EFFECTS OF TIME-COMPRSSED NARRATION AND REPRESENTATIONAL ADJUNCT IMAGES ON CUED-RECALL, CONTENT RECOGNITION, AND LEARNER SATISFACTION

ALBERT DIETER RITZHAUPT  
University of North Carolina, Wilmington  

ANN BARRON  
University of South Florida

ABSTRACT
The purpose of this study was to investigate the effect of time-compressed narration and representational adjunct images on a learner’s ability to recall and recognize information. The experiment was a 4 Audio Speeds (1.0 = normal vs. 1.5 = moderate vs. 2.0 = fast vs. 2.5 = fastest rate) × Adjunct Image (Image Present vs. Image Absent) factorial design. Three-hundred five research participants were recruited from a public, southeastern university in the United States. Results showed statistically significant differences at 2.5 times the normal audio speed, in which performance on cued-recall and content recognition tasks was significantly lower than other audio speeds. Furthermore, representational adjunct images had a significant positive effect on cued-recall, but not content recognition. Participants in the normal audio speed and image present groups were significantly more satisfied than those in other treatments. Recommendations to various stakeholders are provided.

INTRODUCTION
Multimedia can be defined as the presentation of information using both words and pictures (Mayer, 2001). Over the past century, there has been tremendous...
growth in interest and research on multimedia, especially relating to learning. The design and delivery of multimedia learning environments are based on principles and guidelines derived from theory, empirical research, and professional experience (Sabatini, 2001). As technology changes, further empirical research and theory development are necessary to demonstrate its efficiency and effectiveness for learning. Because technology advances at such a rapid pace, the process of conducting sound empirical research and developing theory is ongoing.

Digitally recorded audio is commonly integrated into multimedia learning environments (Moreno & Mayer, 2002). Audio can be broken into three main elements: narration (speech), sound effects, and music (Beccue, Vila, & Whitley, 2001). Narration is the dialog that can be used to deliver an instructional message. Narration, unlike its textual counterpart, is inherently time-dependent. In fact, the use of narration can actually increase the time required by a learner to complete a multimedia program (Barron & Kysilka, 1993; Koroghlanian & Sullivan, 2000).

The goal of an instructional designer in a business or industry setting is to maximize a learner’s comprehension and satisfaction, while minimizing the amount of time the learner will spend on a learning task. The philosophy behind this goal is simple: time is money, and in a business or industry setting, both time and money are limited resources. This goal may not be the same in the context of higher education, however. Faculty members and instructional designers in higher education often try to develop learning materials that will pique their students’ interests and engage them in learning material for longer durations. After all, time on task is a well-documented instructional requirement for effective learning (Stallings, 1980). However, students in higher education often have a conflicting goal in which they may attempt to minimize the amount of time on task with the maximum level of comprehension.

Both of the aforementioned scenarios pose an interesting instructional design and research problem. Previous research shows that conversational speech typically takes place at approximately 150 words per minute (wpm) (Benz, 1971; Nichols & Stevens, 1957), and has demonstrated that normal speech can be increased to 200 to 300 wpm, with minimal loss in comprehension (Barabasz, 1968; Foulke & Sticht, 1967; Goldhaber, 1970). If multimedia materials can potentially increase the amount of time a learner spends on a learning task, then students, business and industry could potentially benefit from the use of time-compression technology to reduce the amount of time on task.

The current body of research on the use of time-compressed speech dates back to the 1950s (Fairbanks, Guttman, & Miron, 1957) and focuses primarily on the comprehension or intelligibility of speech at various speeds, while controlling for other relevant variables. A separate, yet related, line of inquiry exists in the area of multimedia learning, which investigates the effects of combining words and pictures in various forms to influence learning (e.g., spoken words versus written words). Time-compressed speech and multimedia learning research are two separate lines of inquiry, though they are investigating similar phenomena.
Though this gap in the body of research still remains, the explosion of distance learning initiatives has prompted faculty members to engage in the development of audio-enhanced instruction. Faculty members are digitally recording voice-over presentations (e.g., PowerPoint with voice), animated screen captures with narration (e.g., Camtasia), or lectures to distribute to personal computers and other portable media devices (e.g., podcasts) so students can learn on demand (Gill, 2007). Apple established iTunes University, a “service for colleges and universities that provides easy access to their educational content, including lectures and interviews, 24 hours a day, 7 days a week” (iTunes, 2007). Numerous higher education institutions have partnered with iTunes University in an effort to develop a wealth of educational materials, primarily in a digital audio format.

Learners in higher education and other arenas have discovered time-compression technology that is integrated into popular consumer products such as iPods or software such as Windows Media Player. The key digital technology that supports the increased or decreased rate of narration, while preserving pitch, in audio files is called a time compression algorithm (Omoigui, He, Gupta, Grudin, & Sanocki, 1999). A major tenet of time compression is to provide learners with the ability to speed up or slow down content based on their learning preferences. In this manner, they can reduce the amount of time spent listening to multimedia with narration (Spencer & Galbraith, 2002).

