The effects of time-compressed audio and verbal redundancy on learner performance and satisfaction

Albert D. Ritzhaupt a,*, Neil D. Gomes b, Ann E. Barron b

a University of North Carolina Wilmington, Department of Instructional Technology, Foundations, and Secondary Education, 601 S. College Rd, Wilmington, NC 28403-5980, United States
b Instructional Technology Program, Secondary Education, College of Education, University of South Florida, United States

Available online 24 March 2008

Abstract

Digital audio is becoming increasingly popular in higher education with faculty digitally recording and broadcasting lectures for students to learn-on-demand. Students have discovered accelerated playback features in popular computer software and use it to reduce the amount of time spent listening to audio-enhanced instruction. In the current study, 183 undergraduates were randomly assigned to one of three audio-enhanced multimedia presentations that were recorded at three speeds (1.0, 1.4, and 1.8). Results show no significant difference on performance across treatments and a significant difference on satisfaction in favor of 1.4 times the normal audio speed. The results also indicate statistical differences in favor of verbal redundancy, in which the same verbal information was presented on both an auditory and visual channel.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Time-compressed audio; Multimedia learning; Verbal redundancy; Learner performance; Learner satisfaction

* Corresponding author. Tel.: +904 859 9799; fax: +910 962 3609.
E-mail address: aritzhaupt@gmail.com (A.D. Ritzhaupt).

0747-5632/S - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2008.02.017
1. Introduction

Multimedia can be defined as the presentation of information using both words and pictures (Mayer, 2001). The design and delivery of multimedia learning is based on principles and guidelines derived from theory, empirical research, professional experience, and the limitations or possibilities of the technologies used for delivery (Sabatini, 2001). As technology changes, further empirical research and theory development is 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 a continual process.

In particular, audio incorporated into multimedia programs is an area that requires ongoing research. Audio that is used for instruction can be categorized into three main elements: narration, sound effects, and music (Beccue, Vila, & Whitley, 2001). Narration is the speech or dialog that can be used to deliver an instructional message; however, when narration is incorporated into a multimedia program, it can actually increase the time required by a learner to complete the program (Barron & Kysilka, 1993; Koroghlanian & Sullivan, 2000).

The goal of an instructional designer is to maximize a learner’s comprehension and satisfaction, while minimizing the amount of time a learner will spend on a learning task. This poses 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); however, studies have also demonstrated that normal speech can be increased to 200–300 wpm, with minimal loss in comprehension (Barabasz, 1968; Foulke & Sticht, 1967; Goldhaber, 1970).

The current body of research on the use of time-compressed narration or speech dates back to the 1950s (Fairbanks, Guttmann, & 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 that have not converged, though they investigate similar phenomena.

Although this gap in the body of research remains, faculty members are digitally recording voice-over presentations (e.g., PowerPoint with voice), animated screen captures with narration (e.g., Camtasia), and lectures to distribute to personal computers and other portable media devices (e.g., iPods) so students can learn-on-demand (Gill, 2007). Students have discovered time-compression technology integrated into popular consumer products (such as iPods) and software (such as Windows Media Player). The key digital technology that supports the increased or decreased rate of speech, while preserving pitch, in audio files is called a time-compression algorithm (He & Gupta, 2001). A major tenet of time-compression research is to provide learners with the ability to speed up or slow down content based on their preferences.

2. Relevant literature

2.1. Time-compressed speech

People can understand speech faster than narrators can speak – even the fastest talkers reach a physiological barrier at about 215 wpm (Beasley & Maki, 1976). Since, in conver-
sation, one is simultaneously listening and composing speech, it was assumed that perhaps another 125–150 wpm of unused processing capacity might be available in simple listening. In fact, Fairbanks et al. (1957) found that good intelligibility is possible when speech is compressed up to 2 times its original time length.

Barabasz (1968) conducted a study with 118 students in a human behavior and development class. Two lectures were used in a rotational research design to control for intergroup differences. The research investigated two different speeds and used both recall (administered after lecture) and retention (administered two weeks later) as dependent measures. The findings suggest that a lecture can be reduced to one-third the time without a significant decline in either recall or retention (Barabasz, 1968).

