



Learning efficacy of simultaneous audio and on-screen text in online lectures

Justin C. W. Debusse, Andrew Hede and Meredith Lawley
University of the Sunshine Coast

This study investigates the application of voice recognition technology to online lectures focusing on the efficacy of the text component of a multimedia presentation. Specifically, participants were provided with online access to multimedia instructional packages comprising an image of the lecturer with accompanying computer slides, plus simultaneous scrolling text of the words spoken during the lecture. Participants' knowledge was measured before and after the lecture presentation. Contrary to cognitive load theory, the results did not show a negative redundancy effect, that is, there were no differences in learning efficacy between the conditions with and without on-screen text. Further, participants found no difference between text edited for semantic breaks compared to unedited text. The implications for online instructional design are that resources are better spent providing a combination of audio and slides rather than text and slides, and that if text is provided then editing for semantic line breaks is not warranted.

Introduction

In recent years voice recognition technology, which automatically translates spoken words into textual data, has found increasing acceptance in modern society in a variety of applications (Attaran, 2000; Buckler, 2001; Christensen & Hughes, 2007; Kim & Lee, 2007; Marshall, 2002). Education is a key area within which voice recognition technology can be usefully applied, commonly through producing transcriptions of spoken lectures which can be displayed simultaneously together with lecture slides and audio narration. Although such approaches may be primarily aimed at students with disabilities (Leitch & MacMillan, 2001; Yong, 2007), they also have the potential to improve the learning of the wider student community. This efficacy improvement is likely to be attractive to adults studying online who may have to fit their learning around a work schedule and who may also be studying using English as their second language. The present research, therefore, aims to assess the efficacy of lecture transcription technology in the increasingly important area of online education (Byrnes & Ellis, 2006).

This paper examines the background to the study including the literature on multimedia effects on learning and specifically the use of simultaneous screen text in multimedia presentations. Based on this review, the gaps in the literature and hence the research questions that form the basis of this study are discussed. The third section outlines the experimental design adopted for the study, with the fourth section reporting the results. Finally, the implications of the findings are considered and conclusions are drawn.

Background

Multimedia effects on learning

Despite the existence of an extensive literature on multimedia and hypermedia since the mid-20th century, the effects on learning are not fully clear (Dillon & Gabbard, 1998; Hsiu-Ting, 2009; Liao, 1998, 1999; McNeil & Nelson, 1991). The most commonly used combination in instructional multimedia is simultaneous auditory and visual input. As one review pointed out in the mid 1990s, "Forty years of research has yielded a hodgepodge of contradictory conclusions" (Lang, 1995; p. 86). That is, the results from half the studies show that learning is better when both audio and video information is presented, while the other half show that learning is worse. Indeed, the five most widely assumed benefits of multimedia learning have been questioned in a major review by Clark and Feldon (2005).

From such research inconsistencies, it is clear that various contingent factors must be involved in the dynamics of multimedia effectiveness. An integrated model of multimedia effects on learning has been proposed to take account of the many contingency variables identified in the literature, namely: visual input, auditory input, learner control, attention, working memory, long term storage, motivation, cognitive engagement, learner style, intelligence and reflection (Hede, 2002a; Hede & Hede, 2002). A number of these contingencies have been further investigated in various studies since the mid-1990s and although their implications for multimedia design have been somewhat clarified (e.g., Mayer & Moreno, 2003) the picture is still not totally clear.

The main effects that have been identified as influencing audiovisual learning efficacy are as follows: split attention effect, modality effect, redundancy effect, segmentation effect, pre-training effect, coherence effect, signalling effect, spatial contiguity effect, temporal contiguity effect and the spatial ability effect (see Mayer & Moreno, 2003). Another effect that further complicates the picture is the expertise reversal effect (Kalyuga & Sweller, 2004; Kalyuga, Ayers, Chandler & Sweller, 2003) whereby learner experience interacts with a number of the other effects to change the direction of their influence on learning. The most comprehensive explanation for the various effects of multimedia presentation is the cognitive theory of learning which is based on three assumptions: 1) that human information processing involves two distinct channels for verbal and visual input; 2) that each channel has a limited capacity; and 3) that learning requires considerable active processing of information (Clark & Mayer, 2003; Mayer & Moreno, 2003).

Of the numerous above effects, let us consider the three which are of paramount importance in their implications for multimedia instructional design. First, the *split attention effect* (Ayers & Sweller, 2005; Sweller, 1999) refers to the finding that inferior learning occurs when one's attention has to be divided between two information sources within the one modality, for example, between visually presented animation plus simultaneous on-screen text (see also Kalyuga, Chandler & Sweller, 1999; Mayer & Moreno, 1998; Mousavi, Low & Sweller, 1995). Note that split attention has a *negative* effect on learning.