**Purpose**

The purpose of this research, therefore, was to investigate the effect of time-compressed narration and adjunct images on learners’ (specifically undergraduate college students) ability to recall and recognize information in a multimedia learning environment. Additionally, this research investigated learners’ satisfaction of time-compressed narration and representational adjunct images used in multimedia learning environments. The overarching goals of this research were to shed light on time-compression technology, help fill the existing gap in the research literature by merging two disjoint bodies of research, and aid students, instructors and instructional designers to better understand time-compression technology when creating or using instructionally sound multimedia.

**PREVIOUS RESEARCH**

In conversational speech, one is simultaneously listening and composing speech. Because one can speak at approximately 150 wpm, while the rate for speed reading is 250 to 300 wpm (Taylor, 1965) and the rate for silent reading is 275 to 300 wpm (Junor, 1992), it is reasonable to hypothesize that there is 125 to 150 wpm of unused processing capacity that might be available for listening during normal speech. This hypothesis has been studied and tested by researchers under a variety of conditions starting as early as the 1950s (Barabasz, 1968;
Fairbanks et al., 1957; Foulke & Sticht, 1967; Goldhaber, 1970; Jester & Travers, 1967; Reid, 1968). Although the results from these studies varied, the findings suggest that speeds somewhere near 275 wpm or more begin to negatively influence the dependent measures (e.g., comprehension, recall, etc.) (Fairbanks et al., 1957; Foulke & Sticht, 1967; Reid, 1968). These studies also underscore control variables that may influence the dependent measures of interest, such as academic level (Goldhaber, 1970), grammatical complexity (Reid, 1968), or repetition (Jester & Travers, 1967). However, these previous research studies did not study the effects of time-compressed speech in the context of multimedia (with both pictures and words) learning environments.

Multimedia learning has been investigated from different perspectives using several different combinations of onscreen text, narration, and picture treatments (Barron & Kysilka, 1993; Kalyuga, Chandler, & Sweller, 1999; Mayer & Anderson, 1991; Moreno & Mayer, 2002; Severin, 1968; Tindall-Ford, Chandler, & Sweller, 1997). Across these studies, the use of verbal redundancy appears to be ineffective when incorporating pictures in the treatment interventions as it results in a split-attention effect. The use of audio-visual interventions (either as still pictures or animations in concert with related narration) appear to be an effective combination. This combination is the premise of the multimedia principle (Mayer, 2001), which has been empirically tested in many studies and posits that better learning occurs with the presentation of pictures and words than from words alone.

**THEORETICAL FRAMEWORK**

Much of the time-compressed speech research predates the growth in multimedia learning research literature. From a theoretical perspective, speech or narration is effectively the same treatment as words communicated through an auditory channel. The tenets of multimedia learning provide a coherent framework and perspective with which to systematically investigate time-compressed speech. Research conducted in this manner can integrate knowledge and serve a multi-disciplinary audience. Mayer’s (2001) model of the cognitive theory of multimedia learning is provided to illustrate previous and current research.

Figure 1 illustrates the previous research on time-compressed narration using Mayer’s model. The solid bold lines indicate that previous research in time-compression had narrowly focused on information entering the auditory/verbal channel, and that the presentation and movement of the information through the auditory/verbal channel is analogous to the learner experiencing cognitive overload as the auditory/verbal channel is stretched to its limits. The perforated line surrounding the visual/pictorial channel illustrates the absence of the simultaneous representation of related visual information.

The presentation of an adjunct picture may be able to represent verbal information, and by doing so, provide the additional nonverbal memory representation
Figure 1. Modified cognitive model for multimedia learning representing previous research in time-compression.
that can be retrieved from memory if an individual’s verbal information is inaccessible (Kullhavey, Lee, & Caterino, 1985). This can be explained by a referential process between the verbal and nonverbal information (Pavio, 1986). Of particular importance is the relationship between the representational adjunct picture and words used in multimedia materials. For instance, a speech about the history of the Chinese government with the simultaneous presentation of a German flag is semantically incongruent and, according to theories of multimedia learning, may interfere with the learning process. Feature-related information should be more easily accessible in memory than nonfeature-related or completely unrelated information (Kullhavey et al., 1985).

Figure 2 visualizes the theoretical basis for this study on time-compressed speech using Mayer’s model. Time-compressed speech still enters and moves through the auditory/verbal channel at an unordinary rate. However, the simultaneous activation of the visual/pictorial channel—represented as the bolded lines in the bottom half of the model—provides the learner another related channel to access relevant information. Because the learner has access to both a verbal and pictorial model to build referential relationships, a stronger learning outcome is predicted to result. The solid lines to the right of the model represent the connection with prior knowledge and the referential link between the verbal and pictorial models.

It was therefore hypothesized that under extreme cases (audio speeds that exceed 350 wpm) the presentation of a representational adjunct picture might greatly improve either the recall or recognition of relevant information and ameliorate the negative effects of increased narration speeds. This provides a rationale to examine the recall and recognition of narrative information when presented with adjunct images at a normal (1.0) speed of approximately 150 wpm, a moderate (2.5) speed of 225 wpm, a fast (2.0) speed of 300 wpm, and a very fast (2.5) speed of approximately 375 wpm.