Goldhaber (1970) studied the effects of compressed speech as a function of academic grade level. In particular, the study looked at speech delivered at 165 wpm and 330 wpm for students in junior high school (80) and college (80), with comprehension as measured by a 20-item test. The results showed main effects for speech and academic level, but no interaction effect was identified. This indicates that individuals with varying levels of formal education perform differently (high school versus middle school), as one might anticipate.

Reid (1968) studied the effects of grammatical complexity and compressed speech on comprehension. He used a form of the Nelson–Denny reading test to make two difficulty levels of grammatical complexity and compressed speech at 175, 275, 325, and 375 wpm. Further, the verbal scholastic aptitude test was used as a covariate. His findings suggest a significant difference for both compressed speech and grammatical complexity and their interaction. Compressed speech did not result in a statistically significant decrease in comprehension until the 375 wpm level.

Previous time-compressed speech studies suggest that speeds somewhere in the range of 270–300 wpm begin to negatively influence the dependent measures of interest (e.g., comprehension, recall, etc.). These studies also underscore extraneous variables that may influence the dependent measures of interest. The studies, however, did not investigate the effects of time-compressed speech in the context of multimedia learning (pictures and words).

2.2. Audio narration in multimedia learning

Multimedia learning has been investigated from many different perspectives and has evolved from simple media comparison studies to a basis for explaining learning. Previous research in multimedia focused on the medium used for delivery rather than the instructional interventions that positively influence learning (Clark, 1983). A handful of studies have attempted to investigate the effects of audio narration in multimedia learning and often have produced inconsistent results, especially pertaining to verbal redundancy, in which the same verbal information is presented on a visual and auditory channel.

Severin (1968) designed a series of treatment conditions using the tenets of cue summation theory, which posits that the addition of a second channel (audio or visual) results in better learning. Severin’s treatment conditions were: audio with relevant pictures, audio and unrelated pictures, picture-only, audio-only, and audio and print. The sample consisted of 246 middle school students with recognition as the dependent measure. Results demonstrated that the related audio and picture condition was significantly different from the audio and print condition, and the picture-only treatment was significantly different
from the audio-only treatment. He concluded that the condition with audio and print was effectively redundant in nature and did not lead to better learning because the information was processed on the same channel, which interfered with the learning process (Severin, 1968).

Barron and Kysilka (1993) examined the effects of three different treatment groups of audio in multimedia learning with a sample of 60 college students: a visual text-based version, full audio and visual version in which the text accompanied a word-for-word narrative description, and a version with both text and audio with the text presented in a synthesized bulleted form. Their findings demonstrated no difference in achievement, with or without the inclusion of an audio channel. However, the treatments with audio required significantly more time to complete than the treatment without audio. Perceptions among the learners were positive and relatively comparable in all treatments.

Moreno and Mayer (2002) studied the effects of verbal redundancy in multimedia learning using narration. Their study consisted of three separate experiments that specifically studied the effects of verbal redundancy using a combination of narration, on-screen text, and pictures (animation) in eight different groups. Results show that students consistently scored better when presented with words in a visual and auditory form, indicating that verbal redundancy had a significant positive effect on retention, transfer, and matching in these experiments, provided there were no other concurrent visual elements (e.g., animations) on the computer screen. These findings appear to conflict with previous research (Barron & Kysilka, 1993; Severin, 1968); however, the Moreno and Mayer study employed audio treatments that were much shorter in duration, provided little learner control, and attempted to control for a split-attention effect.

3. Theoretical rationale and purpose

Mayer’s cognitive theory of multimedia learning is based on three tenets: dual-channels, limited capacity, and knowledge construction. The first tenet, dual processing, suggests that humans have multiple separate channels for processing visual/pictorial and auditory/verbal information (Mayer, 2001; Mayer, 2003). The second tenet suggests that humans’ processors have a limited capacity to process information at any given instance in time. The third tenet is that humans have knowledge-constructing processors that receive, organize, and connect incoming information with existing knowledge (Mayer, 2001; Mayer, 2003).

