeated measures ANOVA revealed no interaction between annotation type and learning style preference ($F(6,411) = 1.139$, *n.s.*) and no effect of learning style preference ($F(2,411) = 0.068$, *n.s.*), but a main effect of annotation type ($F(3,411) = 4.491$, $p < .05$). *Post hoc* LCD pair-wise comparisons revealed that the Combined Group and the Sound Group scored significantly higher than

the Text Group, indicating that the vocabulary auditory input facilitated sentential listening comprehension. As for the time of measurement, the differences between the two posttests were not significant ($F(1,411) = 0.006$, *n.s.*), suggesting a strong retention of auditory input over two weeks.

4.2 Discussion

4.2.1 Vocabulary learning

The present study found that two-mode and multiple-mode of annotation were not reported to contribute to significant better performance, contradicting to the results in previous studies (Lin & Tseng, 2012; Shahrokni, 2009; Yoshii and Flaitz, 2002). Textual information alone was argued to facilitate learning if translation was clear and precise enough. It is likely that Chinese translation presented in the four types of annotation in the present study well-informed our English beginners of word meanings. Also, Lin and Tseng (2012) maintained that multimedia annotations incorporating translation and video were conducive to learning nouns conveying difficult cultural concepts. In the present study, the target words were selected based on the word list for senior high school students. Though they may be slightly beyond the participants' current ability to learn, they were not really beyond the students' ability to understand the word meaning with the help of Chinese translation.

Despite the fact that multimedia annotation did not significantly contribute to superior performance in recall and retention of word meaning, its effectiveness in enhancing vocabulary acquisition should not be neglected. In the present study the Combined Group still scored the highest in the two recognition posttests. Compared to the students receiving only textual information, the learners accessing additional pictorial and audio stimuli better recalled word meanings because they were more likely to trace the images in the recognition word tests (Yoshii, 2006). Additionally, the presence of word sound may also allow the learners to attend to the associations of word form and word meaning. It is delightful to know that the Sound Group obtained higher score than the Text and the Picture Group in the two recognition posttests. This reemphasizes the inseparable role of sound in vocabulary learning. For our beginners, not being able to associate word forms with their sounds means incomplete learning of new words; the presence of word pronunciation helps our beginning adolescents acquire word meanings.

4.2.2 Sentential listening comprehension

The finding that learners in the Combined Group and the Sound Group achieved better sentential listening comprehension than those in the Text Group strengthened the potential and value of multimedia annotation in English learning. Multimedia annotations that incorporate audio-plus-pictorial or audio messages in addition to textual messages assisted learners in acquiring and retaining new words efficiently. More importantly, they are probable to facilitate integration of vocabulary knowledge of new words into learners' linguistic system and further induce better performance on a complicated listening task which required learners to complete by activating knowledge of the previously-learned words.

The superiority of multimedia annotations incorporating textual, pictorial, and audio messages in facilitating listening comprehension confirmed Mayer's generative theory of multimedia learning (1997). Desirable learning outcomes could be achieved because learners construct verbal and visual representations of a word (Mayer, 1997) and thus had more than one route for retrieval of new words (Chun & Plass, 1996). Moreover, Brett (1995) contended that learning was enhanced because concurrently presented verbal, visual, and audio input in multiple annotations may "be more finely tuned" and "be more readily converted to "intake"" (p.81). In other words, the multiple messages available in the Combined Group were more likely to be integrated into the learner's existing linguistic system and were stored in long-term memory. In this sense, the incoming information introducing the new words was not only input but becomes part of the schema, which could be activated for dealing with a new task. In addition, connecting input presented in varied modal allowed learners to recall knowledge from long-term memory based on multiple cues (Baggett, 1989).

The fact that the Picture Group did not have significant better performance than the Text Group in the listening tests was quite surprising. It has been acknowledged that presenting verbal and

visual information enhanced vocabulary learning (Chun & Plass, 1996; Kost, Foss, & Lenzini, 1999; Shahrokni, 2009; Yoshii & Flaitz, 2002). It seemed that the superiority of dual-modal annotations incorporating textual and pictorial messages in enhancing learning disappeared when the goal was to accomplish tasks requiring listening skills. The results indicated that although the connection between textual and pictorial information of a word contributed to the retrieval of word meanings, it failed to help our beginners to construct the spoken form of a new word. Baddeley (1992) maintained that the channel in which articulatory messages were processed was necessary for acquiring native and second-language vocabulary. This suggested that supplementary audio input is particularly essential and is likely to amplify the effectiveness of multimedia annotations. The results in the present study further revealed that multimedia annotation did not assist our English beginners in acquiring spoken form of a new word without the presence of audio input.

In addition to the Combined Group, the Sound Group was reported to outperform the Text Group in the listening comprehension posttests. It is worth noting that the learners benefited from multimedia annotations particularly providing audio input and then achieved outstanding listening comprehension. In the discussion of effectiveness of multimedia annotations, test type was suggested to influence performance on vocabulary test to certain extent. Kost et al. (1999) claimed that there was corresponding effect of annotation type and test mode. That is, students accessing textual information recalled word meaning of the target words better when they were given translations in the multiple-choice test, whereas students accessing graphical information showed better retention when they were prompted with graphic cues. In line with Kost et al. (1999), Yoshii (2006) suggested that there was an interaction effect between annotation mode and test type. He posited that the learners having access to both textual and pictorial messages were able to retrieve word images; thus, they outperformed those of single-mode in the definition-supply word test instead of the recognition test.