**METHOD**

**Design and Participants**

The experiment uses a 4 Audio Speeds (1.0 = normal vs. 1.5 = moderate vs. 2.0 = fast vs. 2.5 = fastest rate) × Adjunct Image (Image Present vs. Image Absent) factorial design. Audio Speed and Adjunct Image both served as between subject conditions. This research design results in eight unique groups or conditions: NP (Normal-Image Present); NA (Normal-Image Absent); MP (Moderate-Image Present); MA (Moderate-Image Absent); FP (Fast-Image Present); FA (Fast-Image Absent); VP (Very Fast-Image Present); and VA (Very Fast-Image Absent). Table 1 displays the participant distribution by Audio Speed and Adjunct Image conditions. All participants completed the study, and thus, there were no missing data.
Figure 2. Modified cognitive model for multimedia learning representing current research in time-compression.
A total of 305 research participants were recruited from courses at a comprehensive southeastern university in the United States, after making prior arrangements with instructors. The students were offered extra credit for their participation in the study. Fifty-five percent of the participants were male, and 92% indicated that English was their primary language. Forty-nine percent of the participants were classified as juniors, 4% were freshman, 19% were sophomore, 26% were seniors, and the remaining indicated other. Participants represented many different colleges, with 41% from health, 26% from engineering, 15% from education, 5% from business, and the remaining 10% from arts and sciences. The average age of the participants was 23.64 (SD = 6.61) with the maximum age of 53 years old.

### Materials and Measures

**Text and Adjunct Images**

A descriptive narrative titled *Discovering Australia* was used for this research study. This narrative was selected because it had been successfully used in prior educational research studies pertaining to multimedia learning environments (Kealy, Alkhabbaz, Subramanian, Bunch, & Spears, 2006), its content was well-suited for the current research intervention, and because the population (undergraduate students in the United States) had limited knowledge about specific destinations in Australia. The text was slightly modified to suit the needs of this study. The text has a Flesch reading ease of 36.4 and a Flesch-Kincaid 12th-grade level reading score (Flesh, 1949). The narrative consisted of 11 passages of approximately 150-words per passage: one introductory passage and 10 passages describing different locations in Australia. Each passage was subdivided into two paragraphs. The first paragraph related to feature-related information, while the second paragraph was nonfeature-related.
Ten semantically related images corresponding to the Discovering Australia text were selected for research intervention. The images were selected based on their appropriateness in representing one paragraph of the verbal information in each of the 10 passages. Levin (1981) suggested that images can serve as decorative, representational, organizational, and transformational. While decorative images serve no purpose and can actually hinder the learning process, representational images mirror part or all of some related text, and have been found to have moderate effects on learning (Carney & Levin, 2002; Levin, 1981). The images used in this study were purposefully representational in nature as they are intended to provide context relating to the passages in the Discovering Australia text.

Expert reviews were sought to evaluate the images in relation to the selected text. Eight experts in multimedia reviewed the text and the images and confirmed that the images were representational in nature and related to some of the textual information from the passage.

The images were incorporated into a multimedia intervention along with each passage for the Image Present groups. Onscreen text, aside from the Discovering Australia title, was not incorporated into the imagery to avoid a split-attention effect. The narrative was digitally recorded by an English speaking male, and subsequently incorporated into the multimedia intervention. For instance, the passage shown in Figure 3 was taken from the Discovering Australia text. The first paragraph contains the information represented in the image (see Figure 4). The second paragraph does not have a related image.

The corresponding picture (Figure 4) illustrates the city of Sydney and the Sydney Harbour Bridge by visualizing the information from the passage and providing a context. The relationship between the verbal description and the image of the Sydney is intended to activate referential processing according to Dual Coding Theory (Pavio, 1986). The second paragraph is not related to the image shown in Figure 4. The other passages and images from the Discovering Australia text maintain a similar design and semantic connection, in which one paragraph relates to a picture, and the second paragraph does not.

The narration was recorded with each passage taking approximately 1 minute (approximately 150 words) and subsequently altered using Audacity, an open source audio recording and editing utility and Windows Media Player 10. Four different audio files were generated: Normal (1.0); Moderate (1.5); Fast (2.0); and Very Fast (2.5). The four audio tracks were integrated into eight different treatment groups, serving as the between subject conditions.

**Criterion Measures**

Twenty constructed-response questions were created from the Discovering Australia text. An example of the feature-related cued-recall item pertaining to Figure 3 and Figure 4 is “Describe some of the characteristics of the Sydney
At roughly the southeast corner of Australia lays Sydney, a city built around water that offers many recreational activities involving the sun, sand, and surf. The city's location also supports the bustling shipping industry of Port Jackson, which is crossed by Sydney Harbour Bridge, the second longest steel-arch bridge in the world. From the south shore of the port juts the downtown area and Circular Quay, the focus for ocean liners, commuter ferries, and the financial district.

Australia was first sighted by the Dutch almost four centuries ago and they were followed by the English explorer Captain James Cook who sighted the country in 1770. It wasn't until eighteen years later that the first colony was established by Captain Arthur Phillip as a place for the many convicts who crowded the debtor prisons of England. Successive waves of convicts contributed to the swelling population of the state until 1868 when Britain finally discontinued penal settlements.
Harbour Bridge.” This item prompts for the recall of feature-related information and can be activated referentially from the verbal or visual information. An example item pertaining to the nonfeature-related content would be “Who was the explorer that sighted Australia in 1770?” This item pertains to verbal information in the passage and does not directly describe the information in the image. No recall items were developed for the introductory passage.

