IN THIS ARTICLE WE INVESTIGATE THE question of how effectively multimedia learning environments can assist second language (L2) students in the process of listening comprehension. In particular, we ask how written and pictorial annotations can aid learners in the acquisition of new vocabulary and in comprehension when they listen to an aural L2 passage. Researchers and practitioners in the field have long identified a need for a deeper understanding of listening comprehension in L2 teaching (e.g., Cauldwell, 1996; Field, 1997; Joiner, 1997; Lynch, 1998; Mendelsohn, 1998), as well as for a better understanding of how the attributes of multimedia can be utilized to enhance listening comprehension (e.g., Brett, 1995, 1997; Hoven, 1999; Joiner, 1997; Lynch, 1998; Meskill, 1996; Pusack & Otto, 1997; Salaberry, 2001). To date, the question of how pictorial and written information may help students’ aural comprehension and vocabulary acquisition in a multimedia environment has not been sufficiently addressed.

Listening comprehension in a L2 is the process of receiving, attending to, and assigning meaning to aural stimuli. It involves a listener, who brings prior knowledge of the topic, linguistic knowledge, and cognitive processes to the listening task, the aural text, and the interaction between the two (Coakley & Wolvin, 1986). Fischer and Farris (1995) viewed listening comprehension as a process whereby students actively construct a mental representation of an aural text based on prior knowledge of the topic and information found within. Hoven (1999) extended this definition by arguing that constructive learning from aural materials can also depend on the students’ interaction with the characteristics of the text, the interlocutor, the task, and the process.

In L2 education, teachers assign audio-based activities in order to develop students’ listening
skills. Typically, listening comprehension activities provided with language textbooks have presented only the aural component of the language with little to no pictorial or written supportive information. Therefore, students’ preferences or needs have often been ignored, potentially leading to poor comprehension. However, in recent years, numerous software packages have been developed by researchers (e.g., Larson & Bush, 1992; Otto & Pusack, 1992) as well as by authors for textbook publishers (e.g., Amon, Muyskens, & Omaggio Hadley, 2000; Heining-Boynton, Cowell, & Torres-Quíñones, 1999) that provide students with various multimedia listening comprehension activities. Otto and Pusack’s (1992) The Listening Tool allows students to browse through a selection of authentic French, German, or Spanish video documents and includes access to numerous embedded comprehension aids. Larson and Bush’s (1992) Montevidisco, a simulated adventure program, allows students among other activities, to visit a Mexican town and encounter a variety of situations with the help of a male or female assistant and to access an online dictionary and a transcript of the material. ¡Atrevete!, developed by Heining-Boynton, Cowell, and Torres-Quíñones (1999), provides different levels of Spanish language instruction with a mixture of aural, pictorial, and written presentations of the information and interactive exercises. Others (e.g., Brett, 1995, 1997; Chung, 1994) have also developed and tested listening comprehension activities in a multimedia environment.

The availability of such listening comprehension activities in a multimedia format raises the question of what effect these different presentation modes have on students’ aural information processing in a multimedia environment. This article therefore focuses on one particular listening strategy, namely the use of multimedia annotations for vocabulary words, and the effect of these pictorial and written annotations that accompany an aural text on students’ vocabulary acquisition and aural comprehension.

**RESEARCH ON LISTENING COMPREHENSION**

Over the last 30 years, educators have moved away from behaviorist strategies in language teaching and now approach language acquisition from a cognitive and constructivist perspective (Bacon, 1992; Chamot, 1995; Fischer & Farris, 1995; Long, 1989; O’Malley, Chamot, & Walker, 1987; Plass, Chun, Mayer, & Leutner, 1998). In this paradigm, listening comprehension activities are no longer restricted to passive drill-and-practice strategies but instead emphasize active construction of meaning based on prior knowledge, linguistic knowledge, interaction with and understanding of the text (Fischer & Farris, 1995; Joiner, 1986, 1997; Lynch, 1995; Mayer, 1992; Pusack & Otto, 1997). Researchers also encourage the use of authentic materials (those developed by native speakers for the native speaker audience) over pedagogically-designed materials because they present a whole, genuine, and realistic approach to the language and culture (Bacon, 1992; Bacon & Finneman, 1990; Joiner, Adkins, & Eykyn, 1989; Raplan, 1996). However, when students listen to a text on an unfamiliar topic, comprehension may be poor because prior knowledge of the material is very low. Thus, aural comprehension may be unattainable even for the best of students (Wolff, 1987). In this instance, it is paramount to supply learners with supportive information that helps them comprehend the aural passage (Garza, 1991). Students need additional hooks onto which they can “hang meaning and make sense of the aural stream” (Meskill, 1996, p. 184).