The process of learning from multimedia involves five cognitive processes: selecting words, selecting images, organizing words, organizing images, and integrating (Mayer, 2003). According to the cognitive theory of multimedia learning, when a learner engages in a multimedia presentation, information is presented as either words or pictures, impinges the eyes and ears of the learner, and is momentarily stored in sensory memory. Next, the learner selects the information from sensory memory and moves it into working memory for organization. If the learner organizes the information in working memory and connects it with prior knowledge, an “integrated learning outcome” results (Mayer, 2003, p. 304).

A multimedia presentation using time-compressed audio presents words to the auditory channel at an increased rate. Thus, the auditory–verbal channel must withstand a disproportionate amount of information, and the amount of verbal information that can be selected for movement into working memory is limited. According to dual-coding theory
(DCT), visually presented information is initially processed in visual working memory, while information presented auditorily is initially processed in auditory working memory (Paivio, 1986). It is therefore tenable that the presentation of limited verbal information on a visual–verbal channel might serve as a secondary cue for the narrative verbal information by activating a referential process between the visual and auditory representations, indicating a coherent verbal representation was formed (Moreno & Mayer, 2002; Paivio, 1986).

This study attempts to explore this phenomenon by investigating the effects of varying audio speeds and verbal redundancy on learners’ performance and satisfaction when they listen to accelerated audio in multimedia presentations. By shedding light on these issues, instructors and instructional designers will be better able to implement time-compressed audio while creating appealing, time-efficient, audio-enhanced multimedia instruction. More specifically, the research question is: “What is the relationship among various audio speeds and verbal redundancy in a multimedia presentation on various criterion measures (performance, satisfaction)?”

4. Method

4.1. Design and participants

The design of the experiment is a 3 audio speed (1.0 = normal vs. 1.4 = fast vs. 1.8 = fastest rate) × 2 redundancy level (redundant vs. non-redundant) × 2 trial (trial 1 vs. trial 2) factorial. Audio speed serves as the between-subjects measure, and both trial and redundancy serve as repeated measures. The three different audio speeds, length of presentations, and the estimated wpm can be seen in Table 1.

The research participants included 183 undergraduate students in introductory information and communication technology literacy courses in a College of Education and College of Business at two public southeastern universities. Participants received extra credit in their courses for their participation. Sixty-three participants completed the session at a distance while the remaining 120 completed the session within a controlled computer lab. The distance participants were enrolled in an online course, and thus, did not have the option to choose between formats.

The average age of participants was 21.41 (SD = 4.8), and 71% of the sample was female. Most participants did not have prior podcasting experience, with 97% reporting they never created a podcast, and 78% reported never listening to a podcast. Ninety-six percent of the sample reported English as their primary language. Participants reported varying prior experience with basic information and communication technology. For example, 72% had experience with spreadsheet applications, 54% with database appli-

<table>
<thead>
<tr>
<th>Audio rate</th>
<th>Presentation length</th>
<th>Estimated wpm*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (1.0)</td>
<td>20:54</td>
<td>150</td>
</tr>
<tr>
<td>Fast (1.4)</td>
<td>14:59</td>
<td>210</td>
</tr>
<tr>
<td>Fastest (1.8)</td>
<td>11:40</td>
<td>270</td>
</tr>
</tbody>
</table>

* Estimated words per minute based on intervention speed.
tions, 88% with computer games, 93% with web browsers, and 100% with email. Interestingly, approximately 90% of the participants reported they had been using the web between 4 and 12 years, and 60% reported they use the web for eleven or more hours per week.

4.2. Materials

4.2.1. Computer tutorial

A narrated, web-based presentation was developed on the topic of Podcasting in Education. This topic was selected because it was relevant to the courses from which the participants were selected, and because podcasting was new, and therefore the majority of the students did not have prior knowledge in the content area. The presentation was developed using the underlying principles of multimedia learning: dual-channels, limited capacity, and working memory (Mayer, 2001; Mayer, 2003). A male voice provided narration throughout the presentation. The narration has exactly 2,453 words written with a Flesh Reading Ease measure of 39.4 and a Kincaid-Flesh 12th grade reading level (Flesch, 1949).