A rubric was developed to assess each individual constructed response item. The rubrics underwent two iterations using pilot study data. One point was awarded for a response that captured the gist of the correct answer, two points were awarded to a response that was more elaborate in nature or precise in nature, and no points were awarded for incorrect or no responses. Correct answers had to contain verbal information from the passage—not simply a description of a picture. Two members of the research team, having no knowledge of the groups, independently scored a small sample (n = 20) of the protocols from Recall-Australia using the rubric, including all 20 items in each protocol. Inter-rater reliability was calculated at 83.5%. Next, the raters resolved scoring differences in conference until inter-rater agreement exceeded 95%, and updated the rubric to reflect the necessary changes. Inter-rater reliability was calculated on two more occasions at 100% and 95%, respectively. Internal consistency reliability was calculated using Cronbach’s alpha at $\alpha = .79$ for these data. The item-to-total correlations for the scale ranged from $r = .16$ to $r = .59$.

Twenty multiple-choice items were created based on the narrative, serving as content recognition. The multiple-choice questions were developed in a consistent format following established guidelines (Gronlund, 1998). Each stem posed one question for learners to consider, and the distracters were written as likely true/false statements with only one correct statement. The responses were scored dichotomously, and internal consistency reliability was calculated for the scale using Kuder-Richardson 20 at K-R 20 = .631 for these data. The item-to-total correlations for the scale ranged from $r = .03$ to $r = .4$. The instruments were reviewed by instructional technology faculty and instructional technology doctoral students for clarity, accuracy, and content validity, and were subjected to two sets of revisions from pilot studies.

The satisfaction instrument contained 14 items, which were adapted from a previous study (Ritzhaupt, Gomes, & Barron, 2008). The instrument was split into two parts. The first part uses a 5-point semantic differential scale with two bipolar adjectives on both sides. For instance, on the left-most side was the word “negative” and on the right-most side was the word “positive.” This scale was slightly modified using established recommendations and common word-pairs (Osgood, Suci, & Tannenbaum, 1957). The second part of the instrument used a modified Likert ranging from Strongly Disagree to Strongly Agree. The items were designed to measure a learner’s satisfaction with the intervention. For instance, one item stated “The narrator spoke clearly in the Discovering Australia tutorial.” The instrument had an internal consistency
reliability at $\alpha = .94$ for these data. The item-to-total correlations for the scale ranged from $r = .61$ to $r = .8$.

**Computer Programs**

Using the digitally recorded story, pictures, and instruments, a computer program was created for each of the eight treatment groups using Authorware 4.0. The computer program included brief instructions and was installed on an equal number of personal computers with headsets attached in a computer lab, with ample space between computers to prevent participants’ casual viewing of alternate treatments. The computer programs were designed for an 800 × 600 screen resolution and, thus, computer monitors were set to this screen resolution. The estimated length and wpm by Audio Speed are shown in Table 2. Usability testing (using the think-aloud protocol) was conducted to assure that the instructions and the tasks to be performed while in the environment were clear (Fonteyn, Kuipers, & Grobe, 1993). The interventions were deemed suitable as the participants indicated they understood the instructions and tasks to be performed.

**Procedures**

After making prior arrangements with course instructors, the researcher visited classes to inform students that they could receive extra-credit toward the course grade by participating in the research study. The purpose and tasks involved in the research were briefly outlined as a sign-up sheet was distributed, passed around, and signed by interested participants and collected by the researcher. Upon the arrival at a research session, participants were randomly assigned to a workstation with the computer program already configured and a headset with equal volume settings. During the research sessions, the researchers again briefly outlined the tasks to be performed and informed participants that the research was a voluntary, anonymous process. Participants who chose to stay initiated the program. After the introduction, participants started the program and read a specific explanation of the tasks to be completed. The following screen included a button that could be

<table>
<thead>
<tr>
<th>Audio speed</th>
<th>Presentation length</th>
<th>Estimated wpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (1.0)</td>
<td>11 minutes</td>
<td>150</td>
</tr>
<tr>
<td>Moderate (1.5)</td>
<td>7.3 minutes</td>
<td>225</td>
</tr>
<tr>
<td>Fast (2.0)</td>
<td>5.5 minutes</td>
<td>300</td>
</tr>
<tr>
<td>Very Fast (2.5)</td>
<td>4.4 minutes</td>
<td>375</td>
</tr>
</tbody>
</table>
clicked for a sound test. At that time, participants adjusted the volume of their program prior to starting the intervention. Next, basic demographic information, including classification, gender, major, age, and whether English is their second language was collected for descriptive purposes.

All participants were presented with a map, shown in Figure 5 on the left, containing feature information about each of the locations discussed in the Discovering Australia text and the introductory passage at the assigned speed. Participants in the NP, MP, FP, and VP groups were presented with a picture and one passage of narration semantically related to the graphical image as well as nonfeature-related content. Participants in the NA, MA, FA, and VA groups were only presented with the narration and a neutral shape of Australia, shown in Figure 5 on the right, without any feature information, to prevent distractions.

Prior to the start of the narration in all groups, a light beep-like sound signaled the participant that the narration would start. Once the narration was complete, a button appeared, vertically-centered at the bottom of all interventions, to move to the next screen. Participants were instructed to use this time to reflect on the information from the passage. To prevent an ordering effect, passages were randomly assigned by the program until all passages had been traversed by the participant.

