CONJOINT PROCESSING OF TIME-COMPRESSIONED NARRATION IN MULTIMEDIA INSTRUCTION: THE EFFECTS ON RECALL, BUT NOT RECOGNITION

ALBERT D. RITZHAUPT
University of Florida

ANN E. BARRON
University of South Florida

WILLIAM A. KEALY
University of North Carolina at Greensboro

ABSTRACT

Although previous research shows verbal recall of time-compressed narration is significantly enhanced when it is accompanied by a representational adjunct picture (Ritzhaupt & Barron, 2008), the reason for this increased performance remains unclear. One explanation, explored in the current study, is based on the Conjoint Retention Hypothesis (CRH), which posits that mentally stored visual information can serve as a secondary retrieval cue that boosts recall of related verbal material. Four groups of participants (N = 153) listened to a compressed audio narration at different rates of speed. The narration was accompanied by visuals, 50% of which were pictorially-related and 50% of which were pictorially-unrelated. Results show the type of information significantly influenced the recall, but not the recognition performance. While CRH provides the most feasible explanation for the increased recall, the generative-recognize view best explains the differences between recognition and recall performance.

© 2011, Baywood Publishing Co., Inc.
doi: 10.2190/EC.44.2.d
http://baywood.com
Humans can understand narration faster than they can speak. Previous research shows that conversational speech typically takes place at approximately 150 words per minute (wpm) (Benz, 1971; Nichols & Stevens, 1957) and that even the fastest speakers reach a physiological barrier at about 215 wpm (Beasley & Maki, 1976). By contrast, comprehension for accelerated speech only begins to significantly decline beyond a rate of 275 wpm (Foulke & Sticht, 1969). In conversation, one is concurrently listening and composing speech. Therefore, in circumstances where one is playing only a listening role, it is logical to assume that unused processing capacity might be available. This hypothesis has been tested by researchers under a variety of experimental conditions starting as early as the 1950s (Barabasz, 1968; Fairbanks, Guttman, & Miron 1957; Goldhaber, 1970). The findings of these various research studies suggest that speech speeds somewhere near 275 wpm or more begin to negatively influence the criterion measures (e.g., recall, recognition, comprehension, retention). However, these previous research studies did not study the effects of time-compressed narration in the context of a multimedia learning environment in which both pictures and words are simultaneously presented (Mayer, 2001).

While previous research largely relied on analog methods of compression, more recent research has employed digital time-compression algorithms in the deployment of multimedia interventions. Ritzhaupt, Gomes, and Barron (2008) investigated the effects of time-compressed narration on learner performance (content recognition) and satisfaction. The research design incorporated three audio speeds at 1.0 (150 wpm), 1.4 (210 wpm), and 1.8 (270 wpm) as between subject conditions while verbal redundancy—whether or not narrated material was duplicated as onscreen text—served as a repeated measure. Although the results indicated no differences in the recognition of narrated content due to variations in its rate of presentation, participants had significantly greater satisfaction with information heard at an audio speed that was 1.4 times faster than a normal pace. The researchers also identified a positive and significant effect in favor of verbal redundancy similar to previous research (Moreno & Mayer, 2002). However, the study did not isolate the effects of visual elements in the intervention.

As an alternative to supplementing compressed audio with linguistic redundancy (i.e., onscreen text), Ritzhaupt and Barron (2008) explored the effects of adding semantic redundancy to the narration through the use of representational adjunct pictures. In the study, four groups of participants heard narrated material at one of four audio speeds (1.0 = 150 wpm, 1.5 = 225 wpm, 2.0 = 300 wpm, or 2.5 = 375 wpm) with an adjunct picture either present or absent. The study purposefully included a speed (2.5) that would exceed the assumed human processing capabilities. The results demonstrated that the presentation of an adjunct picture significantly influenced the cued recall of the narrated content, but not the recognition of the information heard. Further, these differences between recall and recognition performance were only evident when presentation speeds exceeded approximately 375 wpm.
Conjoint Retention Hypothesis