Research on technology-based support of listening comprehension has focused on different approaches to enhance students’ listening skills or to compensate for the lack of such skills. Applying Faerch and Kaspar’s (1986) notion that when prior knowledge of an aural text is low, meaningful interaction with the aural material is essential in order to help students construct meaning from technology-based environments, researchers have found that students’ comprehension improves when they have access to the tools they need to process an aural passage and when they can interact with the material (Garza, 1991; Joiner, 1986, 1997; Mendelsohn, 1998; Pusack & Otto, 1997; Raplan, 1996). For example, Guíllory (1998) reported that verbal information, provided in the form of captions, helped students better comprehend a video. She found that students receiving keyword captions could comprehend the aural texts equally as well as those seeing full text captions. Other researchers found that pictorial information could promote comprehension in listening activities, both as advance organizers and as annotations to the text (Chung, 1994; Herron, Hanley, & Cole, 1995; Mueller, 1980; Powlens, 1992; Severin, 1967). Several studies also stressed students’ use of appropriate listening strategies to enhance comprehension in an aural learning environment (Joiner, 1997; Lynch, 1995, 1998; Mendelsohn, 1995, 1998; Raplan, 1996).

Other findings that could inform the design of
strategies in listening comprehension come from research on cognitive processing of multimedia information in reading comprehension. Though listening involves a different perceptual channel than reading, information processing in reading and listening is similar (Rost, 1990) because both use symbolic representations in the form of modality-specific verbal codes and include similar cognitive processes that help learners develop mental models of the material (Chun & Plass, 1997a). However, to some extent, readers are better able to overcome the difficulty of a text than are listeners because they can select from familiar words found within the text to guide them through the passage and can use the stability of the written word to slow their rate and review challenging passages. Unfortunately, these strategies do not exist with the traditional audiotape (Kozma, 1991) or even with today’s unimodal digital sound files. The aural presentation of words is not as easily understood as is the written presentation, and students must somehow compensate for the lack of adequate bottom-up information (Field, 1998). One method that has been successfully used in reading comprehension in a multimedia environment is the introduction of pictorial and written annotations as aids that accompany a written text (Chun & Plass, 1996a, 1996b, 1997a; Garza, 1991; Wolf, 1987). Research on information processing of written texts in a multimedia environment suggests that when students work in such an interactive environment and access multiple annotation types (pictorial and written) that accompany a reading comprehension passage, learning is more likely to take place than when annotations are not available (Chun & Plass, 1996a, 1996b, 1997a; Plass, Chun, Mayer, & Leutner, 1998, in press). Will a similar outcome occur when processing information aurally in a multimedia environment? By making multiple presentation modes of information, (i.e., pictorial and written) available to learners for the processing of an aurally presented text, will multimedia effects similar to those in reading comprehension occur? A theoretical foundation for these questions on listening comprehension that has been successfully used in research on reading comprehension is provided by the Generative Theory of Multimedia Learning (Mayer, 1997, 2001).

LISTENING COMPREHENSION WITH MULTIMEDIA

Mayer’s (1997, 2001) Generative Theory of Multimedia Learning, based on Paivio’s (1971, 1986) Dual Coding Theory, posits that in order to comprehend a text meaningfully, students must select relevant written (“verbal”) and pictorial (“visual”) information from it, organize the written information into a coherent verbal mental representation, organize the pictorial information into a coherent visual mental representation, and then integrate these newly created mental representations with one another by building referential connections between them and integrating them into a mental model (see Figure 1).

Using Mayer’s (1997, 2001) theory, Plass, Chun, Mayer, and Leutner (1998) studied the effects of choice of annotations (pictorial and written) on comprehension and vocabulary acquisition from a German reading text. In their study, students could choose to see an English translation of a German word on the screen (written annotation) or to view a picture or video clip representing the selected word (pictorial annotation). Students performed best on recall protocol and vocabulary tests when they selected pictorial and written annotations while reading the computer-based text than when they accessed written annotations alone, pictorial annotations alone, or no annotations at all. Thus, learning is enhanced when students can select pictorial and written information, organize the information in working memory, build referential connections between pictorial and written information, and integrate them with their prior knowledge.

In a similar manner, Chun and Plass (1996a, 1996b) examined the effects of annotations with different forms of media on vocabulary acquisition and comprehension from a L2 reading passage. They found that pictorial information in addition to written information helped support micro- and macrolevel processing in L2 computer-based reading activities. When written and pictorial units were presented contiguously, the learners were able to hold the verbal and visual mental representations simultaneously in verbal and visual working memory. The two kinds of representations in memory allowed the students to build referential connections between the two systems, which resulted in multiple retrieval routes to the vocabulary items and thus additive effects on recall of the information.

Though not applied to a computerized multimedia environment, research has long indicated that processing supportive information (pictorial or written) with an aural text positively affects students’ listening comprehension (e.g., Baltova, 1999; Carlson, 1990; Chung, 1994; Guillory, 1998; Herron, Hanley, & Cole, 1995; Mueller, 1980; Pouwels, 1992; Severin, 1967). Severin (1967) de-
determined that students who processed a listening passage with pictures performed better on post-treatment activities than students who received sound only or sound and unrelated pictures. Mueller (1980) factored student proficiency levels into his study of listening comprehension in a multimodal environment and determined that less proficient students performed best when an image was present, but more proficient students showed very little difference between listening with or without images. He suggested that a single-mode approach is sufficient for high-prior-knowledge students, but that dually-coded information could help low-prior-knowledge learners fill in the gaps that would otherwise exist in their prior knowledge. Chung (1994) also examined the effect of the presence or absence of video or images on students’ comprehension of an aural passage. The students’ comprehension was higher when a passage was presented either aurally with a video or aurally with a single picture than when the material was presented aurally with multiple visuals or with no visuals present. Baltova (1999) examined the effects of English audio and French subtitles (reversed format) and French audio with French subtitles (bimodal format) on students’ comprehension of and vocabulary acquisition from a French video. Comprehension of the video did not differ significantly in the different presentation types, but the students learned significantly more vocabulary when the material was presented in the bimodal format. Likewise, Guillory (1998) examined the effects of various keyword captioning methods on students’ comprehension of a digitized video. The students who viewed the video with keyword captions performed just as well as those with full text captions available to them. This finding led her to conclude that “with smaller amounts of texts in the visual channel, learners are less likely to encoun-
This study examined the potential of multimedia annotations for enhancing listening comprehension skills. Specifically, we investigated how the choice of written and pictorial annotations in a listening comprehension activity affects students’ comprehension and acquisition of new vocabulary from an aurally-presented, authentic historic text.

