Supporting Listening Comprehension and Vocabulary Acquisition with Multimedia Annotations: The Students’ Voice

LINDA C. JONES
University of Arkansas

ABSTRACT
This study extends Mayer’s (1997, 2001) generative theory of multimedia learning and investigates under what conditions multimedia annotations can support listening comprehension in a second language. This paper highlights students’ views on the effectiveness of multimedia annotations (visual and verbal) in assisting them in their comprehension and acquisition of vocabulary from aural texts. English-speaking college students listened to a 2 min 20 sec historical account in French presented by a computer program. Participants were randomly assigned to one of four listening treatments: the aural text (a) with no annotations, (b) with only verbal annotations, (c) with only visual annotations, and (d) with both visual and verbal annotations. For purposes of this paper, 20 students were purposively selected to participate in interviews. Overall, students remembered word translations and recalled the passage best when they had selected both verbal and visual annotations while listening. Students’ voices reflected these results and revealed that they should have options for viewing material in both a visual mode and a verbal mode in a multimedia listening comprehension environment. This study provides qualitative evidence for a generative theory of multimedia learning that suggests that the availability and the choice of visual and verbal annotations in listening comprehension activities enhances students’ abilities to comprehend the material presented and to acquire vocabulary.

KEYWORDS
Multimedia, Listening Comprehension, Choice, Amount of Invested Mental Effort

INTRODUCTION
Listening comprehension activities provide students with the aural component of the target language to help them better hear the intricate sounds, enunciations, and content and develop their abilities to communicate with others in a target language. Educators try to help students enhance their listening skills by assigning them videotape, audiotape or computer-based activities to complete either at home or in the language lab setting. With these materials, students can practice hearing vocabulary words, sentence structures, and dialogues in the target language.
For years, educators and publishers followed a unimodal approach to listening comprehension and presented aural texts without visual or verbal/textual supportive information. Students were often frustrated by such activities (Jones Vogely, 1998) for any number of reasons including lack of prior knowledge of the topic, the comprehensibility of the speaker, the materials reviewed, the lack of visual information, or even the technological design employed. Certainly, when we utilize technology-based listening comprehension materials, our ultimate goal is to help students develop their language skills. However, if the technological design does not offer helpful comprehension aids (e.g., visual aids), then many students’ preferences or needs are ignored, potentially leading to poor comprehension. Thus, when students struggle with the material or the technology used, we find that “the more they fail, the more helpless they feel, and the less effort they come to invest …” (Salomon, 1983, p. 43).

Technology and language teaching have changed in recent years. Now, second language (L2) multimedia packages developed by researchers (e.g., Larson & Bush, 1992; Otto & Pusack, 1992) and by textbook publishing companies provide students with various listening comprehension activities and the learning aids needed to process them. Researchers have also called for an increase in research on L2 listening comprehension (e.g., Cauldwell, 1996; Field, 1997; Joiner, 1997; Lynch, 1998; Mendelsohn, 1998) and an increase in research on technology to better understand how we can utilize the attributes of multimedia to enhance various aspects of language learning, including listening comprehension (e.g., Brett, 1995, 1997; Hoven, 1999; Joiner, 1997; Jones & Plass, 2002; Lynch, 1998; Meskill, 1996; Purdy, 1996; Pusack & Otto, 1997; Salaberry, 2001).

To answer these calls, this paper specifically investigates students’ thoughts and opinions concerning how verbal and visual annotations in a multimedia environment can assist them in their acquisition of new vocabulary from and comprehension of an aural L2 passage. Many related studies in the past pursued a purely quantitative approach (Chun & Plass, 1996a, 1996b, 1997a; Mayer & Sims, 1994; Plass, Chun, Mayer, & Leutner, 1998). While highly pertinent, these studies left many questions unanswered concerning students’ views and experiences in a multimedia environment. This paper, therefore, more closely examines the students’ voices, qualitative data that can provide immeasurable and even unanticipated information concerning the design and effectiveness of multimedia tools for listening comprehension. The results establish a clearer picture of the current needs and attitudes of L2 students in a multimedia environment and clarify how developers can better design listening comprehension materials and activities to enhance students’ aural comprehension and make them feel as though they are indeed learning.

To bring the qualitative perspective to light, the research strategies used for both the quantitative and qualitative components will be fully disclosed while the statistical information, more completely discussed in Jones & Plass (2002), will be summarized. Students’ voices will be represented by their exact quotes, and
their remarks will be further highlighted by an appropriate literature review. It is through research such as this that we can promote change in our pedagogical and technological strategies and can find ways to facilitate students’ acquisition of L2 aural skills in a multimedia environment.

**METHODOLOGY**

**Quantitative Component**

Participants and Design

The participants in this study were 171 English-speaking students enrolled in second semester beginning French at the University of Arkansas. At the time of testing, their mean estimated French GPA was 2.92 (4.0 = A, 3.0 = B, 2.0 = C, 1.0 = D, 0.0 = F).

A pretest/posttest control group (between-subject) design was employed to observe the effects of two factors—the absence or presence of verbal annotations and the absence or presence of visual annotations—on students’ comprehension of the aural passage and their acquisition of vocabulary. Participants were randomly assigned to one of four treatments: the aural text (a) without access to annotations, (b) with only verbal annotations available, (c) with only visual annotations available, and (d) with both visual and verbal annotations available.

Dependent Measures and Scoring

The effects of the four aural treatments on students’ comprehension and vocabulary acquisition were measured using an immediate multiple-choice vocabulary posttest and an immediate recall protocol posttest (dependent measures) then again, as delayed tests, 3 weeks later. The multiple choice vocabulary posttest was made up of 25 of the 27 keywords visible on the five computer screens of each treatment. The keywords were selected from the aural passage because of their importance in the text and because they could be represented in both a visual and a text-based format. Using the split-half reliability method, the internal consistency of the vocabulary test was .82.