In trying to understand why representational adjunct pictures influenced recall performance but not recognition, the researchers considered Kulhavy, Lee, and Caterino's (1985) Conjoint Retention Hypothesis (CRH) as the basis for a possible explanation. This theory attempts to account for the considerable research evidence that verbal information (i.e., material heard or read) is remembered significantly better when it is accompanied by adjunct displays (e.g., pictures, maps, graphs, animations) (Carney & Levin, 2002; Kulhavy, Stock, & Kealy 1993; Winn, 1990). Essentially, CRH is an extension of Paivio’s (1986) Dual Coding Theory (DCT), which states that information is mentally processed as two separate but mutually accessible codes: one linguistic and the other imaginal (Sadoski & Paivio, 2004). CRH further stipulates that because text with corresponding graphics has the capacity for being conjointly retained in long-term memory (LTM) storage, the latter serves as a secondary retrieval cue for information if an initial search of verbal storage proves unsuccessful. Graphical images theoretically aid recall of semantically-related verbal information because they can be held in working memory during a concurrent search of verbal storage in LTM without exceeding human processing capacity. An important provision of CRH is that adjunct displays only facilitate recall of verbal material that is semantically related to the visual imagery (Kulhavy, Stock, & Kealy, 1993).

Much of the research associated with CRH has not made a distinction between recall and recognition tasks as criterion measures of memory for what was read or heard. On the contrary, studies investigating various aspects of CRH have incorporated strikingly different dependent measures. Some researchers, for instance (Robinson, Katayama, & Fan, 1996; Robinson, Robinson, & Katayama, 1999), employed recognition tasks (e.g., multiple-choice questions) as the dependent measure. Alternatively, other researchers (Griffin & Robinson, 2000; Schwartz, Ellsworth, Graham, & Knight, 1998) have studied CRH through experiments involving verbal recall (e.g., free recall, cued recall) as the performance criterion.

Recall versus Recognition Tasks

The relationships and differences between recall and recognition memory tasks have been central to cognitive psychology research for more than a century (Haist, Shimamura, & Squire, 1992; Hall, 1983; Hollingworth, 1913; McDougall, 1904). Both recall and recognition are similar in that they depend on the construct of declarative memory (Haist et al., 1992), which can be described as a part of human memory that can store factual statements. Declarative memory is, in a sense, analogous to the linguistic elements in DCT in which verbal information either in print or auditory form is stored.

Prior research and practice have shown that learners generally find recognition tasks to be easier than recall tasks (Haist et al., 1992; Hall, 1983). Haist,
Shimamura, and Squire (1992) explain there are two general explanations for why recognition tasks are typically easier for learners than recall tasks. One explanation, known as the strength theory (McDougall, 1904), suggests that recall tasks require more information to be stored concurrently in memory than do simple recognition tasks. That is, the strength theory suggests there is a threshold at which recall requires greater demands than recognition on memory.

An alternative explanation, known as generative-recognize view (Hollingworth, 1913), suggests recall requires two separate internal processes:

1. the retrieval of information from LTM; followed by
2. a familiarity decision (Haist et al., 1992).

The familiarity decision requires an individual to match between the information encoded at the time of learning event and the information that is available at the time of making a decision. In contrast to recall tasks, the recognition task only requires the learner to make a familiarity decision.

PURPOSE OF RESEARCH

Because previous research has demonstrated that the presence of a representational adjunct image improves recall of textual information (Ritzhaupt & Barron, 2008), the primary purpose of the current study was to test whether picture-related information in a multimedia presentation could be better recalled and recognized under accelerated narrative speeds. Thus, the primary research question is: when learners hear a time-compressed narration that is supplemented with a representational adjunct picture, will they show improved recall and recognition performance for picture-related narrative information? Conceivably, the availability of information in pictorial storage should be particularly beneficial when narrative presentation rates approach speeds that potentially challenge verbal processing. A secondary purpose of this research is to examine whether the conjoint processing of the accelerated narration in a multimedia presentation would influence both recall and recognition performance. Prior research has used both tasks as criterion measures with mixed results when investigating the CRH (Griffin & Robinson, 2000; Robinson et al., 1996, 1999; Schwartz et al., 1998); thus, this research aims at further exploring the differences between recall and recognition tasks.

METHOD

Design and Participants

The experiment was a 4 Audio Speeds (1.0 = normal, NP vs. 1.5 = moderate, MP vs. 2.0 = fast, FP vs. 2.5 = very fast, VP) × 2 Type Information (Pictorially-Related vs. Pictorially-Unrelated) factorial design with Audio Speed serving as a
between-subject condition and Type Information serving as a repeated measure. Participants \((N = 153)\) were recruited from a public, southeastern university in the United States after making prior arrangements with instructors, and they received extra credit for their participation. Participants were randomly assigned to a treatment group resulting in the following distribution: NP = 37, MP = 38, FP = 38, and VP = 40.