After completing the 11 passages of narration, participants completed three 3-column addition problems designed to clear short-term memory in an effort to test the more durable effects of the intervention. Participants were then instructed that during the recall task they would have 45 seconds to provide a response. Next, the participant provided answers to the cued-recall items. If the participant exceeded the time limit, any information that had been typed was secured as the response. The screen indicated how much time they had left with an animated countdown clock.

Figure 5. Feature information map and non-feature information map.
Next, participants were provided instructions about the recognition task and this was followed by the multiple-choice items. Presenting the cued-recall items before the content recognition items was selected to prevent a testing effect. The items were randomly assigned until all items were traversed. Finally, the participants completed the satisfaction items.

Participants assigned to the faster audio speeds, Fast and Very Fast, were assigned an additional on-screen reading activity of approximately 500 words in length. This activity served as a buffer while participants in the normal and moderate audio groups finished their treatment. This approach diminished the influence of participants being distracted by other participants leaving the room. Finally, participants were thanked for participating in the study. Figure 6 illustrates the instructions, intervention and data collection sequence.

Data Analysis

The content recognition, cued-recall, and learner satisfaction scores were scaled from zero to one. Prior to conducting any inferential statistics, basic descriptive statistics were investigated. A Levene’s test was used to test for the assumption of homogeneity of the variance, and the skewness and kurtosis were used to evaluate the normality assumption. The data are assumed to be independent because of the methodical assignment procedures. Finally, a series of ANOVA procedures using Type III sum-of-squares were conducted with a Tukey follow-up procedure if significant differences were identified. Both descriptive and inferential statistics are provided in tables summarizing the information.

RESULTS

To assure the assignment procedure did not result in inequitable assignments to a demographic group, the Chi-Square test ($\alpha = .05$) was executed across treatment assignments, gender, classification, and college. These results show that the gender, college classification, and college of a participant in the sample did not differ significantly from the hypothesized values $\chi^2 = 8.673 \ (p = .277)$, $\chi^2 = 26.151 \ (p = .565)$, and $\chi^2 = 47.452 \ (p = .078)$, respectively. This is an indication that the assignment procedure was not biased on these particular demographics.
Cued-Recall

Table 3 presents the scaled mean, standard deviation and 95\% confidence intervals for cued-recall performance by Audio Speed and Adjunct Image. A Levene’s test was conducted to test for homogeneity of the variance. For the cued-recall scale, it was calculated at $F(7, 297) = 3.451, p < .01$, which indicates that the variance of the recall measure was not equal across groups. Though not a satisfactory finding, ANOVA is robust to violations of homogeneity of the variance (Stevens, 1990). The data were entered into an Audio Speed × Adjunct Image factorial with both Audio Speed and Adjunct Image serving as between subject conditions and cued-recall serving as the dependent measure.

The interaction effect for Audio Speed and Adjunct Image was not statistically significant at $F(3, 297) = 0.13, p = .95$, partial $\eta^2 < .01$. Only 0.1\% of the variability can be attributed to this interaction effect as shown the minute partial $\eta^2$. There was a significant difference on cued-recall based on the Audio Speed as there is a statistically significant main effect at $F(3, 297) = 12.96, p < .01$, partial $\eta^2 = 0.12$. Though the main effect is statistically significant, the partial $\eta^2$ of .12 indicates only 12\% of the variability can be explained.

The main effect for Adjunct Image is at $F(1, 297) = 5.59, p = .02$, partial $\eta^2 = 0.02$, indicating that the presentation of a picture resulted in a statistically significant difference. The partial $\eta^2$ of .02 shows that only 2\% of the variability can be explained by the presence of a representational adjunct image. A Tukey HSD follow-up procedure revealed the participants in the Very Fast audio condition performed significantly less than the Normal ($Mean\ Difference = –0.13, p < .01$), Moderate ($Mean\ Difference = –0.09, p < .01$), and Fast ($Mean\ Difference = –0.10, p < .01$) audio conditions.

<table>
<thead>
<tr>
<th>Adjunct image</th>
<th>Absent</th>
<th></th>
<th>Present</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio speed</td>
<td>Mean</td>
<td>SD</td>
<td>C.I. LB</td>
<td>C.I. UB</td>
<td>Mean</td>
</tr>
<tr>
<td>Very Fast</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.20</td>
<td>0.18</td>
</tr>
<tr>
<td>Fast</td>
<td>0.24</td>
<td>0.16</td>
<td>0.20</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.23</td>
<td>0.15</td>
<td>0.19</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Normal</td>
<td>0.28</td>
<td>0.16</td>
<td>0.24</td>
<td>0.32</td>
<td>0.32</td>
</tr>
</tbody>
</table>

C.I. LB = Confidence Interval Lower Bound (95\%); C.I. UB = Confidence Interval Upper Bound (95\%).
Content Recognition

Table 4 illustrates mean, standard deviation, and scaled and 95% confidence intervals for content recognition by Audio Speed and Adjunct Image. The Levene’s test was calculated at $F(7, 297) = 1.78, p = .09$, which indicates that the variances of the content recognition measure was equal across groups. The data were examined in an Audio Speed × Adjunct Image factorial with both Audio Speed and Adjunct Image serving as between subject conditions and recognition serving as the dependent measure.