The recall protocol comprehension test instructed students to summarize, in English, the aural passage they listened to. Two French professors chose the 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. The interrater reliability of this measure was .97.

Materials

The listening comprehension software was developed using Adobe Premiere 4.2 and Authorware 4.0. The apparatus for presenting the materials consisted of a 22-station Macintosh computer language lab, arranged such that students could view only their own computer screen.
The computer-based materials presented a 2 min 20 sec aural reading of an authentic encounter between LaSalle and the Quapaw Indians in 1682 (Buzhardt & Hawthorne, 1993; see Appendix). This historic text was chosen because of its rich visual depiction of the encounter and its unavailability in English. It was digitally recorded using the voice of a female native French speaker.

Each treatment began with an introductory screen that placed the historical event in context (advance organizer) and instructed students on how to use the program. The five separate screens that followed contained a total of 27 keywords, positioned on the left side of the screen, accompanied by ellipses to indicate missing words and thus to emulate the flow of the dialogue. Audio play buttons were positioned to the left of each text segment such that students could have equally available access to the predived syntactic chunks of the passage (Meskill, 1996; O’Malley, Chamot, & Küpper, 1989). An icon of a speaker was present on each screen so that students could drag a keyword to it to hear the word pronounced.

The four treatments differed as follows:

1. In the Control group (Trt. 1), no annotations for the key vocabulary words were available except for the option of hearing them pronounced.
2. In the Visual group (visual annotations only, Trt. 2), a camera icon was present to the left of the pronunciation icon. Students could drag a keyword to the camera icon to view its image.
3. In the Verbal group (verbal annotations only, Trt. 3), a text icon was present to the right of the pronunciation icon. Students could drag a keyword to the text icon to view its text-based translation.
4. The Visual and Verbal group (visual and verbal annotations, Trt. 4) contained both the camera and the text icons, in addition to the pronunciation icon. Students could drag a keyword to the camera icon to view its visual annotation or to the text icon to view its verbal annotation of the keyword (see Figure 1).

The visual annotations consisted of 14 color drawings and 13 color photos. The text annotations were written in bold, 14 point Helvetica font. Students could select any annotation available in their treatment as often as desired, before, during, or after presentation of each aural segment.
Students first participated in this segment of the study during two consecutive 50-minute class periods (25 minutes in the first class period, 50 minutes in the second). On the first day, students filled out a paper-and-pencil demographic questionnaire and were then given 8 minutes to complete a vocabulary pretest. On the second day, students were randomly assigned to one of the four treatment groups. They were given 14 minutes to listen to the passage and were instructed to look up all annotations available in their treatment. They were then given 8 minutes to summarize all that they could from the passage and then 8 minutes to complete a vocabulary posttest, identical to the pretreatment vocabulary test. Three weeks later, students completed unannounced delayed vocabulary and comprehension tests that were administered in the same way as the immediate measures.

**Qualitative Methods**

Participants

Twenty students were purposively selected to participate in interviews so as to gain a clearer understanding of the effects of the different media types on students’ comprehension and vocabulary acquisition (see Table 1). Students were selected based on the treatment group in which they worked and the extreme results in
their comprehension and vocabulary acquisition tests. All interviewees were asked to report on the strategies they used and the effects of these strategies on their comprehension and vocabulary acquisition.

Table 1
Number of Students Interviewed Per Treatment

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Number of students interviewed per treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) None</td>
<td>4</td>
</tr>
<tr>
<td>(2) Visual</td>
<td>5</td>
</tr>
<tr>
<td>(3) Verbal</td>
<td>5</td>
</tr>
<tr>
<td>(4) Verbal and Visual</td>
<td>6</td>
</tr>
</tbody>
</table>

Intervisio
The interviews lasted from 20 to 45 minutes each and took place after all dependent measures were completed. During each interview, the researcher used prepared questions but did not strictly adhere to them since, more times than not, the interviewees provided responses to the questions or discussed unanticipated topics without prompting. Once all interviews were completed, they were transcribed and coded for closer analysis.

Throughout each interview, the researcher remained unbiased and accepted each student’s remarks openly. Member checks ensured that what the students said indeed reflected their true beliefs. Unfortunately, all interviews took place 4 to 5 weeks after students’ experiences with the treatments, an unavoidable situation of “give and take” so as to eliminate any influence of the interviews on the delayed comprehension and acquisition tests. Therefore, the “lateness” of the interviews potentially lessened students’ expression of their experiences with their respective treatments.

Data Analysis

All interviews were closely analyzed to identify any emergent data patterns. Since the data were processed electronically, the researcher looked for consistent global themes; in particular, information that underscored students’ experiences with the material and their attitudes toward the different annotations available. Once prepared, the transcripts were more closely reviewed numerous times to identify further unanticipated patterns and individual remarks relating to the quantitative results. The identified information was numerically coded and organized based on themes which were more deeply analyzed to reveal any subtleties that could further support the statistical outcomes. Primary themes were retained for discussion based on their relevancy, either in a supportive or a contradictory manner. Qualitative analysis thus provided a more triangulated approach, and exact quotes are presented in the results and discussion section below.
RESULTS AND DISCUSSION

The underlying theme of this study focused on aural information processing in French as influenced by access to verbal (word translations) and/or visual (images) presentation modes. It was believed that students would comprehend an aural passage and acquire vocabulary best when they had access to and actively selected visual and verbal annotations accompanying the aural material, while students who had access to only one mode (visual or verbal) would perform at a moderate level, and those without access to any annotated information would perform the poorest. Similar research on the comprehension of written texts in a multimedia environment suggests that when students access multiple annotation types (visual and verbal) that accompany a reading comprehension passage, learning is more likely (Chun & Plass, 1996a, 1996b, 1997a; Plass et al., 1998). However, these studies were conducted without inclusion of students’ opinions toward the material with which they worked. The focal point of the study presented here is: What can students tell us to better clarify how visual and verbal annotations might help them to comprehend information aurally presented in a multimedia environment? To understand students’ views more fully, each hypothesis will be presented followed by a summary of the quantitative results obtained. The remainder of the discussion will then focus on students’ views of the helpfulness of visual and verbal annotations for supporting listening comprehension in a multimedia environment.