Fifty-three percent of the participants were male. Ninety-four percent of the participants indicated English was their primary language. Four percent of the participants were freshmen, 20% sophomore, 50% juniors, 27% seniors, and the remaining participants were either graduate students or other. The ages of the participants ranged from 17 to 54 with an average age of 23.77 (SD = 6.86).

Materials

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; Ritzhaupt & Barron, 2008), its content was well-suited for the current research intervention, and the population had limited knowledge about specific destinations in Australia. The text has a Flesch reading ease of 36.4 and a Flesch-Kincaid twelfth-grade level reading score (Flesh, 1949). The narrative consisted of an introductory screen and 10 passages (of approximately 150 words per passage, as people speak approximately 150 wpm) describing different locations in Australia. Each passage was subdivided into two paragraphs. The first paragraph was composed of picture-related information, while the second paragraph was nonpicture-related information. The Discovering Australia text can be characterized as low intrinsic cognitive load (Sweller & Chandler, 1994), meaning that the subject matter is not intellectually challenging or difficult to comprehend. This is in contrast to other multimedia interventions that purposefully attempt to describe scientific subject-matter of cause-and-effect systems (Mayer, 1997), such as how lightning works or how a brake system works.

Ten semantically related pictures corresponding to the Discovering Australia text were selected for research intervention. The pictures were selected based on their appropriateness in representing one paragraph of the verbal information in each of the 10 passages and were previously rated by a panel of experts based on their representational quality (Ritzhaupt & Barron, 2008). Levin (1981) suggested that pictures can serve as decorative, representational, organizational, and transformational. Representational pictures mirror part or all of some related text, and have been found to have modest effects on learning (Carney & Levin, 2002; Levin, 1981). The pictures used in this study were purposefully representational in nature as they were intended to provide context relating to the passages in the Discovering Australia text.
For example, Figure 1 illustrates the city of Perth and the black swans in the Swan River to aid in visualizing the information from the passage in Figure 2. The relationship between the verbal description and the image of Sydney is intended to activate referential processing according to DCT (Pavio, 1986). The second paragraph in Figure 2 is not related to the image shown in Figure 1. 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.

**Criterion Measures**

Twenty constructed-response questions were created from the Discovering Australia text. An example of the picture-related recall (cued task) item pertaining to Figure 1 and Figure 2 is “What body of water does the city of Perth run along?” This item prompts for the recall of picture-related information and can be activated referentially. An example item pertaining to the nonpicture-related content would be “Why were the Aborigine people in Southwestern Australia displaced from their homelands?” This item pertains to verbal information in the passage and does not directly describe the information in the picture. A rubric was developed to assess each constructed response item based on a “gist” protocol (Schreiber, Verdi, Patock-Peckham, Johnson, & Kealy, 2002). 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. Interrater reliability was calculated on two occasions at 100% and 95%, respectively. Internal consistency reliability was calculated using Cronbach's alpha at $\alpha = .77$ for these data.

Twenty multiple-choice items were created based on the narrative serving as recognition (content). The multiple-choice questions were developed in a consistent format following established guidelines (Gronlund, 1998). Each stem posed a question for participants to consider, and the distracters were written with only one correct statement. Further, each item followed a similar design in: picture- vs. nonpicture-related. For example, a picture-related item pertaining to Figure 1 and Figure 2 would be:

Which body of water does the city of Perth run along?

a. The Coral sea  
b. The Great Australian sea  
c. The Swan River  
d. The Timor sea

A nonpicture-related item pertaining to Figure 1 and Figure 2 would be:

Which of the following best describes why Aborigine people in Southwestern Australia were displaced from their homelands?

a. The hostile encounters with European settlers  
b. The despairing droughts in the region  
c. The industrialization of the waterways  
d. The spreading malaria plague
The responses for the multiple-choice questions were scored dichotomously and internal consistency reliability was calculated for the scale using Kuder-Richardson 20 (KR-20) formula. This yielded a value of K-R 20 = .60 for these data, which was considered a satisfactory degree of reliability for a 20-item test.

**Procedures**

After making prior arrangements with teachers, a member of the research team visited classes to inform students that they could receive extra-credit by participating in the research program. 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. At the start of the research sessions, the researchers, again, briefly outlined the tasks to be performed and informed participants that the research was a voluntary, anonymous process. 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 clicked for a sound test. At that time, participants adjusted the volume of their program. Next, basic demographic information, including classification, gender, major, age, and whether English is their second language was collected for descriptive purposes.