The results show there was not a statistically significant interaction on content recognition ($F(3, 297) = 0.37, p = 0.77$, partial $\eta^2 < .01$). Less than 1% of the variability can be attributed to this interaction effect as shown by the trivial partial $\eta^2$. The results show there is a significant difference on content recognition based on the Audio Speed as there is a statistically significant main effect at $F(3, 297) = 9.74, p < .01$, partial $\eta^2 = 0.09$. The main effect, analogous to the cued-recall measure, has a relatively small partial $\eta^2$ and, in this case, shows that the audio speed explains 9% of the variability.

There was not a significant difference on content recognition based on the presence or absence of a representational adjunct image ($F(1, 297) = 3.26, p = .07$, partial $\eta^2 = 0.01$). As in the cued-recall scenario, the main effect for recognition has a minor partial $\eta^2$ for the Adjunct Image condition, and thus only explains 1% of the variability. The Tukey HSD follow-up procedure shows the participants in the Very Fast audio condition performed significantly less than the Normal (Mean Difference $= –0.12, p < .01$), Moderate (Mean Difference $= –0.08, p < .01$), and Fast (Mean Difference $= –0.10, p < .01$) audio conditions.

<table>
<thead>
<tr>
<th>Adjunct image</th>
<th>Absent</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio speed</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Very Fast</td>
<td>0.42</td>
<td>0.13</td>
</tr>
<tr>
<td>Fast</td>
<td>0.50</td>
<td>0.17</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.48</td>
<td>0.16</td>
</tr>
<tr>
<td>Normal</td>
<td>0.53</td>
<td>0.15</td>
</tr>
</tbody>
</table>

C.I. LB = Confidence Interval Lower Bound (95%); C.I. UB = Confidence Interval Upper Bound (95%).
Learner Satisfaction

Prior to analysis, the results were scaled from 0 to 1 to improve the interpretability of the results and serve as an overall indicator of learner satisfaction. The calculation was the summation of the participant item responses divided by the summation for the highest possible response (all on a 5-point scale). Figure 5 illustrates the scaled mean, standard deviation, and 95% confidence intervals of learner satisfaction measure by the independent variables.

Following the same procedures, Levene’s was calculated at $F(7, 297) = 2.07$, $p = .05$, which indicated that the variance of the satisfaction measure was equal across groups. Again, the data were entered into an Audio Speed × Adjunct Image factorial. There was not a statistically significant interaction between the four Audio Speed and two Adjunct Image conditions on learner satisfaction ($F(3, 297) = 0.92$, $p = .43$, partial $\eta^2 = .01$). Again, at 1%, the interaction effect explains very little of the variability.

There was, however, a significant difference on learner satisfaction based on Audio Speed at $F(3, 297) = 54.73$, $p < .01$, partial $\eta^2 = .36$. The main effect, unlike the cued-recall or content recognition measures, has a larger partial $\eta^2$, indicating time-compression speed explained 36% of the variability in learner satisfaction. There was also a significant difference on learner satisfaction based on the presence or absence of a representational adjunct image. The main effect for Adjunct Image is at $F(1, 297) = 5.26$, $p = .02$, partial $\eta^2 = .02$. This main effect only explains 2% of the variability in learner satisfaction. The Tukey HSD follow-up procedure shows the participants in the increased Audio conditions were all significantly less satisfied than any time-compressed speed other than normal. This finding is illustrated in Table 6.

<table>
<thead>
<tr>
<th>Audio speed</th>
<th>Mean</th>
<th>SD</th>
<th>C.I. LB</th>
<th>C.I. UB</th>
<th>Mean</th>
<th>SD</th>
<th>C.I. LB</th>
<th>C.I. UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Fast</td>
<td>0.41</td>
<td>0.15</td>
<td>0.37</td>
<td>0.46</td>
<td>0.41</td>
<td>0.14</td>
<td>0.37</td>
<td>0.46</td>
</tr>
<tr>
<td>Fast</td>
<td>0.48</td>
<td>0.14</td>
<td>0.43</td>
<td>0.52</td>
<td>0.55</td>
<td>0.16</td>
<td>0.51</td>
<td>0.60</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.56</td>
<td>0.16</td>
<td>0.52</td>
<td>0.61</td>
<td>0.60</td>
<td>0.15</td>
<td>0.55</td>
<td>0.64</td>
</tr>
<tr>
<td>Normal</td>
<td>0.68</td>
<td>0.14</td>
<td>0.63</td>
<td>0.73</td>
<td>0.72</td>
<td>0.09</td>
<td>0.68</td>
<td>0.77</td>
</tr>
</tbody>
</table>

C.I. LB = Confidence Interval Lower Bound (95%); C.I. UB = Confidence Interval Upper Bound (95%).
The results obtained from this research show that both audio speed and adjunct images influence cued-recall, independently. However, the two independent variables did not interact significantly in influencing cued-recall. This finding is inconsistent with the predictions that under the highest audio-speed conditions, the presentation of a representational adjunct image would have the greatest effect on cued-recall. The significant difference identified from audio speed was an expected outcome as previous research has shown that a learner’s ability to comprehend information begins to decline somewhere near 275 wpm (Fairbanks et al., 1957; Foulke & Sticht, 1967; Reid, 1968). The partial $\eta^2 = 0.12$ shows that approximately 12% of the variability can be explained by the audio speed, which is an interesting finding in that the fast audio treatments were approximately 300 wpm. This may be an indication that the digital time-compression algorithms employed have substantially improved the intelligibility of the audio content from older methods of time-compression used in the 1950s and 1960s (e.g., SOLA). The Tukey HSD follow-up procedure confirmed the significant differences were identified between the Very Fast speed compared to the Fast, Moderate, and Normal speeds. This suggests the ceiling may have been raised due to substantial improvements to time-compression technology, and that the ceiling effect is somewhere in the range of 300 to 375 wpm.