Hypothesis 1

The first hypothesis argues that students who complete a listening comprehension activity that contains a choice of visual and verbal annotations will recall more propositions of a listening text than those who complete listening tasks with single annotations (visual or verbal) or no annotations. The quantitative results of the study support this hypothesis.

To summarize, a multivariate analysis of variance (MANOVA), computed with the number of correct answers on the immediate recall protocol posttest as the dependent measure and the presence or absence of visual and verbal annotations as the between-subjects factors, revealed a main effect for visual annotations, $F(1,167) = 70.02, MSE = 1137.20, p < .001, \eta^2 = .274$, and a main effect for verbal annotations, $F(1,167) = 16.60, MSE = 269.52, p < .001, \eta^2 = .065$. Post hoc comparisons (Tukey HSD) revealed statistically significant differences, showing that students’ performance was lowest when no annotations were available (Trt. 1, $M = 3.2, SD = 2.9$), but was highest when visual information (Trt. 2, $M = 9.2, SD = 4.0$) and when visual and verbal information were available (Trt. 4, $M = 10.9, SD = 4.6$). Students with visual and verbal annotations performed significantly better than those who received verbal annotations alone (Trt. 3, $M = 6.52, SD = 4.3; p < .05$) and better than those who received no annotations (Trt. 1, $p < .01$). The difference in performance of students in the Visual and Verbal group and students in the Visual group approached significance ($p = .052$).
Multivariate analysis of variance with the number of correct answers on the delayed recall protocol posttest as the dependent measure and the presence or absence of visual or verbal annotations as the between subjects factors revealed a main effect for visual annotations $F(1,133) = 24.61, \text{MSE} = 243.184, p < 0.001, \eta^2 = 0.160$, but not for verbal annotations. Post hoc comparisons (Tukey HSD) revealed statistically significant differences showing that students’ performance was lowest when no annotations were available (Trt. 1, $M = 2.66, SD = 2.77$) and when verbal annotations were available (Trt. 3, $M = 3.29, SD = 3.23$); the difference between these two groups was not significant. Students’ performance was highest when visual annotations were available (Trt. 2, $M = 5.58, SD = 3.48$) and when visual and verbal annotations were available (Trt. 4, $M = 5.78, SD = 2.89$); the difference between these two groups was not significant. However, a main effect for the availability of visual annotations demonstrates that differences in comprehension were statistically significant ($< .001$) for those who received visual annotations (Visual group and Visual and Verbal group) and those who did not (Control group and Verbal group). Students’ remarks supported these results and thus the effectiveness of multiple annotations and even visual annotations alone for comprehending the aural passage. Relevant themes emerged including the benefits of selection from and interaction with different processing modes and the strength of visual annotations for attaining a deeper level of understanding of the aural material.

Since prior knowledge of the vocabulary was very low ($M = 3.08, SD = 3.40$), understanding the text was quite difficult for those without access to any annotations. Allyn, Jan, Daniel and Nina (Control group, Trt. 1) were particularly cognizant of the absence of helpful annotations and the difficult nature of the material. Allyn (Trt. 1), for example, referred to the treatment as “cruel and unusual” while Jan (Trt. 1) exclaimed that accessing keywords for pronunciation alone did not help his comprehension:

I had a tough time with that and it got kinda repetitive and it was pretty long and I guess like I was expecting some more visual cues … basically, we could just drag and hear the word again which didn’t help a whole lot.

The absence of annotated clues and low prior knowledge of the vocabulary prevented him from effectively processing the aural input. Despite the efforts of Allyn (Trt. 1) to continue “listening to it [the passage] to try to figure out how they were using that [a word],” low prior knowledge of the vocabulary and an absence of annotated information prevented even repetitive listening from clarifying the aural passage: “If I knew the words then being able to repeat it would help a lot more” (Daniel, Trt. 1). The students simply could not construct meaning without annotations, leading Nina to lose focus of the task all together: “It was really irritating … am I supposed to be doing something that I don’t realize? I don’t know … with no extraneous information … I twiddled my thumbs a little” (Nina, Trt. 1). Nina’s frustration demonstrates that the more students fail to comprehend, the more helpless they feel and the less effort they come to invest in a given passage.
The students’ poor results on the immediate recall protocol posttest ($M = 3.21$) and their delayed recall protocol posttest ($M = 2.66$) reflect the difficult nature of the control group (Trt. 1) and led one student to exclaim: “It made me feel like I don’t know any French” (Allyn, Trt. 1). Overall, their comprehension of the passage was significantly lower ($p < 0.001$) than those who had access to visual annotations.

The Control group (Trt. 1) and the Verbal group (Trt. 3) were significantly different from each other on the immediate recall protocol posttest (Tukey HSD = 0.001) but not on the delayed recall protocol posttest (Tukey HSD = 0.821). Despite their statistical similarities, students in the Verbal group (Trt. 3) believed that the translations of the keywords helped them understand the aural passage:

I found it to be a big help. The key words going through gave what the meanings of the words were and that would help me when I went back and listened … it really helped me understand and I could pick up on the words I hadn’t necessarily picked up on before because I knew where the conversation was headed and what it was about … if I wouldn’t have had the meaning of the word there … it would’ve been like making a stab in the dark (Max, Trt. 3).

Keyword translations provided information directly related to the passage and allowed students to more richly utilize metacognitive, cognitive and interactive strategies. Doug (Trt. 3), for example, would select from the keyword translations to construct meaning from the text:

I paid attention to the keywords because that was there and that’s what I knew and that’s what I didn’t know but then once I learned the keywords then I’d try to go back and put it in the sentence and see if I could recognize the whole sentence as well.