Basing our argument on Mayer’s (1997, 2001) Generative Theory of Multimedia Learning, we argue that access to supportive information represented by pictorial annotations (visual mode) and written annotations (verbal mode) will increase students’ comprehension of an aural passage and their acquisition of new vocabulary. In particular, we propose two hypotheses:

1. Students who complete a listening comprehension activity that contains a choice of pictorial and written annotations will recall more vocabulary words from the listening passage than those who receive only one type of annotation (pictorial or written) or no annotations.

2. Students who complete a listening comprehension activity that contains a choice of pictorial and written annotations will recall more propositions from the listening text than those who receive only one type of annotation (pictorial or written) or no annotations.

These hypotheses are based on the idea that students can select from and build referential connections between pictorial and written information and therefore have multiple retrieval routes to the information, resulting in enhanced vocabulary acquisition and aural comprehension. Thus, when students choose more than one annotation type (pictorial and written), learning and subsequent retrieval are more likely; when they do not work with additional annotations, comprehension and vocabulary acquisition are less likely.

**Method**

**Participants and Design.** The participants in this study consisted of 171 students (59 males, 112 females) enrolled in a second-semester beginning French class at a university in the south central United States. All of the students were nonnative speakers of French and were fluent in English. The students participated in this study as a regular class activity. Their mean estimated French GPA at the time of testing was 2.92; 4.0 = A, 3.0 = B, 2.0 = C, 1.0 = D, 0.0 = F.

Each student completed a pretreatment vocabulary test to determine his or her knowledge of the vocabulary in this study. All students were found to have low prior knowledge of the vocabulary with an average score of 3 out of a maximum score of 25, M = 3.08, SD = 3.40.

A pretest/posttest control group (between-subject) design was employed in order to observe the effects of two factors, the absence or presence of written annotations and the absence or presence of pictorial annotations, on the students’ acquisition of vocabulary and on their comprehension of the aural passage. In this 2 × 2 design, the participants were randomly assigned to one of four treatments of the listening text: (a) with no annotations available, (b) with only written annotations available, (c) with only pictorial annotations available, and (d) with both pictorial and written annotations available. The participants in these four groups did not differ in their performance on the vocabulary pretest: control group N = 42, M = 3.02, SD = 3.68; pictorial annotations only group N = 41, M = 2.29, SD = 2.12; written annotations only group N = 44, M = 3.48, SD = 3.57; pictorial and written annotations group N = 44, M = 3.45, SD = 3.87; F(3, 167) = 1.11, n.s.

**Dependent Measures and Scoring.** The effects of the four treatments on the students’ vocabulary acquisition and comprehension of the aural pas-
sage were measured through a written multiple-choice vocabulary posttest and a written recall protocol posttest, which were both administered immediately after the treatment and again, as delayed tests, after 3 weeks. The vocabulary posttest, identical to the pretreatment vocabulary test, was made up of 25 of the 27 keywords present in the aural passage that were visible on the five computer screens of each treatment (see Figure 2). Five English distractors, along with the correct English translation, were provided for each French word in order to reduce the potential to guess the correct answers. Each correct response received one point, with a maximum possible vocabulary test score of 25 points. The internal consistency of the vocabulary test, determined using the split-half reliability method, was .82. The use of a written vocabulary test instead of an oral test may limit the external validity of this study somewhat, but it does not affect its internal validity given that the participants in all treatment groups took the same test.

The recall protocol comprehension test instructed the students to summarize, in English, the passage they listened to during their treatment activity. We used a recall test in English in order to avoid measuring the students’ language production competency in addition to their comprehension, a method used successfully in related studies (e.g., Chun & Plass, 1996b; Lee, 1986).

Two experienced instructors of French chose 63 propositions that represented the idea units of the passage. The test was scored based on the number of correct propositions given by each student, up to a maximum of 63 points; a proposition could only be scored one time. The interrater reliability of the recall protocol measure was .97.

Materials. A Power Macintosh G3 was used to develop the listening comprehension module using Adobe Premiere 4.2 (Adobe, 1994) and Authorware 4.0 (Macromedia, 1997). The apparatus for presenting the materials consisted of a 22-station Macintosh computer language lab, arranged in such a way that the students faced away from the center of the room and thus could view only their own computer screen.