The second effect to consider is the *modality effect* which denotes the finding that multimedia presentation results in superior learning when pictorial information (i.e., visual modality) is accompanied by narration (i.e., auditory modality) rather than by

on-screen text (i.e., dual verbal presentations in the visual and auditory modalities) (Low & Sweller, 2005; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Moreno, Mayer, Spires & Lester, 2001). Thus, the modality effect is *positive* insofar as two separate modalities can be better for learning than one (see also Mayer & Sims, 1994; Mayer & Anderson, 1991; Tindall-Ford, Chandler & Sweller, 1997). However, a reverse modality effect was reported by Tabbers, Martens and van Merriënboer (2004) in a non-laboratory setting, again demonstrating the existence of complex confounding variables in multimedia instruction. Of relevance in this context is the issue whether the main operational modalities in multimedia processing are on the one hand, visual-versus-auditory (as proposed by Penney, 1989) or on the other hand, pictorial/visual-versus-verbal (viz., both visual text and auditory narration) as proposed by Pavio (1990) and Baddeley (1998) (see also Mayer, 2001).

The third key multimedia effect is the *redundancy effect* but again the literature is not clear cut. Mayer and Moreno (2003, p. 49) define the redundancy effect as the finding that “students understand a multimedia presentation better when words are presented as narration rather than as narration and on-screen text”. Thus, redundancy of information has a *negative* effect on multimedia learning. Other researchers who define redundancy as a negative effect include Sweller (2005), Kalyuga (2000) and Kalyuga, Chandler and Sweller (2004). However, a contrary finding is reported by Moreno and Mayer (2002, p. 161) who define the ‘verbal redundancy effect’ as “Words presented in both the visual and auditory modalities enhance learning as compared to words presented in only one modality”. On this definition the redundancy effect is seen as having a *positive* influence on learning.

Of particular relevance to the present research is the negative redundancy effect which has been explained in terms of cognitive load theory (Kalyuga et al., 2004; Pass, Renkl & Sweller, 2003; Pass, Tuovinen, Tabbers & Van Gervan, 2003). On this explanation, the limited capacity of the two input channels will be overloaded when the same verbal information in textual and auditory form has to be processed using only one channel (viz., verbal/auditory) (Kalyuga, 2000; Kalyuga et al., 2004). Such overloading can be avoided if the multimedia presentation involves audio narration and on-screen animation, because the two sources of information can be processed by separate channels (Kalyuga et al., 1999). A recent study by Pociask and Morrison (2008) showed that the cognitive load of multimedia instructional materials can be effectively minimised by using instructional and message design strategies.

Mayer and Moreno (2003) argue that overload may be reduced by approaches such as: 1) using the modality principle, whereby audio narration is presented instead of on-screen text to reduce the visual channel load (Clark & Mayer, 2003); 2) applying the redundancy principle (Mayer, Heiser & Lonn, 2001) of avoiding the use of identical written and spoken words to accompany animation; 3) allowing the user to break the presentation up into segments to avoid memory overload; 4) giving guidance in selecting and organising information through cues such as an emphasis on important words in speech; and 5) presenting audio and visual elements simultaneously to reduce memory loading. The combination of written and auditory information, therefore, violates both the modality and redundancy principles, and will generally result in overload which impairs learning. However, in cases where students have difficulty understanding spoken words or when the pacing of the material is not fast (Clark & Mayer, 2003), simultaneous audio and visual information may be experienced as non-redundant and overload may be avoided.

Use of simultaneous text in online multimedia lectures

The use of simultaneous screen text in lectures has received only limited research attention. A recent study appears to contradict the modality and redundancy principles, with its finding of no significant difference in student performance using lecture slides with audio compared to lecture slides with transcription (Day, Foley & Catrambone, 2006). Their findings also indicate that adding audio or transcription to lecture-slide-only presentations makes no significant improvement in learning efficacy. However, a combination of video, audio and lecture slides proved significantly better than all other approaches. The authors suggest that cognitive load theory and the cognitive theory of learning are of only limited application to lengthy presentations in lecture style (Day et al., 2006). In a separate study Fang, Xu, Brzezinski and Chan (2006) also challenge the redundancy principle. In this study short audio narrations of relevant information were found not to interfere with visual processing of on-screen textual information during web browsing. The researchers see this result as having implications for the use of simultaneous auditory and visual information on small screen, mobile devices.

Some research on the use of simultaneous on-screen text in live lectures has focused on student reaction and satisfaction measures rather than upon online learning efficacy. One study examined student reaction to simultaneous on-screen text presented during live lectures in three university courses (Hede, 2002b). A key problem with the use of speech recognition software during live lectures is that inaccuracies in the on-screen text cannot be edited out, and thus can cause distraction to students. The results of this study indicated that the majority of students found text inaccuracies to be distracting (range of 82% to 94% across the three courses) while only a small proportion reported the on-screen text as being helpful (range of 11% to 18%) and overall no more than one in four considered that the technology had a positive effect on their learning from the lectures (range of 12% to 25%) (Hede, 2002b).