Fig. 1 provides a screen shot from the Podcasting in Education presentation. The presentation has eight screens and advances automatically with an average of two seconds latency from the end of one screen to the start of the next screen, irrespective of the audio speed. The advancement from one screen to the next is triggered by the completion of the narrative information for a particular screen. The transition latency between screens (elapsed time from the end of one screen to the start of the next) was not included in the calculation of the estimated words per minute (see Table 1).

The presentation was designed to provide both redundant information and non-redundant information, where redundant information was defined as on-screen text and narration. To illustrate how verbal redundancy was manifested in the presentation, Fig. 2
provides segments of the narration for the screen illustrated in Fig. 1, and corresponding performance items. Verbal redundancy is implemented as the same key verbal information presented simultaneously both on-screen and within the narration. Non-redundant verbal information is only presented in the narrative content, and is not presented on-screen. Therefore, verbal redundancy is not manifested as a complete replication of the same verbal information, but rather the same keywords to activate the referential process of the visual and auditory representations.

4.2.2. Instruments

Three instruments were used to collect relevant information: background survey, performance exams, and a satisfaction survey. The background survey collected information related to a participant’s prior computing experience, age, previous experience with podcasting, and whether English is their primary language. Some of these characteristics could potentially serve as confounding variables (e.g., whether English is primary language).

The performance measure included 10 multiple-choice items that directly measured exclusively narrative content (non-redundant), along with 10 items that measured content that was presented through simultaneous on-screen text and narration (redundant). Each item had four distracters, and participants were instructed to select the “best” option. The items and distracters were randomized in both trial 1 (pretest) and trial 2 (posttest) formats. The Kuder–Richardson 20 for the 20-item trial 2 performance quiz was K–R 20 = .67. The instruments were implemented using web-based HTML forms, a server-side scripting language (PHP 4.0), and a relational database (MySQL).

The satisfaction instrument consisted of 15 items from a previous study that were adapted for the current study (Lyskawa, 1998). The first part of this instrument uses a five point scale with two bipolar adjectives on both sides. For instance, on the left-most side is the word “obscure” and on the right-most side is the word “clear.” The second part of the instrument uses a modified Likert scale, ranging from strongly disagree to strongly agree. The items were designed to measure a learner’s satisfaction with the intervention. For instance, one item states “The narrator spoke clearly in the Podcasting in Education tutorial.” The Cronbach’s $\alpha$ for the satisfaction survey was $\alpha = .92$. The internal consistency reliability for the overall satisfaction survey was high, and the posttest had a marginally acceptable internal structure.
4.3. Procedures

The research intervention was delivered in two formats: distance and traditional. The traditional procedure included verbal instructions from a session moderator who visited classes in a controlled computer lab. The distance format included text instructions and screenshots for participants to review before starting the online tutorial. Students enrolled in an online computer literacy course represented the distance format (n = 63). None of the participants were prompted for any identifying information, and all participants were assured of anonymity.

Prior to beginning the Podcasting in Education presentation, each participant completed the background survey and a trial 1 performance exam. Trial 1 was administered as a way to control for inter-group differences, and intended to demonstrate the fidelity of the intervention with the expectation that performance would significantly improve on the trial 2 exam. All participants were randomly assigned to a time-compressed presentation intervention. Each participant was automatically assigned a unique sequential integer, and the computer programs used modulus arithmetic to randomly assign each participant to a different audio speed treatment (1.0, 1.4, and 1.8). Participants were not informed of which speed they were assigned, and the researchers were also unaware (a double-blind random assignment). After completing the presentation, participants responded to a trial 2 performance exam and a satisfaction survey, which was designed to measure learners' attitudes toward the audio-enhanced presentation and their comfort level with their assigned audio rate. Fig. 3 illustrates the procedures for the research study.