Deb (Trt. 3) also followed a particular strategy to process the aural text:

I listened through it trying to figure out as many words as I could. I think I dragged each word before I listened to it. And then, I would listen to it … I tried to maybe put the words that I knew in context. So, and then … I’d go back through and listen through it you know parts of it just to get more meaning.

Because students differ, the freedom of interaction meant that students could follow their preferred processing paths through the aural treatment. Despite this flexibility, the presence of so many new, unfamiliar words hindered Jerry’s (Trt. 3) ability to understand:

I’m not to the point yet where I can pick out what they are saying unless it was words that we’ve been working on in class recently or have worked on … like I could’ve listened to that thing on the computer a hundred times and I wouldn’t have ever known what they were saying.
Other than Jerry’s struggles, the Verbal group (Trt. 3) expressed a more positive view of the activity and believed that keyword translations were sufficient for comprehending the passage. However, the immediate recall protocol ($M = 6.52$) and delayed recall protocol posttest results ($M = 3.29$) did not reflect this more positive outlook.

All recall protocol posttest results for the Visual group (Trt. 2) and the Visual and Verbal group (Trt. 4) differed significantly from those of the Verbal group (Trt. 3, $p < 0.050$) and the Control group (Trt. 1, $p < 0.001$). However, the Visual group (Trt. 2) did not differ significantly on the immediate recall protocol (Tukey HSD = 0.203) or on the delayed recall protocol posttest (Tukey HSD = 0.992) from those who processed both visual and verbal annotations (Trt. 4). Despite their similarities, access to words annotated in both visual and verbal modes lead to the highest recall protocol mean scores. In other words, the “‘good’ mixture of narration and visuals [resulted] in a richer and more durable memory representation” (Baggett, 1989, p. 120). These results support and extend Mayer’s (1997, 2001) generative theory of multimedia learning since the acquisition of new knowledge and comprehension of the aural material was greatest when students processed the text, selected from the relevant verbal and visual information available, organized the verbal and visual mental representations of the annotations into a coherent mental representation, and then integrated this representation into their existing mental model to help them most successively construct meaning from the aural passage. Since a connection exists between “an element in the text and some other element that is crucial to the interpretation of it” (Baggett, 1989, p. 108), students had multiple avenues from which they could recall information from memory.

To understand more clearly why the Visual group (Trt. 2) and the Visual and Verbal group (Trt. 4) differed significantly from the others, we must first explore the effects of selecting from both visual and verbal annotations on students’ aural comprehension. In the Visual and Verbal group (Trt. 4), students could choose between the different annotation types; they developed two different mental representations (visual and verbal) in addition to a mental representation of the aural input and, thus, integrated the organized representations of the different modes into their mental model of the aural passage. As Lowell (Trt. 4) commented,

> It helped to have the pictures there and to have an explanation … just seeing the two together and it was just, like when I would go back and you asked me words later on and I would just be able to think, oh I would go back in the story and just like remember pictures of the canoe or the words or something like that.

Chris (Trt. 4), in fact, followed a particular strategy to process the aural input and thus construct meaning:

> Whenever I went through it first, I just did translation. And then, I went back and did the word with the picture and then I think I went through it a third
time and did all three together. I’d look at the word, listen to the translation, and then I’d look at the picture again.

Overall, students in the Visual and Verbal group (Trt. 4) received the highest mean scores on the immediate recall protocol posttest ($M = 10.89$) and the delayed recall protocol posttest ($M = 5.78$).

“Choice,” an important component of Mayer’s (1997, 2001) generative theory of multimedia learning, argues that students may freely select from the information available to process the text. Since students learn efficiently in different ways (Reinert, 1976), multimedia environments that provide both visual and verbal annotations of keywords may be most effective since students can choose the annotation type that best suits their needs (Plass et al., 1998) and can review this information more than once, thereby further reinforcing their learning (Chun & Plass, 1996a). To demonstrate the power of multimodal information, Garza (1991) found that including L2 text-based subtitles with an L2 video enhanced students’ comprehension. The redundancy of the material in aural, verbal, and visual modes meant that students could focus on and select from those modes that best facilitated their aural comprehension. Chun and Plass (1996b) also looked at the effects of annotations with different forms of media on reading comprehension of an L2 reading passage and concluded that visual information in addition to verbal information helped support micro- and macrolevel processing. The presence of visuals and text lead to the building of referential connections between the two systems, resulting in more retrieval routes to the vocabulary items and additive effects on recall (Chun & Plass, 1996a).

The choice of different annotation modes implies that students actively interacted with the technology, as though engaged in a conversation: “I felt like I had a French speaker at my disposal only I didn’t have to ask them to repeat, I would just click on it and listen as many times as I wanted” (Bob, Trt. 4). Though Doug (Trt. 3) worked only with word translations, he, too, found the computer to be more “social” than the audio tape:

On those [tape exercises], you hear a word you don’t know then that’s it, you don’t know it … you’ve just got to figure it out and try to look it up if you even know how it’s pronounced but with the computer you are able to have a conversation with the computer. You can take the keywords up there and get the definition … extremely helpful.

Meskill (1996) argues that an increase in interaction with the different modes of input available will lead to greater integration of the aural message into a learner’s developmental system and thus greater recall of the material. Faerch and Kaspar (1986) also argue that comprehension is best ensured through interaction and negotiation. Since the Visual group (Trt. 2) and the Verbal group (Trt. 3) had access to single annotations, some interaction was possible. However, the multiple annotation modes available in the Visual and Verbal group (Trt. 4) provided easier access to redundant information from the aural passage, encouraged more
interaction, and further reinforced their learning (Chun & Plass, 1996a).