These findings contrast with those of a study using the same speech recognition technology (*viz.*, *Via Scribe*) which found that 94% of students reported that the on-screen text improved their understanding of the live lecture (Leitch & MacMillan, 2001). One key difference between the two applications is that the lectures in the former study (Hede, 2002b) used both projected computer slides and simultaneous on-screen text, whereas only on-screen text was used in the latter. This use of a single rather than dual visual input in addition to the audio lecture may have decreased the likelihood of verbal/auditory channel overloading in the latter study. However, this difference is insufficient to account for the large discrepancy in results between the two studies (*viz.*, 18% versus 94% finding simultaneous on-screen text helpful) and further research is needed to determine how simultaneous on-screen text impacts on learning from live lectures for students without a hearing disability (Hede, 2002b).

The problem of on-screen text inaccuracies causing distraction can be overcome by using stored online applications which enable editing of the text file that accompanies a video lecture or that is provided as a stand alone learning resource. Also, editing can be used to modify the scrolling of the simultaneous text on the screen. The *Via Scribe* software automatically inserts a line break whenever the lecturer pauses for longer than a pre-set time. Because most lecturers have both semantic and non-semantic pauses in their speech, this software produces numerous non-semantic line breaks that may make it more difficult for a learner to comprehend the text as it scrolls

continuously on the screen (see Box C in Figure 1). On the other hand, if cognitive load theory applies, non-semantic line breaks may reduce the redundancy of the simultaneous auditory and textual information. To test these possibilities, the present study compared the impact on learning of unedited line breaks in simultaneous text with a condition in which non-semantic line breaks had been removed wherever possible.

Rationale and hypotheses

The rationale for this study was to investigate the efficacy of simultaneous text within an online multimedia lecture context. The research tests a prediction of cognitive load theory, namely, that text processed by a student via the visual modality will disrupt their learning if the same information is simultaneously processed via the auditory modality as both rely on the verbal/auditory channel (Clark & Mayer, 2003; Kalyuga, 2000; Kalyuga et al., 2004). However, cognitive load theory would propose that students who have difficulty with the English language may avoid overload in this situation because for them there is not full redundancy (i.e., they do not fully comprehend the verbal information they hear and see at the same time). Particularly if they find it difficult to understand spoken English, the presence of the simultaneous text on screen may assist their comprehension rather than disrupting it, as would occur with fully redundant information from the audio and video sources. Further, because simultaneous on-screen text with semantically edited line breaks will be more redundant than text with unedited breaks, cognitive load theory would predict less learning disruption in the latter case. Although in both conditions there will be two sources of verbal information that have to be processed by a single channel, the less semantic overlap with unedited text should result in a lower cognitive load.

The specific hypotheses (expressed in the null form) tested in this study were:

- H₁ There is no difference in learning when simultaneous on-screen text is presented together with an audio lecture in an online multimedia presentation as compared with a no screen text condition;
- H₂ Students' proficiency in the English language does not affect their level of learning under the two conditions, namely, with and without on-screen text; and
- H₃ There is no difference in the learning impact of simultaneous text and audio between semantically edited versus unedited line breaks in the on-screen text.

Methodology

Considering the hypotheses under investigation and the need to control and manipulate different conditions, an experimental approach was selected as the most appropriate design. The many contingent factors known to influence multimedia effectiveness were controlled by maintaining all conditions constant except for two text treatments and also by randomly assigning learners to three experimental groups (viz., no text, unedited text, and edited text).

Students enrolled in the various postgraduate business courses offered by a regional university were invited to undertake an optional online instruction in scholarly referencing via a multimedia presentation of approximately 30 minutes. The presentation adopted the format of a multimedia computer screen with three instructional boxes (see Figure 1), namely:

- Box A – Static image of the lecturer providing the audio instruction;
- Box B – *PowerPoint* slides presented sequentially;
- Box C – Simultaneous text of the words being spoken by the lecturer (using *Via Scribe* software, which automatically produces a transcription of the spoken words).

In addition to the visual components outlined above, all groups could hear the lecture presentation as an accompanying audio file. The presenter was carefully chosen to ensure that the lecture was presented in a clear, well-paced, conventional Australian accent voice.

Participants were randomly assigned to the three experimental groups with identical instructional materials except for variation in the text box (Box C) on the screen. The variations of Box C were as follows:

- Group 1 – Text with unedited line breaks (line breaks determined by pauses in lecturer's speech as in standard *Via Scribe* text files);
- Group 2 – Text with edited line breaks (line breaks edited to give semantically grouped text where possible);
- Group 3 – No simultaneous text (text box blank on the computer screen).