5. Results

Prior to inferential analysis, performance scores were scaled from 0 to 1 to improve the interpretability of the results (proportion correct on multiple-choice performance assessments), which served as the dependent measure for performance in this study. Skewness and kurtosis for trial 1 and trial 2 performance were .90 and 1.68, and −.26 and −.32, respectively. The data did not exhibit any severe departures from normality. To ensure there were no inherent differences between delivery formats (the experiment was delivered both in computer lab setting and from a distance), an ANOVA procedure was conducted on Format for trial 2 performance and satisfaction. The results indicated there was no significant difference between the trial 2 performance of the students from the two formats at $F(1,182) = 3.092$, $p = .08$, partial $\eta^2 = .017$ or on satisfaction $F(1,182) = 2.193$, $p = .09$, partial $\eta^2 = .016$. Since there was no main effect on either performance or satisfaction, it was deemed tenable to pool the data from both formats.

The performance measures were analyzed with respect to participant age, prior experience with podcasting, and whether English was the participant’s primary language. Mild correlations were identified between trial 1 ($r = -.052$, $p = .486$) and trial 2 ($r = -.048$, $p = .528$).
p = .519) performance and age, thus age was not included as a potential statistical control. English as a second language \(F(1,182) = .937, p = .334\), and prior podcasting experience \(F(1,182) = .286, p = .594\) did not exhibit statistically significant effects on the trial 2 performance results, and thus, were also discounted as confounding variables.

The scaled performance data were entered into a trial × redundancy × audio speed repeated measures ANOVA with trial and redundancy serving as within subjects effects, and audio speed serving as a between-subject condition. The results indicate that trial was statistically significant \(F(1,180) = 360.947, p < .01\), partial \(\eta^2 = .667\). As anticipated, Performance after the Podcasting in Education tutorial improved significantly from trial 1 to trial 2 with approximately 67% of the variability explained. The mean scores and standard deviations for Performance on trial 1 and trial 2 were .407 (.132) and .63 (.159), respectively. Table 2 provides the number of participants, means, and standard deviations of both performance and satisfaction by trial and audio speed.

When investigating the effects of audio speed in isolation, the results indicated the absence of a statistically significant main effect for audio speed \(F(2,180) = 2.158, p = .118\), partial \(\eta^2 = .023\). However, the means and standard deviations between groups on trial 2 performance did show mild variation (see Table 2). The results also indicated there was no significant interaction effects \(F(2,180) = .355, p < .702\), partial \(\eta^2 = .004\). This indicates that audio speed did not interact with trial or the prior knowledge between groups.

Redundancy must interact with trial to demonstrate the effects of redundant versus non-redundant information as the effects can only be detected after participants experience the intervention. The results indicate that the redundancy × time interaction was statistically significant \(F(1,180) = 158.13, p < .01\), partial \(\eta^2 = .468\). This interaction effect indicates that on-screen text and narration redundancy was superior to narration only from trial 1 and trial 2 performance (post intervention). Approximately 47% of the variability was explained by the interaction as demonstrated by the partial \(\eta^2\). Fig. 4 illustrates the greatest changes from trial 1 to trial 2 on the redundancy condition across audio speeds. As can be gleaned, the redundancy condition consistently outperforms the non-redundant condition across the audio speed treatments.

One possibility is that the learners acquired information primarily from the redundant information during the intervention. This would suggest participants did not acquire information in the faster audio conditions. Thus, an ANOVA was executed on audio speed for only non-redundant information on trial 2 performance. No statistically significant differences were identified: \(F(2,180) = 1.769, p = .173\), partial \(\eta^2 = .019\). This suggests the learners were able to acquire audio-only information regardless of the audio speeds.