The computer-based materials presented an aural reading, lasting 2 minutes, 20 seconds, of an authentic historic encounter between LaSalle and the Quapaw Indians as it was written in 1682 (Buzhardt & Hawthorne, 1993; see Appendix A). It was 331 words in length and was digitally recorded using the voice of a female native French speaker. This text was chosen by the French in-

In addition, there was an icon of a loud-speaker present on each screen. The students could drag a keyword to the speaker icon in order to hear its pronunciation.

The four treatments used in this study differed as follows. For the None group (control), no annotations for vocabulary words were available except for the option to listen to the pronunciation of the word. For the pictorial annotations group (pictorial annotations only), a camera icon was present to the left of the speaker icon. Students could drag a keyword to the camera icon to view a pictorial representation of the keyword. In the written annotations group (written annotations only), a text icon was present to the right of the speaker icon. Students could drag a keyword to the text icon to view a written translation of the keyword. The materials for the pictorial and written annotations group (both pictorial and written annotations) contained both the camera and the text icons, in addition to the speaker icon. The students could drag a keyword to either the camera or the text icon in order to view a pictorial or a written annotation of the keyword (Figure 2). The students could select either or both icons and therefore have access to both representations of a keyword.

The visual annotations consisted of 13 color photos and 14 color drawings. The text annotations were written in bolded, 14-point Helvetica font. The students were allowed to select any annotation available in their treatment for any number of times, before, during, or after presentation of each aural segment. Given that pictures do not require translations in order to be understood by the learner, the text annotations were translations of the vocabulary words into English, allowing for a comparison between the effectiveness of pictorial and written annotations. Similar decisions were made in other related studies where written annotations in the L2 were found to be much more difficult to understand than pictures (e.g., Chun & Plass, 1996a).

Procedure. The students participated in this study during their normally scheduled French classes. Classes met for 50 minutes per day, 3 times per week, and the entire procedure required two 50-minute class periods on two consecutive days (25 minutes in the first class period, 50 minutes in the second). On the first day, the students first filled out a paper-and-pencil demographic questionnaire at their own rate. They were then given 8 minutes to complete a vocabulary pretest. On the second day, the students were randomly assigned to one of the four treatment groups and were given 14 minutes to listen to the
passage. They were instructed to look up all annotations available in their treatment. They were then given 8 minutes to summarize everything they could remember from the passage and then 8 minutes to complete a vocabulary posttest, identical to the pretreatment test. Three weeks later, without any additional opportunities to listen to the aural passage, the students completed unannounced delayed vocabulary and comprehension tests during two consecutive class periods. The students were again given 8 minutes to summarize everything they could remember from the passage in a delayed recall protocol, and then, 2 days later, they were given 8 minutes to complete a delayed vocabulary posttest. Both delayed measures were identical to the immediate measures.

Results

Hypothesis 1: Students who complete a listening comprehension activity that contains a choice of pictorial and written annotations will recall more vocabulary words from the listening passage than those who receive only one type of annotation (pictorial or written) or no annotations.

The Generative Theory of Multimedia Learning predicts that the acquisition of vocabulary words while listening to a narration is best when both verbal (written) and visual (pictorial) information are processed. The construction of connections between the verbal and visual representations of the information will result in an increased likelihood that the learner will have two kinds of retrieval routes for the word, which is expected to result in better performance on a vocabulary test.

Immediate Vocabulary Test. A multivariate analysis of variance (MANOVA) was computed with the number of correct answers on the immediate vocabulary posttest and on the immediate recall protocol posttest as dependent measures and the presence or absence of pictorial and written annotations as the between-subjects factors. The results for the immediate vocabulary posttest revealed a main effect for written annotations, $F(1, 167) = 82.38, MSE = 1561.27, p < .001, \eta^2 = 0.243$, and for pictorial annotations, $F(1, 167) = 70.90, MSE = 1343.56, p < .001, \eta^2 = 0.209$, and an interaction effect for pictorial and written annotations, $F(1, 167) = 18.71, MSE = 354.61, p < .001, \eta^2 = 0.055$.

Post hoc comparisons (Tukey HSD) of the immediate vocabulary test scores of the four treatment groups revealed statistically significant differences, showing that student performance was lowest when no annotations were available ($N = 42, M = 8.10, SD = 4.3$), and highest when both pictorial and written annotations were available ($N = 44, M = 19.75, SD = 3.2$); see Figure 3. In fact, the students in the pictorial and written annotations group significantly outperformed all others (for None group, $p < .001$; for written annotations group, $p < .05$; for pictorial annotations group, $p < .05$). When either pictorial or written annotations were available (written annotations group: $N = 44, M = 17.02, SD = 5.6$; pictorial annotations group: $N = 41, M = 16.59, SD = 4.0$), the students’ performance was higher than when no annotations were available.

![FIGURE 3](https://example.com/figure3.png)

**FIGURE 3**
Immediate and Delayed Vocabulary Posttest Scores by Treatment Type
(for written annotations and for pictorial annotations groups, \( p < .001 \)), but the results for these two groups did not differ significantly from one another.