Before viewing the multimedia instructional presentation, participants were asked to take a pre-test on their knowledge of scholarly referencing. After undertaking the instruction, they were asked to complete a post-test of their knowledge within 30 minutes of viewing the presentation. Participants were also asked to complete a brief questionnaire about their learning experience.

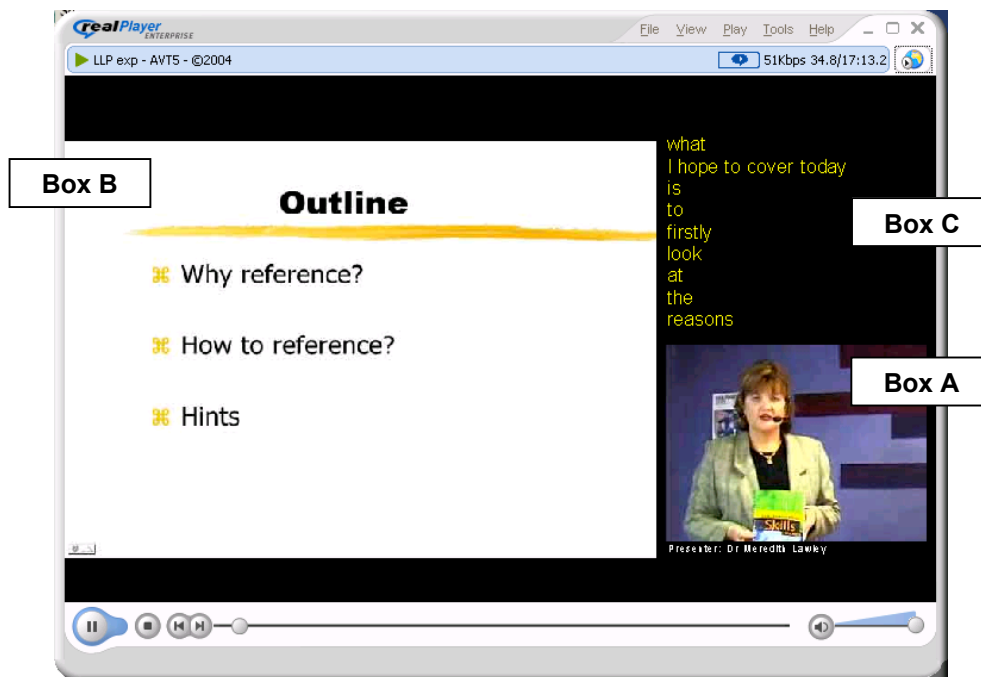


Figure 1: Screen layout showing the three types of visual instructional material (A: image of the lecturer, B: computer slides, C: simultaneous text of the audio lecture)

Operationally, the study was administered within the *Blackboard* learning management system. Participants were given access on their 'welcome page' to one of three versions of an additional 'course' which was offered as an optional instruction in scholarly referencing to be undertaken within a six-week period. When they accessed the course online they were presented initially with an announcement providing information about the study and instructions regarding the pre- and post-test procedures. Participants were advised to take the knowledge tests relying solely on their memory and not on any notes or other reference material. It should be noted that although it was possible for participants to 'cheat' on the tests, it was assumed that any effect of such behaviour would be non-systematic across the three experimental conditions.

The tests and the follow up questionnaire were administered via the 'survey' facility within *Blackboard* which provides anonymity for participants. The instructional presentation was accessed via a simple screen button available when they accessed their online material.

Measures

The pre- and post-tests of knowledge about scholarly referencing consisted of one set of 20 multiple choice questions with identical questions across the three experimental groups. The multimedia presentation as well as the test questions were based on the recommended resources provided to students for referencing.

In addition to the pre- and post-tests, participants were also asked to complete a brief questionnaire on their attitudes and preferences about the multimedia presentation as well as some demographic items. Specifically, questions were asked about: participants' preferred learning style (reading study material, listening to presentations with visual aids or a combination of reading and listening); self reported measures of competency in reading and understanding English; self ratings of the percentage of time during the presentation that participants looked at each part of the screen (lecturer image, *PowerPoint* slides, text box); and finally ratings of perceived usefulness of various aspects of the presentation such as hearing, seeing and reading.

Results

Profile of participants

Out of approximately 400 enrolled postgraduate students, a total of 60 accepted the invitation to participate in the research and 48 completed all three phases of the field experiment (the pre-test and post-test of knowledge about scholarly referencing, plus the questionnaire on aspects of their learning). There were 17 participants in both Group 1 and Group 2 and 14 participants in Group 3 based on random allocation (see Table 1). The present group sizes (*viz.*, 14-17) compare well with related studies employing an experimental design, namely: 10 per group in the study by Tindall-Ford et al. (1997); 8-12 in Kalyuga et al. (1999); 16-22 in Mayer et al. (2001) and 17-19 in Moreno and Mayer (2002). A follow up of the 12 students who commenced but did not complete the present study indicated that almost all dropped out because of Internet connection problems unrelated to their treatment group (with the remainder for unstated reasons).