<table>
<thead>
<tr>
<th>Audio speeds</th>
<th>Mean difference</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Fast</td>
<td>Fast</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>-0.29</td>
</tr>
<tr>
<td>Fast</td>
<td>Moderate</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>-0.19</td>
</tr>
<tr>
<td></td>
<td>Very Fast</td>
<td>0.10</td>
</tr>
<tr>
<td>Moderate</td>
<td>Fast</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>Very Fast</td>
<td>0.17</td>
</tr>
<tr>
<td>Normal</td>
<td>Fast</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Very Fast</td>
<td>0.29</td>
</tr>
</tbody>
</table>

DISCUSSION

The results obtained from this research show that both audio speed and adjunct images influence cued-recall, independently. However, the two independent variables did not interact significantly in influencing cued-recall. This finding is inconsistent with the predictions that under the highest audio-speed conditions, the presentation of a representational adjunct image would have the greatest effect on cued-recall. The significant difference identified from audio speed was an expected outcome as previous research has shown that a learner’s ability to comprehend information begins to decline somewhere near 275 wpm (Fairbanks et al., 1957; Foulke & Sticht, 1967; Reid, 1968). The partial $\eta^2 = 0.12$ shows that approximately 12% of the variability can be explained by the audio speed, which is an interesting finding in that the fast audio treatments were approximately 300 wpm. This may be an indication that the digital time-compression algorithms employed have substantially improved the intelligibility of the audio content from older methods of time-compression used in the 1950s and 1960s (e.g., SOLA). The Tukey HSD follow-up procedure confirmed the significant differences were identified between the Very Fast speed compared to the Fast, Moderate, and Normal speeds. This suggests the ceiling may have been raised due to substantial improvements to time-compression technology, and that the ceiling effect is somewhere in the range of 300 to 375 wpm.
The significant difference on cued-recall identified from representational adjunct images was also an expected outcome as this is the basis for the multimedia principle of learning (Mayer, 2001). This research lends credence to the multimedia principle having a durable effect in educational research, even in conditions with accelerated information traversing the auditory/verbal channel. The proportion of variability explained by the presentation of an adjunct image is only 2% (partial $\eta^2 = 0.02$). This may be attributable to the purposeful design of the instrument having 10 feature-related and 10 nonfeature-related items. This research also demonstrates the value of other principles of multimedia learning. As mentioned, care was taken in this research to avoid the split-attention effect and also enforced the redundancy principle, which suggests individuals learn better from narration and images than from narration, images and onscreen text (Mayer, 2001).

Unlike the cued-recall findings, only audio speed significantly influenced content recognition. The main effect for adjunct image and the interaction effect between audio speed and adjunct image were hypothesized to have a statistically significant effect on content recognition. The audio speed had a minor partial $\eta^2 = 0.09$, and, again, the Very Fast condition resulted in significantly poorer performance than the other three audio speeds. The adjunct image not having a significant effect may be attributable to three things. First, the weak internal consistency reliability for this sample (K-R 20 = .63). Second, the purposeful design of instrument having 10 feature-related and 10 nonfeature-related items would have diminished the image effects. Third, there was a substantial amount of overlap in the confidence intervals for the two adjunct image conditions, which can be seen in Table 4. The lack of an interaction effect for both the cued-recall and content recognition task provides strong evidence that the two conditions may not interact in a meaningful way as to positively influence learning.

The mean scaled performance on the content recognition task is substantially higher than cued-recall, even though it was testing similar content. This is an indication that the nature of the recognition task may be “easier” for learners to complete when compared to cued-recall, and also an indication that guessing may have contributed to measurement error on the content recognition task. As pointed out by Haist, Shimamura, and Squire (1992), there are two general explanations for why recognition tasks are typically easier for learners than recall. One explanation, known as the Strength Theory (McDougall, 1904), suggests that recall tasks require more information in working memory than does recognizing. An alternative explanation, known as Generative-Recognize Theory (Hollingworth, 1913), suggests recall requires two processes: the retrieval of information from memory followed by a familiarity decision (Haist et al., 1992). In contrast, the recognition task only requires the familiarity decision.

The results from the current study can be best explained with the Generative-Recognize Theory. The strength and direction of the relationship between cued-recall and content recognition was $r = .79$ ($p < .01$), a strong and positive
correlation. This relationship indicates the parallelism of the content between the cued-recall and content recognition tasks in relation to the familiarity decisions. The items on both the cued-recall and content recognition instruments probed for similar information, but in different forms. The learners consistently performed better on the content recognition task as opposed to the cued-recall task, which may be attributable to the additional processing required to retrieve information into working memory.