Delayed Vocabulary Test. A MANOVA with the number of correct answers on the delayed vocabulary posttest and on the delayed recall protocol posttest as dependent measures and the presence or absence of pictorial or written annotations as the between-subjects factors revealed for the delayed vocabulary posttest a main effect for pictorial annotations, \( F(1, 133) = 41.22, \) \( MSE = 740.38, \) \( p < .001, \eta^2 = 0.20, \) for written annotations, \( F(1, 133) = 21.11, \) \( MSE = 379.28, \) \( p < .001, \eta^2 = 0.105, \) and an interaction effect for written and pictorial annotations, \( F(1, 133) = 4.76, \) \( MSE = 85.58, \) \( p < .05, \eta^2 = 0.024. \) Due to attrition, the number of participants in the delayed posttests was 137, with \( N = 34 \) for the None group (control), \( N = 36 \) for the pictorial annotations group, \( N = 34 \) for the written annotations group, and \( N = 36 \) for the pictorial and written annotations group.

Post hoc comparisons (Tukey HSD) of the delayed vocabulary test scores of the four treatment groups revealed statistically significant differences, showing that when no annotations were present, the students’ performance again was the lowest \( (M = 6.33, SD = 3.69) \). When pictorial annotations \( (M = 12.45, SD = 4.20) \) or written annotations \( (M = 11.15, SD = 4.90) \) were available, the students’ performance was significantly higher than when no annotations were available (for written annotations and for pictorial annotations groups, \( p < .001 \)), but the two groups did not differ significantly from one another in performance. When both pictorial and written annotations were present, the students’ performance was highest \( (M = 14.08, SD = 4.02) \), and the difference to students with written annotations available was statistically significant \( (p < .05) \), whereas the difference to those with pictorial annotations available was not; see Figure 3.

In summary, we found both for immediate and for delayed vocabulary tests that the students’ performance was highest when both pictorial and written annotations were available and lowest when neither type of annotation was available. The students who had both written and pictorial annotations available had a higher performance than the students who had only pictorial annotations available in the immediate vocabulary posttest, but they did not differ in their performance in the delayed vocabulary posttest. These findings are in line with our hypothesis and with the predictions of Mayer’s (1997, 2001) Generative Theory of Multimedia Learning, though we did not predict that the effect of pictorial annotations would persist over time while that of written annotations would not.

Hypothesis 2: Students who complete a listening comprehension activity that contains a choice of pictorial and written annotations will recall more propositions from the listening text than those who receive only one type of annotation (pictorial or written) or no annotations.

This hypothesis is based on the Generative Theory of Multimedia Learning (Mayer 1997, 2001) which posits that text comprehension includes the processes of selecting relevant information, organizing information in working memory into meaningful verbal (written) and visual (pictorial) mental representations, and integrating these representations with one another and with prior knowledge in long-term memory. When pictorial and written annotations are available, referential connections can be established between the verbal and the visual representations in working memory and learning is enhanced.

Immediate Comprehension Test. The MANOVA for the immediate posttests reported above revealed for the immediate comprehension posttest a main effect for pictorial annotations, \( F(1, 167) = 70.02, \) \( MSE = 1137.20, \) \( p < .001, \eta^2 = 0.274, \) and a main effect for written annotations, \( F(1, 167) = 16.60, \) \( MSE = 269.52, \) \( p < .001, \eta^2 = 0.065. \) No interaction effect was found for these factors.

Post hoc comparisons (Tukey HSD) of the immediate recall protocol scores of the four treatment groups revealed statistically significant differences, showing that the students’ performance was lowest when no annotations were available \( (M = 3.2, SD = 2.9) \). Performance was highest when pictorial annotations were available \( (M = 9.2, SD = 4.0) \) and when pictorial and written annotations were available \( (M = 10.9, SD = 4.6) \); see Figure 4. Students with pictorial and written annotations performed significantly better than students who received written annotations \( (M = 6.52, SD = 4.3; p < .05) \) and those who received no annotations \( (p < .01) \). The difference in performance between students in the pictorial and written annotations group and students in the pictorial annotations group was marginally significant \( (p = .052) \). Examples of students’ recall in the no annotations group and the written and pictorial annotations group are included in Appendix B.

Delayed Comprehension Test. The MANOVA for the delayed posttests reported above revealed for
the delayed comprehension posttest a main effect for pictorial annotations—$F(1, 133) = 24.61$, $MSE = 243.18$, $p < 0.001$, $\eta^2 = 0.160$—on the students’ recall of propositions. There was no main effect for written annotations, nor was an interaction effect present for availability of written and pictorial annotations.

Post hoc comparisons (Tukey HSD) of the delayed recall protocol scores of the four treatment groups revealed statistically significant differences, showing that the students’ performance was lowest when no annotations were available ($M = 2.66$, $SD = 2.77$) and when written annotations were available ($M = 3.29$, $SD = 3.23$). The performance of the students in these two groups was not significantly different; see Figure 4. The students’ performance was highest when pictorial annotations were available ($M = 5.58$, $SD = 3.48$) and when pictorial and written annotations were available ($M = 5.78$, $SD = 2.89$), and the difference in performance between these two groups was not statistically significant. However, the main effect for the availability of pictorial annotations means that differences in comprehension were statistically significant ($p < .001$) for those students who received pictorial annotations (the pictorial annotations group and the pictorial and written annotations group) and those who did not (None group and written annotations group).