The profile of the participants is here reported in two sections. The first section addresses general issues such as the study location, preferred learning style, English

language competency and overall usefulness of the instructional presentation (Table 1). It should be noted that while the number of questionnaire responses varies due to missing data, all 48 participants completed both the pre- and post-tests of their knowledge. The second section focuses on the time spent by participants in each group on the various visual and auditory components of the presentation (Table 2).

Table 1: Profile of participants

Characteristic		Total (n = 48)	Group 1 Unedited text (n = 17)	Group 2 Edited text (n = 17)	Group 3 No text (n = 14)
Study location	Direct online	24	11	9	4
	Offshore (Fiji)	1	1	0	0
	Offshore (Malaysia)	14	4	2	8
Preferred learning style	Reading	2	1	0	1
	Listening	3	0	1	2
	Combination	33	14	10	9
English first language	Yes	26	9	10	7
	No	13	7	1	5
Difficulty understanding English	None	34	14	11	9
	Slight	5	2	0	3
Difficulty reading English	None	31	13	10	8
	Slight	7	2	0	4
	Moderate	1	1	1	0
Overall usefulness* (means)		3.05	2.93	2.93	3.33
Specific usefulness* (means)	Hearing audio	3.15	3.20	3.00	3.25
	Seeing PPT slides	3.08	3.33	2.92	2.92
	Reading text	2.78	2.87	2.67	N/A
	Seeing presenter	2.13	2.13	1.77	2.50

Notes: Differences in numbers between profile subtotals and totals reflect missing data.

*0 = not at all useful, 1 = slightly useful, 2 = moderately useful, 3 = quite useful, 4 = very useful

Of those participants who responded to the section about their profile, approximately 40% were studying at offshore locations (Fiji or Malaysia) (see Table 1); over 86% of participants used a combination of learning styles; 66% had English as their first language; 87% reported no difficulty understanding English; and 79% reported no difficulty reading English. Considering each group in relation to these criteria, Group 1 had a fairly even balance of study locations represented, while Group 2 had more students studying directly (typically Australian students) and Group 3 had more Malaysia based students. In all three groups the 'combination' learning style (reading and listening) was predominant.

The distribution of participants with English as a first language reflected the pattern of study location. Thus, Group 1 comprised almost equal numbers of those who spoke English as a first language and those who did not, whereas Group 2 comprised mainly those with English as a first language and Group 3 mostly had members with non-English as their first language (see Table 1). Regardless of whether English was a first language or not, all three experimental groups had the majority of their members reporting no difficulties in reading or understanding English.

Overall in terms of usefulness, participants found the multimedia presentation to be quite useful (mean = 3.05), with Group 3 (no text) reporting the highest level of

usefulness (mean = 3.33). This may reflect the large proportion of offshore, non-native English speakers in this group compared to the other two groups.

In terms of the usefulness of specific components of the multimedia presentation (i.e. hearing the presentation, seeing the *PowerPoint* slides, reading the simultaneous text and seeing the presenter), all were considered useful but hearing was rated the most useful by two of the three groups, Groups 2 and 3 (see Table 1). The order was generally consistent across groups; that is, hearing the presenter and seeing the *PowerPoint* slides were consistently the most useful, seeing the presenter was consistently the least useful and reading the text was the second least useful, with the one exception being Group 1 where seeing the *PowerPoint* ranked higher than hearing the audio.

In summary, while participants were randomly allocated to the groups, the profile of group members was different on the key variables of English as a first language and study location. The impact of this needs to be considered when interpreting the results.

In addition to asking participants to rate the usefulness of the various components of the multimedia presentation, the questionnaire also asked them to estimate the proportion of time spent on each of the components of the presentation. Results from Table 2 show that the patterns of time spent on the various components of the multimedia match the ratings of usefulness in Table 1. Thus, the majority of participants (56%) spent 75% or more of their time listening to the presentation compared to only 38% who spent 75% or more of their time looking at the *PowerPoint* slides. The comparative figures for reading and looking at presenter were 29% and 14% respectively.