After the demographic survey, the computer displayed a map of Australia for all participants (see Figure 3) and the introductory text at the assigned speed. On subsequent screens, participants were presented with a picture and one passage of narration. No other information was presented on the auditory or visual channels to control for a split attention effect (Chandler & Sweller, 1991). 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 to move to the next screen. To prevent an ordering effect, passages were randomly assigned by the program.

After completing the 11 passages of narration, participants were presented with three 3-column addition problems designed to preclude rehearsal of information in working memory immediately before testing. Participants were then instructed that they would have 45-seconds to provide responses to short-answer questions (recall task). 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. Next, participants were provided instructions about the multiple-choice, recognition task. The recall items were presented before the recognition items to prevent a testing effect. The items were presented in random order, one at a time until all items were completed.

Participants assigned to the faster audio speeds, FP and VP, 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 speeds 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 4 illustrates the instructions, intervention and data collection sequence.

RESULTS

Prior to analysis, scores for both recall and recognition were scaled from 0 to 1 to improve the interpretability of the results (percent correct). Skewness and kurtosis for both the recall and recognition measures were within an acceptable range (−1 to 1). A Levene’s test was conducted on both recall and recognition to test for homogeneity of the variance, and results showed no violations at $F(3, 149) = 2.175, p = .093$ and $F(3, 149) = 2.175, p = .093$, respectively. Since the
treatments were randomly assigned, it is tenable to assume there were no violations of independence.

**Recall Performance**

Table 1 illustrates the percent correct recall results by audio speed and information type. As can be gleaned, participants consistently scored higher on picture-related items as compared to nonpicture-related items. The recall scores were entered into a 4 Audio Speeds × 2 Type Information repeated measures ANOVA. The main effects for Type Information and Audio Speed were both significant at $F(3, 149) = 265.58, p < .01, \eta_p^2 = 0.64$ and $F(3, 149) = 8.31, p < .01, \eta_p^2 = 0.14$, respectively. Type information explained approximately 64% of the variability, while the audio speed explained approximately 14%. However, the interaction effect at $F(3, 149) = 1.964, p = .122$ was not statistically significant, indicating a non-relationship between the speed of the narration and type of information presented in the narration. A Tukey follow-up procedure showed that every group scored significantly higher than the fastest audio speed condition (VP), which equates to approximately 375 words per minute.

**Recognition Performance**

Table 2 shows the percent correct recognition results by audio speed and information type. Overall, participants scored higher on the recognition items than on the recall items. However, unlike recall, participants did not naturally respond higher to picture-related information. The recognition scores were entered into a 4 Audio Speeds x 2 Type Information repeated measures ANOVA. The main effects for Audio Speed was significant at $F(3, 149) = 6.786, p < .01, \eta_p^2 = 0.12$ and explains approximately 12% of the variability. However, the main effect for Type Information and the interaction effect were not significant at $F(1, 149) = .685, p = .41$, and $F(3, 149) = .18, p = .91$, respectively. As in the case with recall, a

<table>
<thead>
<tr>
<th>Combined</th>
<th>Picture-related</th>
<th>Nonpicture-related</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>NP</td>
<td>0.32</td>
<td>0.13</td>
</tr>
<tr>
<td>MP</td>
<td>0.28</td>
<td>0.13</td>
</tr>
<tr>
<td>FP</td>
<td>0.28</td>
<td>0.14</td>
</tr>
<tr>
<td>VP</td>
<td>0.18</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Tukey follow-up procedure showed that every group scored significantly higher than the fastest audio speed condition.

**DISCUSSION**

Interpretation of the results must be viewed within the limitations and delimitations of this study. Neither criterion measure employed in this research attempted to measure higher levels of learning. Essentially, the items purposefully targeted declarative knowledge. Therefore, these results will not generalize to content that might be characterized high intrinsic cognitive load (Sweller & Chandler, 1994). Further, the recognition performance measure lacked strong internal consistency reliability (> .7).

The only forms of control for a test effect were the random completion of items and the purposeful ordering of the recall task before the recognition task. The participants in this research were primarily English speaking college students and, thus, these results may not generalize to other populations. Finally, the representational images used in this research were selected based on their degree of relationship to the text according to a panel of experts and were not psychologically normed (i.e., association or complexity). In light of the stated limitations, this research has resulted in several interesting findings.