Likewise, the finding of the superiority of the Combined and Sound Group in the listening comprehension tests in the present study confirmed corresponding effect of annotation type and test mode suggested by Kost et al. (1999). The listening comprehension tests may favor learners accessing audio input. Lynch (1998) stated that phonetic, phonological, and lexical along with other four sources of information were needed for comprehension of listening messages. Learners must recognize word sound first so as to further make sense of an oral description. The audio information available in the annotations informed the learners of the word sound, that is, the phonetic information. As a result, learners are more likely to identify the spoken target words and then understand the oral descriptions. Moreover, our results showed that the English beginners received noticeably highest scores when they were concurrently presented with pictorial and audio input together with textual input. It might be because the accompanying pictures provided the learners with additional routes to recall word meanings (Kost et al., 1999) and to associate spoken word form with word image with ease. Our learners were able to identify and recall word meaning, sound, and images efficiently in a listening task.

5. Conclusion

The aim of the present study was to investigate in what way presenting different modes of multimedia annotations influenced recognizing new words as well as comprehending new words in a spoken sentence. Although the differences were not found in learning word meanings, annotation type had a significant effect on affecting performance of sentential listening comprehension. To be specific, learners consulting multimedia annotations that provided textual as well as audio messages or that presented textual, pictorial, as well as audio messages achieved better listening comprehension immediately after and two weeks after the treatment than those reading textual information. It was also found that learners' learning style preference did not significantly influence the effectiveness of different types of multimedia annotations in both vocabulary learning and sentential listening comprehension.

The findings in the present study reinforced the importance of incorporating audio input into multimedia annotations. English beginners tend to ignore correct word pronunciation and thus fail to associate spoken word form with written form. As a result, they are more likely to encounter greater barriers when they contrive to make connections between word form and meaning. Provision of

additional audio input such as word sound is suggested in that it informs learners of precise pronunciation and equips them with phonetic and phonological knowledge of new words. Provision of multimedia annotations incorporating textual, pictorial, and auditory messages should also be promoted because it enables learners to retrieve and retain not only word form but also word meaning efficiently.

The effectiveness of presenting multiple modes of multimedia annotations in assisting listening comprehension was made evident in the present study. Yet, there are still some limitations of the present experimental design. First, the participants were from junior high schools in northern and southern Taiwan. Future research need to investigate how learners at intermediate or tertiary level are assisted by using multimedia annotations in improving vocabulary recognition and sentential listening comprehension. Second, the present study assessed the learners' ability to comprehend sentential oral descriptions after they learned the new words from the contents presented in multimedia annotations. Future research can further investigate whether learners are able to comprehend longer oral passages, in which target words are essential to construct meanings.

References

Aldera, A., & Mohsen, M. (2013). Annotations in captioned animation: Effect on vocabulary learning and listening skills. *Computers & Education, 68*(1), 60-75.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology, 5*(1), 202-232.

Baddeley, A. (1992). Working memory. *Science, 255*, 556-559.

Baggett, P. (1989). Understanding visual and verbal messages. In H. Mandl & J. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 101-124). Amsterdam: Elsevier.

Brett, P. (1995). Multimedia for Listening Comprehension: The design of a multimedia-based resource for developing listening skills. *System, 23*(1), 77-85.

Chen, N. S., Hsieh, S. W., & Kinshuk. (2008). Effects of short-term memory and content representation type on mobile language learning. *Language Learning and Technology, 12*(30), 93-113.

Chun, J. M. & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*, 183-198.

Hsu, C.-K., Hwang, G.-J., Chang, Y.-T., & Chang, C.-K. (2013). Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices. *Educational Technology & Society, 16* (1), 403–414.

Kinsella, K. (1993). Perceptual Learning Preferences Survey. In J. M. Reid (Ed.) *Learning Styles in the ESL/EFL Classroom* (pp.221-224). Boston: Heinle & Heinle Publishers.

Ko, Y. W. (2002). *Perceptual style preferences and their relationship to English achievement and learning strategies of junior high EFL learners in Taiwan.* Unpublished master's thesis, National Kaohsiung Normal University, Kaohsiung, Taiwan.

Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals, 32*(1), 89-113.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. Modern Language Journal, 73, 445-464.

Lin, C. C. & Tseng, Y. F. (2012). Videos and animations for vocabulary learning: A study on difficult words. *The Turkish Online Journal of Educational Technology, 11*(4), 346-355.

Lynch, T. (1998). Theoretical perspectives on listening. *Annual Review of Applied Linguistics, 18*, 3-19.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Pavio, A. (1986). *Mental representation: A dual-coding approach.* New York: Oxford University Press.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*, 25-36.

Shahrokni, S. A. (2009). Second language incidental vocabulary learning: The effect of online textual, pictorial, and textual pictorial glosses. *TESL-EJ, 13*(3). Retrieved June 1, 2010, from http://www.tesl-ej.org/wordpress/issues/volume13/ej51/ej51a3/

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

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 19*(3), 287-307.

Webb, S. (2005). The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27, 33-52.
Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*, *13*(2), 48-67.
Yeh, Y., & Wang, C. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal, 21*(1), 131-144.
Yoshii, M. (2006). L1 and L2 glosses: their effects on incidental vocabulary learning. L*anguage Learning & Technology*, *10*(3), 85-101.
Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal*, *20*(1), 33-58.