Table 2: Reported time spent on various activities during the online lecture presentation

Activity	Time spent	Total (n = 48)	Group 1 Unedited text (n = 17)	Group 2 Edited text (n = 17)	Group 3 No text (n = 14)
Listening to audio	0%	3	1	1	1
	25%	0	0	0	0
	50%	7	1	0	6
	75%	15	9	3	3
	100%	12	4	6	2
Looking at <i>PowerPoint</i>	0%	4	1	1	2
	25%	5	3	2	0
	50%	10	4	2	4
	75%	10	4	1	5
	100%	8	3	4	1
Reading text	0%	2	1	1	N/A
	25%	3	2	1	
	50%	6	5	1	
	75%	10	4	6	
	100%	4	3	1	
Looking at presenter	0%	10	5	1	2
	25%	12	4	2	5
	50%	8	4	2	3
	75%	5	1	1	1
	100%	2	1	4	1

Note: Differences between profile subtotals and totals reflect missing data

In summary, all groups rated the presentation as useful, with the auditory component generally most useful. Of the visual components seeing the *PowerPoint* slides was rated more useful than reading the text (either edited or unedited) and seeing the presenter was the least useful. On the basis of this profile the three specific research questions will be addressed next.

Hypothesis testing

The main analysis compared the change in post-test performance over pre-test performance across the three experimental groups. The means are shown in Table 3 (i.e., the mean differences in scores out of 20 in the two tests of knowledge about referencing). While the mean for the unedited text condition (Group 1) appears higher than in the other conditions (2.82 versus 2.24 and 2.0), an analysis of variance indicated this is not statistically significant ($F = 0.52$, $df = 2$, NS). The means of the groups are very close, and so the variations within each group would have to be very small to be significant. However, the coefficient of variation is 82%, 119% and 96% for Groups 1 to 3, respectively.

Table 3: Improvement in learning (post-test minus pre-test scores) across treatment conditions

Group	Mean	n	Std. deviation
Group 1 – Unedited text	2.82	17	2.32
Group 2 – Edited text	2.24	17	2.66
Group 3 – No text	2.00	14	1.92
Total	2.38	48	2.32

Specifically, our first hypothesis suggested that there would be a difference in learning when simultaneous text is presented on screen in addition to an audio lecture in an online multimedia package. In terms of our experimental groups, two groups (1 and 2) had simultaneous text and one group (3) had no text. This hypothesis predicted a difference between Group 1 and Group 3 and as well as difference between Group 2 and Group 3. Statistically, as highlighted above, there were no significant differences between these groups and hence this hypothesis is not supported.

It was not possible to conduct a statistical test of the second hypothesis because the three experimental groups did not include sufficient numbers of students who had difficulty with English language.

The third hypothesis predicted a difference in the learning impact of simultaneous text and audio between semantically edited versus unedited line breaks in the text. In terms of the experimental groups, this predicted a significant difference between Group 1 and Group 2. Again, no significant difference was identified between these two groups and hence this hypothesis is not supported.

The learning efficacy score (post-test minus pre-test) is likely to be affected by the pre-test score, since as participants get nearer to the maximum possible score within the pre-test the amount by which they may improve on this (represented by post-test minus pre-test) will diminish. To measure this effect, the relationship between pre-test score and the learning score (post-test minus pre-test) was tested using a one-way analysis of covariance. A significant relationship between pre-test and learning score was discovered ($F_{1,44} = 20.69$, $p < .01$). This relationship is significantly negative ($B = -0.47$, $t = -4.55$, $p < .01$) meaning that for lower pre-test values the learning score is high,

and that as pre-test score increases the learning score decreases. This confirms that, as expected, the nearer a student gets to the maximum pre-test score the less they are able to improve upon it in the post-test.

Given the relationship between pre-test score and learning efficacy score, it is necessary to determine whether, when the pre-test score is taken into account, there is a significant relationship between experimental group and learning score. The ANCOVA showed no significant differences in learning efficacy score among the three groups ($F_{2,44} = 0.71$, NS) after accounting for the significant effect of pre-test scores ($F_{1,44} = 20.69$, $p < 0.01$). Students who scored more highly in their pre-test showed less improvement, as would be expected. The incorporation of pre-test as a covariate also accounts for the temporal design of the experiment. Further, the observed power for group within the between-subjects effects of the ANCOVA is relatively low (0.16); this may be caused by small sample size and/or by large variations in scores within groups.

A further potential effect upon the group means is the language background of the participants. This was measured by using an independent samples t test to determine whether learning scores are significantly different between participants whose first language is English and those whose first language is not English. The test showed no significant difference ($t_{.05, 34} > 0.99$, $p > .05$) between the English and non-English participants, suggesting that the effect of language background on the groups was not strong.