In summary, the participants’ comprehension in the immediate test was highest when they received pictorial annotations and written annotations. Both groups that received pictorial annotations understood the aural text better than the group that received written annotations; the None group, without any annotations, had the lowest comprehension. In the delayed test, the pattern of comprehension remained the same except for the difference in performance between the written annotations group and the None group, which was statistically significant in the immediate test but not in the delayed test. Likewise, the difference between the pictorial annotations group and the pictorial and written annotations group was (marginally) significant in the immediate test but did not reach significance in the delayed test. These findings are in line with our hypothesis and with the predictions of the Generative Theory of Multimedia Learning (Mayer, 1997, 2001), though, as for vocabulary acquisition, we did not predict that the effect of pictorial annotations would persist over time whereas the effect of written annotations would not.

**DISCUSSION**

The research questions addressed in this study concerned the use of multimedia annotations to aid learners in listening comprehension. In particular, we were interested in whether previous research in reading comprehension, which had found a multimedia effect for the availability of different types of annotations, could be replicated for listening comprehension activities. Basing our work on Mayer’s (1997, 2001) Generative Theory of Multimedia Learning, we hypothesized that students who completed a listening comprehension activity that contained both pictorial and written annotations would acquire more vocabulary and remember more ideas from the text than those who completed the listening task with sin-
ingle annotations (pictorial or written) or no annotations.

Vocabulary acquisition, in strictest terms, involves rote learning or bottom-up processing where students focus on memorizing definitions for words. This concept leads to the assumption that direct translations of L2 keywords should provide students with the microlevel information necessary for comprehension. However, Kellogg and Howe (1971) suggested that foreign words associated with visual imagery or objects are learned more easily than those without such comprehension aids. Plass et al. (1998) further contended that, in rote learning, retrieval of words may be more difficult than input and that multiple annotation types can facilitate this retrieval process by providing multiple access routes to the information in long-term memory.

The results of this study show that as the students listened to the aural passage, they acquired more vocabulary when they had both pictorial and written annotations available ($M = 19.75, SD = 3.2$) than when they had only written annotations ($M = 17.02, SD = 5.6$) or only pictorial annotations ($M = 16.59, SD = 4.0$) available. Vocabulary learning was lowest when the students had no annotations available for vocabulary words in the listening text ($M = 8.10, SD = 4.3$). The results of the delayed test show the same pattern as the results of the immediate test, with the exception that the difference in vocabulary learning between the pictorial and written annotations group and the pictorial annotations group was statistically significant in the immediate test but did not reach significance in the delayed test (pictorial and written annotations group $M = 14.08, SD = 4.02$; pictorial annotations group $M = 12.45, SD = 4.20$; written annotations group $M = 11.15, SD = 4.90$; None group $M = 6.33, SD = 3.69$).

These results are an indication that the effects of multimedia annotations on students’ vocabulary acquisition found in L2 reading comprehension activities can also be found in listening comprehension activities, thus extending Mayer’s (1997, 2001) Generative Theory of Multimedia Learning to listening comprehension. According to this theory, the students who selected from words annotated in both pictorial and written modes were able to build more referential connections between the verbal (written) and visual (pictorial) mental representations of the vocabulary words and were therefore able to recall more of the vocabulary in a follow-up vocabulary posttest than the students in the other three groups. In this study, the effect of written annotations on vocabulary acquisition in the immediate vocabulary posttest accounted for 24.3% of the variance, and the effect of pictorial annotations for 20.9% of the variance, both unusually strong effects. However, the effect of pictorial annotations in the delayed vocabulary posttest, administered 3 weeks after the immediate test, still accounted for 20.0% of the variance, whereas the effect of the written annotations accounted only for 10.5% of the variance, resulting in the disappearance of the differences in vocabulary acquisition between the pictorial annotations group and the pictorial and written annotations group that was found for the immediate vocabulary posttest. Such an effect has not been observed in previous research on L2 listening comprehension or reading comprehension, where the effect of written annotations is usually stronger than that of pictorial annotations (Chun & Plass, 1996a, 1996b; Plass et al., 1998).

We also hypothesized that students who completed a listening comprehension activity that contained pictorial and written annotations would recall the passage better than those who completed the passage with single annotations (pictorial or written) or no annotations available. In the immediate recall protocol posttest of our study, the students who listened to the aural passage and received pictorial and written annotations ($M = 10.9, SD = 4.6$) showed a marginally significantly better comprehension of the aural passage than students who received pictorial annotations alone ($M = 9.2, SD = 4.0$), and a significantly better comprehension than those who received written annotations ($M = 6.52, SD = 4.36$) or no annotations ($M = 3.2, SD = 2.9$). A similar pattern was found for the delayed test, except for the fact that the difference in performance between the written annotations group and the None group, which was statistically significant in the immediate test, did not reach statistical significance in the delayed test (pictorial and written annotations group $M = 5.78, SD = 2.89$; pictorial annotations group $M = 5.58, SD = 3.48$; written annotations group $M = 3.29, SD = 3.23$; None group $M = 2.66, SD = 2.77$). In addition, the difference between the pictorial annotations group and the pictorial and written annotations group that was (marginally) significant in the immediate test was not significant in the delayed test.