The strength of images affected students’ comprehension of the aural passage as well: “It helped a lot because a couple of times it was hard to understand but seeing the pictures mainly was the big thing that helped” (Edie, Trt. 2). Rebecca’s (Trt. 2) remarks were particularly striking:

I don’t necessarily think giving the definition out right immediately and seeing that with the word; it’s like you have to figure it out, you have to go through a process, you have to kind of want it, you know you see the word [picture] and you want to know what it means and so by just that I think I feel like it kind of clicks something in your head and you remember you tend to remember it a little bit more than rather if it is the word definition … it makes you question it or even if you’re really not that interested in it but if you want to succeed at what you are doing and you are trying at what you are doing … I really like the pictures a lot … it allows you to visualize it better.

Rebecca’s description reinforces the theory that visual annotations demand but also support deeper processing of a proposition than do verbal annotations; since visuals carry a structural message that compliments the language presented (Baggett, 1989; Kozma, 1991), they carry more properties with the corresponding events of the real world than do text (Kozma, 1991) and, thus, make learning more efficient (Oxford & Crookall, 1990). Therefore, if listening entails the construction of mental representations and interpretations, it makes sense that visuals would more strongly support such a process than words (Meskill, 1996) since “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). Analogic representations (images) are mapped directly 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, 1997, p. 8).

Research has long indicated that processing images with an aural text positively affects students’ aural comprehension (Carlson, 1990; Chung, 1994; Mueller, 1980; Pouwels, 1992; Severin, 1967). Severin (1967) determined that students who processed a listening passage with pictures performed better on posttreatment activities than did those 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 that more proficient students’ performance showed very little difference between listening with or without images. He suggests 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 are otherwise absent. Kozma
(1991, p. 192), therefore, argues that

People can construct a mental representation of the semantic meaning of a story from either audio or visual information alone, but it appears that when presented together each source provides additional, complementary information that retains some of the characteristics of the symbol system of origin . . . Audio may be sufficient for those knowledgeable of a domain, but visual symbol systems supply important situational information for those less knowledgeable.

Research in listening comprehension (e.g., Chung, 1994; Hudson, 1982; Mueller, 1980; Omaggio-Hadley, 2000; Severin, 1967), in tandem with the quantitative and qualitative results in this study, highlight the helpfulness of images for aural processing and help demonstrate why comprehension rates in the Visual group and the Visual and Verbal group were significantly greater than those in the Verbal group and the Control group, \( p < 0.05 \). The “bushier” format was key. To clarify, Baggett (1989) argues that when we hear the verbal representation “brown leaf,” we have two “words” before us: “brown” and “leaf.” When we see a “brown leaf,” we are more cognizant of its size, shape, color, use, and environment. This visual information simply carries more information than does text and allows for greater comprehension and retention. When we examine the recall protocol measures, students who accessed visual annotations understood and retained their knowledge of the passage best because the dense/deep quality of images allowed residual memory to remain. The visual annotations improved students’ memory of the aural input, both short term and long term.

Students’ comments supported the quantitative results for the first hypothesis in that the Control group (Trt. 1) had the lowest opinion of the activity as well as the lowest scores on all recall protocol posttests; the absence of annotations prevented them from understanding the passage. The Verbal group (Trt. 3) was more positive and believed that the translations of the keywords helped them construct meaning. However, these students did not perform as well as those who had visual annotations available to them. The Visual group (Trt. 2) believed that visual annotations allowed for deeper processing of the aural passage and, thus, longer retention of the material. Images did, in fact, lead to deeper processing; the mapping of visual input with the aural input into a mental model resulted in stronger recall than did the mapping of words. The Visual and Verbal group (Trt. 4) further believed that the ability to interact with the computer and to choose from multiple annotations provided more than one route to the information and created a more individualized, “socially” interactive approach. Students could select from and make two and even three connections between the verbal, visual, and aural mental representations to help them construct meaning. Students thus provided a more supportive view of the helpfulness of multiple modes, images, and interaction for aural comprehension.

Students’ construction of knowledge from an aural text requires “learning on a
higher level, including understanding words in context as well as propositions” (Plass et al., 1998, p. 27). Thus, the deeper processing required of and supported by images leads to greater knowledge construction. Initially, this process seems different from vocabulary acquisition where students often rely on rote memorization of word translations to learn the meanings of words. Despite the shallower reputation of vocabulary acquisition, the availability of different annotation types (visual and verbal) and the amount of invested mental effort (Salomon, 1983) that students give to a particular annotation mode suggest that deeper processing may also be key to acquiring vocabulary. Therefore, it is a discussion of the second hypothesis and vocabulary acquisition in an annotated listening comprehension environment to which we now turn.

**Hypothesis 2**

The second hypothesis examines how students best acquire vocabulary when the availability of annotation types varies from treatment to treatment. It was proposed that students who completed a listening comprehension activity that contained visual and verbal annotations would acquire more vocabulary words than those who completed the listening task with single annotations (visual or verbal) or no annotations. To summarize the statistical results, a multivariate analysis of variance, computed with the number of correct answers on the immediate vocabulary posttest as the dependent measure and the presence or absence of visual and verbal annotations as the between-subjects factors, revealed a main effect for verbal annotations, $F(1,167) = 82.38$, $MSE = 1561.27$, $p < .001$, $\eta^2 = .243$; and for visual annotations, $F(1,167) = 70.90$, $MSE = 1343.56$, $p < .001$, $\eta^2 = .209$; and an interaction effect for visual and verbal annotations, $F(1,167) = 18.71$, $MSE = 354.61$, $p < .001$, $\eta^2 = .055$. Post hoc comparisons (Tukey HSD) additionally showed that student performance was lowest when no annotations were available (Trt. 1, $N = 42$, $M = 8.10$, $SD = 4.3$), and highest when both visual and verbal annotations were available (Trt. 4, $N = 44$, $M = 19.75$, $SD = 3.2$). In fact, students in the Visual and Verbal group significantly outperformed all others (for Control group, Trt. 1, $p < .001$; for Verbal group, Trt. 3, $p < .05$; for Visual group, Trt. 2, $p < .05$). When either visual or verbal annotations alone were available (Verbal group, Trt. 3: $N = 44$, $M = 17.02$, $SD = 5.6$; Visual group, Trt. 2: $N = 41$, $M = 16.59$, $SD = 4.0$), students’ performance was higher than when no annotations were available (for Verbal group and for Visual group, $p < .001$), but did not differ significantly from one another.