The results obtained from this research show that both the audio speed and type of information influence recall, independently. However, the two independent variables did not interact significantly to influence recall. The significant difference identified from audio speed was an expected outcome as previous research has shown that a learner’s ability to retain information begins to decline somewhere near 275 wpm (Barabasz, 1968; Fairbanks et al., 1957; Goldhaber, 1970). Specifically, a significant difference was detected at a speed of approximately 375 wpm with approximately 12% of the variability being explained by the audio speed. This is roughly 100 words more per minute than in previous

<table>
<thead>
<tr>
<th></th>
<th>Combined</th>
<th>Picture-related</th>
<th>Nonpicture-related</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>NP</td>
<td>0.56</td>
<td>0.15</td>
<td>0.44</td>
</tr>
<tr>
<td>MP</td>
<td>0.53</td>
<td>0.14</td>
<td>0.54</td>
</tr>
<tr>
<td>FP</td>
<td>0.54</td>
<td>0.15</td>
<td>0.54</td>
</tr>
<tr>
<td>VP</td>
<td>0.43</td>
<td>0.13</td>
<td>0.56</td>
</tr>
</tbody>
</table>
research. The finding provides further evidence that the digital time-compression algorithms employed have substantially improved the intelligibility of the audio from older analog methods of time-compression (Ritzhaupt & Barron, 2008).

In terms of recognition, only audio speed significantly influenced the dependent measure, while the type of information was insignificant. Notably, the performance on the recognition task is also substantially higher than recall, even though it was measuring similar content. Again, the audio speed accounted for about 12% in the variability of performance differences in which the VP (375 wpm) condition resulted in significantly poorer performance than the other three audio speeds.

These results provide evidence in support of the CRH in that learners were able to better recall picture-related information. On the recall task, approximately 64% of the variability was explained by the type of information, which attests to the strength of conjointly encoded information and its availability for recall. However, the effect was only durable for the recall of picture-related information and did not manifest itself with the recognition of picture-related information. This finding has serious implications for cognitive psychology and educational research in general. Why is it that pictures facilitate the recall of information, but fail to facilitate the recognition of information?

One possible explanation is that the verbal information stored in long-term memory has a referential connection to the pictorial information via a dual-coding process. When the learner is presented with the recall task, the pictorial information serves as a secondary cue to retrieve the relevant verbal information into memory, which is the basis of the CRH (Kulhavy et al., 1985). However, when the learner is presented with the recognition task, the need for the pictorial information to serve as a secondary retrieval cue may be unnecessary since only a familiarity decision must be made as explained by the generative-recognize view (Hollingworth, 1913). Therefore, the conjoint processing may not assist when a learner is presented with recognition task as opposed to recall task.

The results from the current study can be best explained with the generative-recognize view of recall and recognition. The strength and direction of the relationship between recall and recognition was \( r = .78 \) (\( p < .01 \)), a strong and positive correlation. This relationship indicates the inherent parallelism of the content between the recall and recognition tasks learners had to make in relation to the familiarity decisions. The items found on both the recall and recognition instruments probed for similar information from the learner, but in different forms. The learners consistently performed better on the recognition task as opposed to the recall task, which speaks of the additional cognitive processing required to retrieve information from LTM (e.g., recall task).

Another topic of interest is that of cognitive load, referring to the amount of information that can be processed in working memory at one instance of time (Chandler & Sweller, 1991). In many ways, time-compressed narration is a deliberate test of cognitive load on a verbal/auditory channel. In reframing the
discussion on the highest audio speed, it appears that humans cannot process speeds of somewhere in the range of 300 to 375 wpm. This has an implication for the design of multimedia instruction and how much verbal information should be presented over a verbal/auditory channel. Designers should refrain from accelerating the speed of instruction past 300 wpm as evidenced by this research.

This research has added to what is known about the nature of time-compressed narration integrated into multimedia instruction and has provided further evidence in support of the CRH as applied to recall tasks. This research also shows that the CRH may not be applicable to the recognition of verbal information, and provides a viable explanation for this phenomenon. Future research is necessary to further explore whether conjoint processing may be applicable to recognition tasks.

REFERENCES


Direct reprint requests to:

Dr. Albert D. Ritzhaupt  
School of Teaching and Learning  
College of Education  
University of Florida  
2423 Norman Hall  
P.O. Box 117048  
Gainesville, FL 32611  
e-mail: aritzhaupt@gmail.com