The time-compressed audio condition had moderate effects on cued-recall and content recognition, explaining approximately 12% and 9% of the variability, respectively. This is an indication that the acceleration of information on the auditory/verbal channel is a generalizable effect on these measures. The adjunct image was less influential on cued-recall and content recognition, explaining only 2% and 1% of the variability, respectively. In this case, the adjunct image condition only significantly influenced the cued-recall measure. Thus, the adjunct image condition appears to be more helpful for the retrieval of information in working memory as opposed to a familiarity decision.

A statistically significant main effect was detected for both the audio speed and adjunct image conditions on learner satisfaction. The audio speed treatment shows 36% of variability explained by the time-compressed narration speed. As the audio playback speed was increased, this effect had a general negative influence on learner satisfaction. This is a powerful message in that the Tukey HSD follow-up procedures showed that all conditions were statistically different from the normal audio speed, which was shown to be most satisfying. The availability of a representational adjunct image had a positive influence on learner satisfaction. The learners presented with the picture were significantly more satisfied with the learning experience. However, the condition only explains 2% of the variability in learner satisfaction, and did not have enough strength to ameliorate the negative influence of the accelerated playback.

Recommendations to Stakeholders

Learners are ultimately the key stakeholders in this research. This research recommends that learners can choose to use time-compression technology, but should exercise this choice with extreme care and caution. The results demonstrate a generalizable negative effect on cued-recall and recognition after audio speeds two times the normal rate or over approximately 300 wpm. Additionally, this research has only employed subject matter that might be described as declarative knowledge or low intrinsic cognitive load. Using time-compression technology with complex subject-matter might lower the ceiling effect.

Learners should first identify which software and hardware devices are available that support time-compression technology. Common consumer products, such as iPods and personal computers with Window Media Player, have the technology readily available. Second, if available in the instruction, learners
should attempt to attend to representational pictures while listening to time-compressed narration as these research results indicate doing so will assist in the retrieval of relevant information into working memory. Third, learners should identify a speed at which they are most comfortable to assure a satisfying learning experience. As learners become comfortable at accelerated speeds, some research suggests that they can increase the playback to higher levels (Friedman, Orr, Freedle, & Norris, 1966; Voor & Miller, 1965).

From an educator and instructional designer perspective, this research highlights the importance of including representational adjunct imagery into instruction—the use of a relevant image was documented to show positive effects on both cued-recall and learner satisfaction. With the explosive growth in educational podcasts (e.g., iTunes University) and other audio-only media, more emphasis should be placed on developing instruction that includes both digital audio and pictures to improve a learner’s ability to retain the information. Current technology for creating podcasts (e.g., mpeg video files) already supports this functionality and can be authored using standard tools. While instructors and instructional designers should not assume their learners will make use of the imagery, providing the imagery and encouraging the use by learners is recommended. This research also encourages instructional designers and educators to be mindful about whether learners will choose to use time-compression technology when engaged in their instruction. Designing instruction with digitally recorded audio stored in appropriate formats (e.g., mpeg, mp3, mp4, etc.) affords learners the option to choose whether or not to use time-compression technology. Learners should not be coerced to use time-compression technology within their courses or training programs as this research shows that learners assigned to faster audio speeds were generally less satisfied.

This research also provides some guidance for future research efforts. For example, it did not provide evidence of a statistically significant interaction between audio speed and adjunct images as predicted. Future research efforts may seek to retest this hypothesis under different conditions to detect whether a representational adjunct image might serve as the secondary cue for verbal information under the fastest time-compressed audio constraints. The current evidence suggests there is no relationship. Ultimately, learners should have the choice to set the audio settings at a comfortable speed. Conceivably, doing so would not only improve a learner’s cued-recall and recognition of the information, but also serve to increase a learner’s level of satisfaction with the instruction. Future research should aim at designing experiments to include learner control as a meaningful variable.

The cued-recall and content recognition instruments included both feature-related and nonfeature-related items. Future research might only include feature-related items to observe the full effects of the multimedia principle in relation to time-compressed audio. Additionally, the topic selected, Australia, is classifiable
as low intrinsic cognitive load (Sweller & Chandler, 1994). As pointed out by Barron (2004), content type is likely a moderating variable.

Levin (1981) suggested images can serve as decorational, representational, organizational, and transformational. Representational images only mirror part or all of some related text, and, in previous research without time-compression, have been found to have moderate effects on learning (Carney & Levin, 2002; Levin, 1981). The use of organization images in the research interventions, such as maps or diagrams that illustrate magnitude (e.g., pie charts), may have had a more powerful impact on the dependent measures of interest.

One of the goals of this research was to connect the time-compressed speech research literature with multimedia research literature. Though the results indicate no relationship between the time-compressed audio speed and adjunct image conditions, the use of multimedia learning to explain previous time-compressed speech research has been documented in this study. Multimedia models can be used to explain, control, and predict the effects of time-compression technology on human learning. Future studies might elect to employ multimedia models as a framework to examine time-compression technology and human learning.

REFERENCES


Direct reprint requests to:

Albert Dieter Ritzhaupt
Assistant Professor, Instructional Technology
Dept. of Instructional Technology,
Foundations and Secondary Education
Watson School of Education
University of North Carolina, Wilmington
601 South College Rd.
Wilmington, NC 28403-5980
e-mail: artizhaupt@gmail.com