Table 2
Performance and satisfaction by trial and audio speed

<table>
<thead>
<tr>
<th>Audio speed</th>
<th>n</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Normal (1.0)</td>
<td>60</td>
<td>0.418</td>
<td>0.123</td>
<td>0.629</td>
</tr>
<tr>
<td>Fast (1.4)</td>
<td>63</td>
<td>0.377</td>
<td>0.131</td>
<td>0.612</td>
</tr>
<tr>
<td>Fastest (1.8)</td>
<td>60</td>
<td>0.428</td>
<td>0.139</td>
<td>0.651</td>
</tr>
</tbody>
</table>

\(M = \text{mean}, \ SD = \text{standard deviation}, \ n = \text{number of participants}.\)
The 15 items from the satisfaction survey were summed and scaled from 0 to 1 to serve as an overall satisfaction indicator, with higher values serving as a sign of greater satisfaction. Skewness for the satisfaction indicator was $-0.20$ and kurtosis was $-0.39$. A Levene’s test $F(2, 180) = 0.471, p = .625$ was conducted to test for homoscedasticity. No violations of normality or equal variances were identified. The mean satisfaction indicators and standard deviations for audio speeds 1.0, 1.4, and 1.8 were $0.687 (0.137)$, $0.712 (0.146)$, and $0.647 (0.151)$, respectively.

An ANOVA was conducted on audio speed, with audio speed serving as a between-subjects effect. The results indicated that audio speed was statistically significant $F(2, 180) = 3.134, p = .046$, partial $\eta^2 = 0.034$ on learner satisfaction. A Tukey HSD follow up procedure indicated that the fast audio speed (1.4) was significantly more (mean difference $= 0.0648, p = .036$) satisfying than the fastest audio speed (1.8), but not the normal speed (1.0).

6. Discussion

Interpretation of the results must be viewed within the limitations and delimitations of this study. The study did not attempt to isolate the effects of the pictorial information in the multimedia interventions. Further, the performance measure was essentially content recognition, and lacked strong internal consistency reliability ($K-R 20 > .7$). The only form of control for a test effect was the scrambling of items and distracters from trial 1 to trial 2.

It was assumed that research participants did not have hearing impairments that might render the audio interventions unintelligible and did not have extensive previous experience with time-compressed narration (e.g., training effect). The results of this study should not be generalized outside of the population of undergraduate students in higher education or populations with similar demographics. Additionally, the type of content employed
in this study would likely be characterized as low intrinsic cognitive load (Chandler & Sweller, 1991), indicating the degree of difficulty of the content may have an interaction effect with the audio speed treatments.

From theoretical and practical perspectives, this research resulted in several important findings. First, the results of this study indicated no statistically significant differences among the three audio speeds on performance. This is an especially important finding as it suggests that learners can manipulate audio playback speeds without adversely influencing performance, and instructional designers can incorporate this novel technology into multimedia programs without fear of impacting achievement.

However, learner performance is only one dimension. Although learners had similar levels of satisfaction with the *Podcasting in Education* presentation irrespective of the audio speed, statistical significance was detected in favor of the fast audio speed at 1.4 times the audio normal rate when compared to 1.8 times the normal audio speed. This variability in satisfaction among various audio speeds suggests that although performance was not statistically significant, learners may have audio speed preferences, highlighting the importance of incorporating learner control in multimedia programs. Indeed, future research studies could incorporate options whereby students can select and adjust their audio speeds.

One of the most important results of the research is that it supports the use of verbal redundancy in multimedia. Statistical significance was detected in favor of information that was presented in both auditory and visual form, a referential process in multimedia learning. This finding is contrary to some research (Barron & Kysilka, 1993; Severin, 1968) and complementary to others (Moreno & Mayer, 2002). Because of these inconsistent findings, further empirical observation is necessary to test the hypothesis. Notably, this study used limited verbal information presented on-screen, while the narrative information was much more comprehensive in nature to minimize split-attention. In addition, the content was primarily conceptual, and the questions were based on content recognition.

The increased popularity of audio being incorporated into multimedia programs, coupled with audio time-compression technology, underscores the importance of this research. Learners are using audio time-compression technology to control the speed of multimedia presentations and reduce the amount of time spent on the learning task (Glabraith, Ausman, Liu, & Kirby, 2003). It is imperative that educators determine at what point (if any) this technology places learners at a disadvantage. In the current study, performance was not adversely influenced at 1.8 times the normal rate, but statistical significance was identified in terms of satisfaction in favor of 1.4 times the original rate. Future research might attempt to determine a threshold, incorporate learner control over the audio speeds, and focus on different types of content.

References