Discussion

In terms of our first hypothesis predicting differences between those receiving simultaneous on-screen text and audio as compared to those receiving audio only, our results contradict the modality and redundancy principles (Clark & Mayer, 2003), which predict that the group viewing no on-screen text (Group 3) would perform better than the others. Indeed, while there was no significant difference in terms of raw averages, Group 3 actually showed the lowest average improvement of the three groups. This may be partially explained by the use of a static image of the lecturer rather than animated video, which would reduce the risk of overloading the visual channel of the learners. Further, the redundancy principle is less appropriate when participants struggle to understand spoken words or the pacing is not fast (Clark & Mayer, 2003); both of these may apply in this case, given that Group 3 had the highest proportion of non-native English speakers. In addition, the content contained within the presentations may also have been of sufficiently low information density to reduce the risk of overloading working memory (Kalyuga, 2000; Kalyuga et al., 1999) or of exceeding the limited capacity of a processing channel (Clark & Mayer, 2003).

The results do, however, agree with those of Day et al. (2006) who found no significant difference between student performance using lecture slides alone compared to the combination of lecture slides and a transcription. Their suggestion that cognitive load theory and the cognitive theory of learning have limited application to long, lecture style presentations is also likely to apply to our scenario.

In terms of our second hypothesis, cognitive load theory suggested non-semantic line breaks may reduce the redundancy of the simultaneous auditory and textual information (Kalyuga, 2000; Kalyuga et al., 1999). Our results did not support this and

indeed raw averages suggest unedited text (Group 1) performed better than edited text (Group 2). One possible explanation for this may be that the unedited on-screen text matched the presenter's pauses (i.e., in the audio input) and hence overload was reduced. Whereas in Group 2 the edited text would not have matched the pauses in the audio as closely, thus increasing overload.

Further insight into reasons for lack of support of our hypotheses may be found in the general profile results. In particular these results showed that the two groups with simultaneous text and audio ranked hearing audio and seeing *PowerPoint* slides as more useful than seeing the text. In addition these two groups spent more time listening and looking at *PowerPoint* slides than they did reading text. Given the consistent lower usefulness and time spent on text, it is then not surprising that differences between edited and unedited text did not occur.

Conclusions

This experimental research has added further support to the existing literature showing that lecture transcription (text) appears to be of relatively low importance to participants, specifically online students rather than those studying face to face. Results suggest that resources are better spent providing audio accompanied by *PowerPoint* slides, with on-screen text providing limited additional value. Given this finding, any differences between edited and unedited text are irrelevant to student learning outcomes. However, if voice recognition technology is being used to generate on-screen text, our findings suggest resources should not be wasted editing text for semantic line breaks.

While this study has provided valuable insights into the learning efficacy of simultaneous screen text and audio in online lectures it has some limitations. Firstly, the experimental groups were relatively small though not unusually so in comparison with other similar studies. However, group size may have impacted on the statistical power of the analysis and hence the ability of the study to detect differences between experimental conditions. Secondly, although participants were randomly assigned to the three experimental groups an analysis of group demographics indicated a larger proportion of non-native English speakers in one group which may also have impacted our results even though this effect was statistically analysed and found to be non-significant. Finally, the task chosen to measure learning efficacy (viz., referencing skills) may have been of too low complexity and may have resulted in relatively small improvements in learning as participants had a fairly high level of knowledge prior to instruction. This shortcoming has, however, been addressed by taking pre-test scores into account; this adjustment does not affect the results. Future research, therefore, should aim to address these limitations in order to resolve the important issues of learning efficacy raised in this study.

References

- Attaran, M. (2000). Voice recognition software programs: Are they right for you? *Information Management & Computer Security*, 8(1), 42-44.
- Ayers, P. & Sweller, J. (2005). The split-attention effect in multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press, New York.
- Baddeley, A. (1998). *Human memory*. Allyn & Bacon, Boston, MA.

- Buckler, G. (2001). Recognizing voice recognition. *Computer Dealer News*, 17(22), 17.
- Byrnes, R. & Ellis, A. (2006). The prevalence and characteristics of online assessment in Australian universities. *Australasian Journal of Educational Technology*, 22(1), 104-125. <http://www.ascilite.org.au/ajet/ajet22/byrnes.html>
- Christensen, J. & Hughes, B. (2007). Voice-enabled IT transformation: The new voice technologies. *IBM Systems Journal*, 46(4), 763-775.
- Clark, R. E. & Feldon, D. F. (2005). Five common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press, New York.
- Clark, R. E. & Mayer, R. (2003). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Wiley, San Francisco, CA.
- Day, J., Foley, J. & Catrambone, R. (2006). *Investigating multimedia learning with web lectures*. GVU Technical Report GIT-GVU-06-25. Georgia Institute of Technology, GA. <http://smartech.gatech.edu/handle/1853/13141>
- Dillon, A. & Gabbard, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control and style. *Review of Educational Research*, 68(3), 322-349.
- Fang, X., Xu, S., Brzezinski, J. & Chan, S. S. (2006). A study of the feasibility and effectiveness of dual-modal information processing. *International Journal of Human-Computer Interaction*, 20(1), 3-17.
- Hede, A. (2002a). An integrated model of multimedia effects on learning. *Journal of Educational Multimedia and Hypermedia*, 11(2), 177-191.
- Hede, A. (2002b). Student reaction to speech recognition technology in lectures. In S. McNamara & E. Stacey (Eds), *Proceedings of the Australian Society for Educational Technology (ASET) International Conference*, 7-10 July, ASET, Melbourne. <http://www.ascilite.org.au/aset-archives/confs/2002/hede-a.html>
- Hede, T. & Hede, A. (2002). Multimedia effects on learning: Design implications of an integrated model. In S. McNamara & E. Stacey (Eds), *Proceedings of the Australian Society for Educational Technology (ASET) International Conference*, 7-10 July, ASET, Melbourne. <http://www.ascilite.org.au/aset-archives/confs/2002/hede-t.html>
- Hsiu-Ting, H. (2009). Learners' perceived value of video as mediation in foreign language learning. *Journal of Educational Multimedia and Hypermedia*, 18(2), 171-190.
- Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australian Journal of Educational Technology*, 16(2), 161-172. <http://www.ascilite.org.au/ajet/ajet16/kalyuga.html>
- Kalyuga, S., Ayers, P., Chandler, P. & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23-31.
- Kalyuga, S., Chandler, P. & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors*, 46(3), 1-15.
- Kalyuga, S., Chandler, P. & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351-371.