In addition to these results, a multivariate analysis of variance with the number of correct answers on the delayed vocabulary posttest as the dependent measure and the presence or absence of visual or verbal annotations as the between-subjects factors revealed a main effect for visual annotations $F(1,133) = 41.22$, $MSE = 760.38$, $p < .001$, $\eta^2 = .20$, for verbal annotations $F(1,133) = 21.11$, $MSE = 379.28$, $p < .001$, $\eta^2 = .105$ and an interaction effect for verbal and visual annotations, $F(1,133) = 4.76$, $MSE = 85.58$, $p < .05$, $\eta^2 = .024$. Post hoc comparisons
(Tukey HSD) further revealed that when no annotations were present, students’ performance again was lowest (Trt. 1, $M = 6.33$, $SD = 3.69$). When visual annotations (Trt. 2, $M = 12.45$, $SD = 4.20$) or verbal annotations (Trt. 3, $M = 11.15$, $SD = 4.90$) were available, students’ performance was significantly higher than when no annotations were available (for Verbal group and for Visual group, $p < .001$), but did not differ significantly from one another. When both visual and verbal annotations were present, students’ performance was highest (Trt. 4, $M = 14.08$, $SD = 4.02$), and the difference to students with verbal annotations available was statistically significant ($p < .05$), while the difference to those with visual annotations available was not.

As with Hypothesis 1, students’ remarks supported this second hypothesis and the helpfulness of multiple annotations for acquiring vocabulary. Once again, relevant themes emerged including the helpfulness of interaction with the annotations, the supportive nature of multimodal materials for vocabulary acquisition, as well as students’ beliefs concerning the amount of invested mental effort (Salomon, 1983) needed to process verbal or visual annotations.

In the strictest terms, vocabulary acquisition involves rote learning whereby students learn or memorize definitions of individual words. This basic idea would lead one to assume that direct translations of L2 keywords would provide students with the tools necessary to understand. However, it may be misleading. Kellogg and Howe (1971) suggest that foreign words associated with visual imagery or actual objects are learned more easily than those without such comprehension aids. Plass et al. (1998) contend that retrieving words is more difficult than inputting them and that multiple annotation types, rather than single mode annotations, can facilitate this process since “the organization and integration of two different forms of mental representations enhance retrieval performance by providing multiple retrieval cues” (p. 34).

Students’ impressions varied as to the helpfulness of images and translations for acquiring vocabulary. The Control group (Trt. 1) received the lowest mean scores on all vocabulary measures and again found the absence of annotated information very challenging: “I thought the French vocabulary and listening parts in the study were difficult” (Jan, Trt. 1). On the other hand, access to annotated cues in the other three treatments helped students process the material: “Instead of having to flip through 1,000 pages in a French book to try to figure out what this word means, you just had to drag it up … it really was helpful” (Doug, Trt. 3). Frances (Trt. 4) had a similar view:

You were doing something and you had pictures to relate to what you were doing and … you didn’t have to look in separate, you didn’t have to go ‘let me flip back to this’; it made it a whole lot more convenient and less frustrating.

Garrett (1991) favors the use of annotations that address individual preferences since they allow students to access the explanations and practice opportunities they need. In this light, students expectedly had differing views towards the
helpfulness of visual or verbal annotations for acquiring vocabulary. For Frances (Trt. 4), images combined with the aural passage helped her later recall word meanings:

Well, a lot of the words sounded similar so, for words that were like totally off, I related like the picture. Like there was one word that described beautiful and it seemed like design, so I related words like that. I forgot what the word was but it sounded similar to design and I kind of related design to meaning beautiful, that’s what the word meant, like decorated or something like that … The pictures helped me to get a feel for the passage. Also, I will probably be able to remember the vocabulary from picture association.

She selected from the annotations available, made connections between the mental representations of the sounds, words, and images and later used these connections to recall the vocabulary. Chris (Trt. 4) also relied on the strong relationship between the visual and verbal mental representations to recall their meaning more easily:

It helped a lot for me personally to be able to … have the translation, but to also see the picture whenever I heard the word just to associate … maybe if I couldn’t remember the hearing text I could remember seeing it on the page and I think it helps to have the picture along with hearing it, just you know like a double reinforcement.

However, students differed concerning the match of annotation types with their preferred processing strategies. Visual annotations, for example, suited Pat’s (Trt. 2) strategy of acquiring new vocabulary: “When I’m learning a foreign language, I try to think of the pictures instead of thinking of the English word and translating; I try to think of pictorially or emotionally what it is and match that to the word.” In contrast, despite his preference for text-based cues, Bob (Trt. 4) looked up some images that subsequently helped him retain and recall two vocabulary words:

There are two that I did recall with visual cues; I remember noting that when I did it because I was surprised that the images stuck with me … one was the culs, the backs of the people, and the other one was the poles, the poles sticking out of the ground, the perches.

Annotations do help students acquire vocabulary (Chun & Plass, 1996a; Knight, 1994; Kost et al., 1999; Plass et al., 1998). Chun and Plass (1996a) looked at the effectiveness of different annotation modes (visual and verbal) for vocabulary acquisition in an L2 reading passage. They found that access to multiple annotations led to even greater vocabulary acquisition; on average, students recalled 77% of the vocabulary words when they accessed multiple annotations. We see similar results in the present study; on average, students recognized 79% of the vocabulary from the aural passage when multiple annotations were present, and
they provided favorable comments for using such a strategy.