Similar to the results for vocabulary learning, these results for text comprehension are in line with the Generative Theory of Multimedia Learning (Mayer, 1997, 2001). According to this theory, the students who selected pictorial and written annotations were able to organize them
into meaningful visual and verbal mental representations in working memory, build referential connections between these mental representations of the annotations and the aural passage, and integrate them with one another into the mental model they constructed of the listening text. This connection between “an element in the text and some other element that is crucial to the interpretation of it” (Baggett, 1989, p. 108) provided the students with multiple retrieval cues to recall their knowledge from long-term memory. The results are also in line with Meskill’s (1996) argument that an increase in the different modes of input available along with the potential for active engagement in or interaction with the input will lead to greater integration of the aural message into a learner’s developmental system and that this integration will subsequently lead to better recall of the material. Faerch and Kaspar (1986) also argued that interaction is needed to help students gain meaning from a text; comprehension of aural texts is best ensured through interaction and negotiation. In this study, annotations in different types of presentation modes of information provided elaborations on the information in the aural passage, which allowed the students greater interaction with the listening text and thus further enhanced their learning (Chun & Plass, 1996a).

In the immediate recall protocol posttest, the effect of pictorial annotations accounted for 27.4% of the variance in comprehension, whereas the effect of written annotations accounted for only 6.5% of the variance. On the delayed recall protocol posttest, the effect of pictorial annotations still accounted for 16.0% of the variance, whereas the effect of written annotations was not statistically significant, accounting for only 0.2% of the variance. This pattern replicates our findings for vocabulary learning, where the effect of pictorial annotations accounted for 20.9% of the variance in the immediate test, and for 20.0% of the variance in the delayed test. The effect of written annotations, strong with 24.3% in the immediate test, accounted for only 10.5% of the variance in the delayed test. It appears from these findings that the pictorial annotations had a stronger and longer-lasting effect than the written annotations, both for vocabulary acquisition and for the comprehension of the aural passage.

It has been argued that images carry a structural message that complements the language presented (Baggett, 1989; Kozma, 1991) and that increases the efficiency of learning (Oxford & Crookall, 1990). Given that listening comprehension entails the construction of mental representations and interpretations, our results, indicating that pictures more strongly support such a process than words (Meskill, 1996), seem reasonable: “the mapping of pictures onto the mental model provides a stronger bond than the mapping of words due to the different representations of their information (analog vs. symbolic)” (Kost, Foss, & Lenzini, 1999, p. 99). Analog representations (images) can be directly mapped onto the mental model and are assumed to be language independent, whereas symbolic representations (text) are sequentially processed and demand “an indirect transformation between the symbolic representation of the text and the analog mental model” (Chun & Plass, 1997a, p. 8). This effect is described in Mayer’s (2001) theory as the multimedia effect, which predicts that information supplied in different presentation modes (such as pictures and text) will result in better learning than information that is presented in just one mode (such as text), as is the case for the group receiving only written annotations (i.e., text, with the aural passage, which also consists of text).

CONCLUSION

In this study, when the students listened to an aural passage, they understood the narration better and learned more vocabulary when they selected from the pictorial and written annotations that were available for keywords in the narration than when they selected only one type of annotation or none. The delayed tests of vocabulary and comprehension, however, showed that students learned best when pictorial annotations were available, independent of the availability of written annotations. According to the Generative Theory of Multimedia Learning (Mayer, 1997, 2001), the students organized the aural passage along with the relevant verbally (written) and visually (pictorial) annotated words into verbal and visual mental representations and integrated the newly-constructed pieces of information with one another and with their mental model. The process of comprehending the aural passage, therefore, focused on establishing relationships between the aural presentation of the material and the written and pictorial annotations available to students, which required both microlevel (vocabulary acquisition) and macrolevel processing (aural comprehension). We found in this study that for these processes, pictorial annotations were more important than written annotations for long-term retention. The results found
in this study replicate those of earlier studies on text comprehension in multimedia environments (Chun & Plass, 1996a, 1996b) and are in line with the Generative Theory of Multimedia Learning (Mayer, 1997, 2001).

Some limitations that are due to the nature of the quasi-experimental research method apply to our findings. For example, in our study it was necessary to limit the number of annotated words available for the students to select in order to obtain a reliable and valid measure of vocabulary acquisition. For an instructional use of an application like the present listening comprehension software, it would be preferable and more authentic for the students to be able to select any word from the text. In addition, the listening materials used in this study were narrated from an authentic written text and were not, due to the nature of the materials, part of an authentic communication. Our results may therefore be confounded by the fact that the narrated aural passage we used had similarities with a written text. A follow-up study might investigate whether the results found in this study can be replicated with authentic oral communication. Such a follow-up study could also use oral outcome measures, instead of a written measure, and could examine additional behavior data, such as lookup patterns and time spent accessing the annotations.

This study offers practical and theoretical implications. On the practical side, it provides evidence relevant to the design of multimedia instruction for L2 learning. In particular, this study suggests that learners should have options for viewing material in both a visual (pictorial) mode and a verbal (written) mode to support the process of aural comprehension, but it shows also that pictorial annotations have stronger effects on vocabulary acquisition and comprehension than written annotations. Because different students learn efficiently in different ways (Reinert, 1976), multimedia environments that provide both pictorial and written annotations of keywords may be most effective for learning because the students can choose the annotation modes that best suit their needs and preferences (Plass et al., 1998, in press) and can review this information more than once, thereby further reinforcing their learning (Chun & Plass, 1996a).