- Kalyuga, S. & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, 96(3), 558-568.
- Kim, J. H. & Lee, S. B. (2007). Speech recognition using multilayer recurrent neural prediction models and HMM. *Control and Intelligent Systems*, 35(1), 9-14.
- Lang, A. (1995). Defining audio/video redundancy from a limited-capacity information processing perspective. *Communication Research*, 22(1), 86-115.
- Leitch, D. & MacMillan, T. (2001). *Improving access for persons with disabilities in higher education using speech recognition technology: Year II progress report*. Unpublished Report, Liberated Learning Project, Saint Mary's University, Halifax, Canada.
http://www.liberatedlearning.com/resources/docs/RC_2001_Year_II_Research_Report.doc
- Liao, Y. (1998). Effects of hypermedia versus traditional instruction on students' achievement: A meta-analysis. *Journal of Research on Computing in Education*, 30(4), 341-360.
- Liao, Y. (1999). Effects of hypermedia on students' achievement: A meta-analysis. *Journal of Educational Multimedia and Hypermedia*, 8(3), 255-277.
- Low, R. & Sweller, J. (2005). The modality principle in multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press, New York.
- Marshall, P. (2002). Voice recognition: Sound technology. *Federal Computer Week*, 16(1), 32.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press, New York.
- Mayer, R. E., Heiser, J. & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93(1), 187-198.
- Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 42-52.
- Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312-320.
- Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86(3), 389-401.
- Mayer, R. E. & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83(4), 484-490.
- McNeil, B. J. & Nelson, K. R. (1991). Meta-analysis of interactive video instruction: A 10-year review of achievement effects. *Journal of Computer-Based Instruction*, 18(1), 1-6.
- Moreno, R. & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, 94(1), 156-163.
- Moreno, R. & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(3), 358-368.
- Moreno, R., Mayer, R. E., Spire, H. A. & Lester, J.C. (2001). The case for social agency in computer-based multimedia learning: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19, 177-214.

- Mousavi, S. Y., Low, R. & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87(2), 319-334.
- Pass, F., Renkl, A. & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Cognitive Psychologist*, 38(1), 1-4.
- Pass, F., Tuovinen, J. E., Tabbers, H. & Van Gervan, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Cognitive Psychologist*, 38(1), 63-71.
- Pavio, A. (1990). *Mental representations: A dual-coding approach*. Oxford University Press, New York.
- Penney, C. G. (1989). Modality effects and the structure of short-term memory. *Memory and Cognition*, 17, 398-422.
- Pociask, F. D. & Morrison, G. R. (2008). Controlling split attention and redundancy in physical therapy instruction. *Educational Technology Research and Development*, 56(4) 379-399.
- Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press, New York.
- Sweller, J. (1999). *Instructional design in technical areas*. Australian Council for Educational Research, Melbourne.
- Tabbers, H. K., Martens, R. L. & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, 74, 71-81.
- Tindall-Ford, S., Chandler, P. & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3(4), 257-267.
- Yong, Z. (2007). Speech technology and its potential for special education. *Journal of Special Education Technology*, 22(3), 35-41.

<p>Dr Justin Debuse, Lecturer, Dr Andrew Hede, Professor, and Dr Meredith Lawley, Associate Professor, Faculty of Business, University of the Sunshine Coast, Maroochydore DC, Queensland 4558, Australia. Email: jdebuse@usc.edu.au, ahede@usc.edu.au, mlawley1@usc.edu.au Web: http://www.usc.edu.au/</p>
--