Despite the helpfulness of different annotation types, students had strong opinions concerning the amount of invested mental effort (Salomon, 1983) needed to process them and thus to learn the vocabulary or to recall the aural passage. Pat (Trt. 2), for example, believed that some images may not have clearly defined the keywords and that translations would have been easier:

Sometimes the English translation would have been nice because see there was a picture and I wasn’t sure what exactly what part of the picture was important … Like the word fog, there was a picture of cliffs and the arrows kinda pointed to the fog but it was awfully hard to point to fog. And I think it took me a couple of times of listening to it before I realized that it was the fog and not say the bluffs that it was pointing to or talking about. So, in some instances, the English translation would be easier than a picture would.

Although she did not have the translation “fog” to influence her understanding of the French word *brume*, through deeper analysis of the image itself, she correctly identified the word. Others remarked that a direct translation was sufficient, “I think that just the textual definition is plenty and anything else is not really very meaningful” (Pam, Trt. 3). Still others commented that it was only out of curiosity that they consulted the images:

I only looked at the visuals because I was curious as to what the computer would do but when I wanted to understand the word, I was much more inclined to go for the translation because that was the most immediately gratifying thing … (Bob, Trt. 4).

For Charlie (Trt. 2), a translation provided instant meaning of a keyword, whereas an image demanded greater effort: “The images, the point of them is still to get across what the word means and a translation is going to do that fine without you having to interpret what the image is trying to tell you.” John (Trt. 4) added that a translation eliminated the need to “figure out” the visual representation: “Seeing the English really helped me out so I didn’t have to figure out anything so it was there, set in stone, this is the way it is, memorize it so you can use it, type thing.” And yet, it may well be the challenge of figuring out meanings of images that lead to greater retention of the vocabulary and greater recall of the passage.

Students’ views of the easy or difficult nature of visual and verbal annotations highlight Salomon’s (1983) theory concerning the amount of invested mental effort students apply to a given task. Specifically, Salomon (1983, p. 42) believes that “the number of nonautomatic elaborations applied to a unit of material …,” or the amount of mental effort students invest in learning is influenced by how they perceive the source of the information. That is, if they perceive that a given task is difficult, they will use more mental effort and therefore more nonautomatic energy to process the material. If they perceive that a given task is easy, they will
invest less mental effort and may potentially learn or retain less as a result.

Cennamo (1993) determined that learners’ perception of television as an easy medium actually interfered with their ability to learn from it. Salomon (1984) also found that when learners perceived television to be easier than print, they invested less mental effort in learning from television and therefore learned less. In this study, many students believed that translations were easier to process than images because they demanded less effort to clarify meaning. However, those who accessed visual annotations outperformed those who did not. Only on the posttreatment vocabulary test was no significant difference present between those who accessed visual annotations and those who did not. Nevertheless, the mean difference from pre- to posttreatment vocabulary tests was greater when students accessed visual annotations than when they accessed verbal annotations: Verbal and Visual group (Trt. 4), mean difference = 16.075; Visual group (Trt. 2), mean difference = 14.34; Verbal group (Trt. 3), mean difference = 13.18; Control group (Trt. 1), mean difference = 5.36. In addition, the Visual group (Trt. 2, \( M = 12.45 \)) and the Visual and Verbal group (Trt. 4, \( M = 14.08 \)) did not differ significantly from each other in the delayed vocabulary test (Tukey HSD = 0.324), though the Visual and Verbal group remained significantly different from the Verbal group (Tukey HSD = 0.012). Students who accessed verbal annotations alone may have expended fewer cognitive resources; they worked in an automatic manner, in which little conscious effort was utilized and experience with the material was fast and effortless, and therefore learned less. In contrast, students who worked with images worked in a nonautomatic manner in which deeper, more effortful mental processing occurred (Salomon, 1983). These students received significantly higher recall protocol scores and better acquired vocabulary. In short, “the rate of forgetting [was] thus a function of the lack of depth and analysis” (Cohen, 1987, p. 45).

Salomon’s (1983) theory helps to explain why the results were significantly greater when students accessed visual annotations. However, it does not completely address the outcomes of this study. An interaction effect was present for both verbal and visual annotations and students’ results on the vocabulary tests, when accessing verbal or visual annotations alone, were not significantly different from each other. There are two possible explanations for this situation. First, students who accessed translations on the computer viewed the same, direct representations of the keywords on each vocabulary test. Their treatment, by default, familiarized them with the English translations and, consequently, gave them an advantage over those who did not have translations (i.e., those who had to “work harder” to interpret meaning). Students without access to verbal annotations did not see direct English translations until they took the vocabulary tests. Despite the presumably “easier” nature of translations, students more successfully acquired and retained vocabulary definitions with images present.

An additional explanation for the similarities between those who accessed only
verbal or only visual annotations is that the multiple-choice vocabulary tests may have influenced the effects of students’ invested mental effort (Cennamo, 1993). Though five English distracters along with the correct response were provided to diminish students’ abilities to guess, the potential to guess correct responses remained. In contrast, the recall protocol tests required students to demonstrate their understanding of the material without helpful clues available. Therefore, as Cennamo (1993, p. 41) suggests,

> When researchers are interested in assessing the effects of the mental effort learners use to elaborate on the content and create meaning from the materials, questions that require constructed responses may be a more sensitive measure than recognition tasks.

The challenge of creating “proof” of understanding may have additionally contributed to the significant differences between the presence or absence of visuals in all recall protocol measures. The potential to guess the correct response on the vocabulary tests, and the advantage of previewing the verbatim translations as a part of the Verbal group (Trt. 3) may have diminished the significant differences, thereby reducing the visibility of students’ invested mental effort in the results of the vocabulary tests. The strength of images cannot be denied, in particular since students in the Visual group (Trt. 2) did not have access to direct translations and yet performed better than those who did.