In this study, the students learned more vocabulary words and comprehended the text better when they could use annotations for individual vocabulary words while listening to a text passage. In other words, even though the annotations were designed to support lower levels of cognitive processing on the vocabulary level, we found an effect of these annotations on text comprehension, which requires higher levels of cognitive processing. Future studies should address the effectiveness of multimedia annotations that are designed to support top-down processing on the phrase or sentence level in comparison to the bottom-up processing supported in this study.

Concerning the selection of existing software or the design of new listening comprehension software, the results of this study suggest that pictorial annotations for words have a stronger effect on vocabulary acquisition and on comprehension than written annotations, and should therefore be made available in such software. Other research has shown, however, that students with different learning preferences and abilities require different types of annotations (Plass et al., 1998) and that students’ learning is facilitated best when they are given a choice of pictorial and written annotations (Plass et al., in press).

On the theoretical side, this study provides evidence for a Generative Theory of Multimedia Learning (Mayer, 1997, 2001) and its extension to listening comprehension in L2 acquisition (Plass et al., 1998) that suggests that the availability and choice of pictorial and written annotations in listening comprehension activities enhances students’ abilities to comprehend the material presented. That is, students listen to an aural passage and select relevant information from the available pictorial and written annotations. Students then organize the pictorial and written annotations as well as the text from the aural passage into visual and verbal mental representations in working memory and build referential connections between them. Finally, they integrate the visual and verbal representations and their connections into a mental model of the passage.

Future research is needed to address issues regarding the amount of cognitive load imposed by aural, visual (pictorial), and verbal (written) information and to assess what roles individual differences play in moderating the effects described in this article (Plass et al., in press). Future research focusing on outcome measures might also study the different recall of vocabulary with aural rather than written cues. In addition, the differences in the larger and more stable effect sizes for visual information and the smaller and less stable effect sizes for written information, which we found both for vocabulary acquisition and text comprehension, have not been found in ear-
lier research on listening comprehension or reading comprehension and therefore warrant further investigation, especially under consideration of the potential moderating effects of individual differences.

ACKNOWLEDGMENTS

All materials for this study were developed by Linda Jones, with help from Louise Rozier as voice talent. The design of the materials was inspired by *CyberBuch* (Chun & Plass, 1997b).

REFERENCES


LaSalle Meets the Quapaws (1682)

On fit traverser les canots sur la gauche à une pointe de sable. On se retrancha le mieux qu’on peut avec des petits bois de tremble, qu’on coupa, dont on fit des palissades. La brume se cessa, et l’on vit un canot de Sauvages venir à nous . . . Mais voyant qu’on ne leur en tirait point, ils s’en retournèrent chercher dire que c’étaient des gens de paix. Ils revinrent 6 sans armes avec le calumet de paix faisant signe qu’on vienne à leurs habitations. Ils présentèrent à fumer à M. de LaSalle et à tous ceux qui étaient autour de lui disant toujours qu’on s’embarque . . . Le lendemain les guerriers et la jeunesse dansèrent le calumet. C’est de s’assembler tous sur la place. Les guerriers mettent leur présents sur des perches comme quand on veut faire sécher du linge. Ils apportent deux grands calumets enjolivés de plumes de toutes couleurs et pleins de cheveux de leurs ennemis. Ils mettent tout cela entre les mains de leurs chefs qui sont assis sur leurs culs et arrangés autour de la place. Ils ont tous des gourdes pleines de cailloux et des tambours qui sont des pots de terre. Ils commencent une chanson qu’ils accompagnent du bruit de leurs instruments. Les guerriers qui ont fait de belles actions vont danser et frapper un poteau avec leurs casse-tête et disent les belles actions qu’ils ont faites . . . cependant les chefs fument, l’un après l’autre dans les calumets, et chacun le présentait à M. de LaSalle et à tous ceux dans la compagnie. Après, ils le prirent et le placèrent au milieu de la place, dansant tous autour de lui au son des instruments et chansons, chacun lui mettant sur le corps son présent qui étaient des peaux de boeufs qui ont de la laine comme nos moutons d’Europe. Si les Français ne l’auraient pas déchargé à mesure de ses peaux, ils l’auraient étouffé sous leurs présents. Il leur fit à son tour présents de haches, couteaux, et rassades.

APPENDIX B
Examples from Recall Protocol Measure of Comprehension

High Score Example: Treatment 4 (23/63).
LaSalle and his men traveled by canoe and arrived at a sandy beach. They built a palisade of wood and met some native Indians. It was foggy. The warriors were peaceful and gave or offered LaSalle a pipe to smoke which was decorated. They gathered in a central place where there were two poles and the natives sat with something on their backs. LaSalle was interested in creating peace. They exchanged gifts. The natives gave LaSalle buffalo skins, wool, and sheep. Then, he gave them hatchets. They sang songs in the camp of the indians. There were many Indian warriors. There were gourds and pebbles which the Indians used.

Low Score Example: Treatment 1 (1/63).
I recall that they went to a port and bought items. Fish, I believe, was one of them. LaSalle had friends, they talked about them at the end. They spoke of a party and they talked about food throughout the passage. They talked about potatoes and a lot of something. They had a feast or party. They spoke of a cassette.