**SUMMARY**

Throughout this study, students performed best when they had access to visual and verbal annotations, moderately well when they had access to visual or verbal annotations alone, and poorest when no annotations were available to them. In most instances, access to multiple annotations led to significantly higher results than the other annotations. Students’ comments supported these results in that those who worked without annotations had the lowest opinion of the activity as well as the lowest scores on all dependent measures; the absence of annotations prevented them from understanding the passage and learning the vocabulary. The Verbal group (Trt. 3) was more positive and believed that the word translations helped them understand and recall the vocabulary. The Visual group (Trt. 2) and the Visual and Verbal group (Trt. 4) believed that visual annotations allowed for deeper processing of the aural passage and thus longer retention of the material and greater vocabulary knowledge. Images, in fact, did lead to deeper processing of the aural text; the mapping of visual input with the aural input into a mental model resulted in stronger recall than the mapping of words. As a result, these two groups rarely differed significantly from each other though access to both annotations consistently led to the highest results on all dependent measures.

Students in the Visual and Verbal group also confirmed that choosing from multiple annotation types created a more individualized, balanced, and interactive
approach to the activity. The choice of annotations led to more than one retrieval route to the information, a strategy that was particularly helpful when either the image or the translation was not clear or when students’ cognitive abilities were low for a particular annotation type. The presence of visual and text annotations allowed students to select from the material and to build referential connections between two, if not three, systems, thereby resulting in more retrieval routes to the vocabulary items and added effects on recall.

The qualitative data also highlighted students’ views of the amount of mental effort needed to process visual or verbal annotations. Overall, those who worked with any annotation type were positive about their treatment. However, some believed that English translations were more effective and efficient for learning vocabulary than images because images required more effort to process word meanings. Students’ remarks appear to corroborate Salomon’s (1983) theory that the less mental effort placed on a given task, the less successful the outcome. Students who believed that verbal annotations were easier to process may have expended less effort with such annotation types and, therefore, retained less vocabulary on the delayed vocabulary test and only slightly understood the passage. Images, on the other hand, were viewed as the more difficult form of annotation with which to work. Students may have used more mental effort to process these annotations, and therefore performed significantly better on all recall protocol measures than those with verbal annotations alone. This same attitude may have helped them to perform better on the delayed posttreatment vocabulary test than those with verbal annotations alone. Students in the Visual and Verbal group (Trt. 4) also reconfirmed that choosing from multiple annotation types created a more individualized and balanced approach to the activity, a strategy that was particularly helpful when either the image or the translation was not clear. This outcome, once again, reconfirms Mayer’s (1997, 2001) generative theory of multimedia learning and demonstrates that the treatment design itself benefited students’ comprehension and vocabulary acquisition more so than did the amount of invested mental effort. The presence of visual and text annotations allowed students to select from the material and build referential connections between multiple systems, thereby resulting in more retrieval routes to the vocabulary items and greater comprehension of the aural passage.

CONCLUSION

In a pilot study conducted during the Fall of 1998, students remarked that they struggled with listening comprehension activities when there was a lack of choice for information. In particular, students struggled with aural exercises when they had no preference for the pedagogical design nor could receive any relevant contextual assistance to help them better comprehend the passage. The absence of choice compromised their ability to understand the spoken message.

Guillory (1998, p. 104) believes that strategies that promote access to annotations could well motivate students to work more with aural activities:
Technology-based language learning environments which present short and interesting authentic video modules and a lexicon containing the keywords in the captions for use as a resource while listening may well motivate learners to spend time outside of the classroom exploring the language ….

In this study, several students confirmed this view:

I liked what we did on the computer better than any listening comprehension I’ve done before. I found it much more helpful … I’d like to see it if maybe they made it a part of the classroom or say once every two weeks the class went to the listening lab and say did that (Pat, Trt. 2).

Doug (Trt. 3), who was not impressed with the typical tape technology used in language teaching, made a similar remark: “I think we need more of them [listening activities]. In my opinion, we need to get rid of all that [the tape booths] out there and get all those computers in there and do that [the computer based activities].” Chris (Trt. 4) furthered this belief by saying

I really liked using the computer, it really helps to be able to hear and see the text and pictures and words. I will definitely try to come to the lab to study if there are programs like this one.

These remarks reflect the educational implication of this study, that we should provide students with the option to select from and process visual and verbal annotations that accompany an aural passage in a multimedia environment. Students’ remarks revealed that choice in educational materials can better ensure that students’ cognitive abilities and learning preferences are considered. Such a strategy would address students’ needs and would permit them to use their own self-directed style of accessing and processing information. Because the students had a more positive outlook and better results with listening activities when provided with multiple annotations and interaction, giving them a more prominent opportunity to succeed certainly seems in order.

A theoretical implication for combining listening comprehension activities with multimedia is that organizing aural material into working memory seems to be aided by students making connections between the visual and verbal systems. As stated by the students themselves, the presence of visual and verbal annotations helped them to link information with the aural message and thus better retain information in long-term memory for later comprehension and vocabulary recall. Students’ voices, in addition to the quantitative results summarized above, extend Mayer’s (1997, 2001) generative theory of multimedia learning to listening comprehension since selection from the material available and construction of referential connections between visual, verbal, and aural representations gave students more retrieval routes to the information to recall vocabulary words and to comprehend the passage.
NOTE

1 The names of the students who participated in the study have been changed.

REFERENCES


**APPENDIX**

LaSalle Meets the Quapaws (1682) (Buzhardt & Hawthorne, 1993, pp. 1-2)

On fit traverser les canots sur la gauche à une pointe de sable. On se retranchait mieux que l’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 la Salle 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 la Salle 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’avaient 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.

**AUTHOR’S BIODATA**

Linda C. Jones is Assistant Professor of Instructional Technology and Director of the Language Learning Center in the Department of Foreign Languages at the University of Arkansas. She developed all materials for this study with help from Louise Rozier as voice talent. The design of the materials was inspired by *CyberBuch* (Chun & Plass, 1997b).

**AUTHOR’S ADDRESS**

Linda C. Jones  
Department of Foreign Languages, KIMP 425  
University of Arkansas  
Fayetteville, AR 72701  
Phone: 479/575-7608  
Fax: 479/575-6795  
Email: lcjones@uark.edu