Effects of Modality Principle in Tutorial Video Streaming

Ali H. AbuSaada

School of Educational Studies, Universiti Sains Malaysia
Email: ali.seaada@hotmail.com

Lily Pei Lin Lee

School of Educational Studies, Universiti Sains Malaysia
Email: lily4ever85@gmail.com

Soon Fook Fong

School of Educational Studies, Universiti Sains Malaysia
Email: sffong05@gmail.com

Abstract

This study aimed to investigate the effects of Modality Principle in video streaming of ICDL course for trainee-teachers. Two modes of video streaming were used, namely video and narration (VN) mode, and video and text (VT) mode of ICDL course. The sample consisted of 202 trainee-teachers in the National Institute for Training (NIT) in Gaza. The results showed that the learners using the VN mode attained significant better achievement than the learners using the VT mode. The learners using the VN mode are learn better and more deeply than the VT mode. The presented information in both visual and auditory modes imposes lesser cognitive demands on learners' without overloading working memory. Thus, more working memory capacity remains available for information processing and improve the learning outcomes.

Keywords: Modality, Video streaming, ICDL.

1. Introduction

Recent researches in multimedia focus on study of learner-centred method based on theories of human learning rather than the study of technologies used to transmit the presentation (Lowe & Schnotz, 2008). The multimedia presentation designers have to take into account how people learn and process the information (Mayer, 2001). With this regards, the current researchers try to comprehend how the use of multimedia in education can assist cognition (McNeill, Doolittle & Hicks, 2009).

Mayer & Sims (1994) mentioned that the use of multimedia presentations improves learning outcomes. This achievement can be explicated through dual-coding theory which suggests that humans have two separate and interconnected channels systems (i.e., verbal and non-verbal) for representing information (Hodes, 1994; Paivio, 1990). Kalyuga, Chandler and Sweller (1999)

pointed out that dual-mode multimedia presentations increase the capacity of effective working memory rather than minimizing the unusual cognitive load.

Recent researches shed the light on how the information is represented to the learner in dual encoding system, such as presenting images along with on-screen text (Mayer, 2001).

Other researches in this area have pointed out that if the multimedia is not designed in an appropriate manner, it can impose undue strain on the learner's information processing system (Kalyuga, Chandler, & Sweller, 1999; Chandler & Sweller, 1991).

According to Mayer (2001), multimedia learning is learning from both words and pictures. The words can be printed (e.g., on-screen text) or spoken (e.g., narration). The pictures can be static (e.g., graphs, illustrations, charts, or photos) or dynamic (e.g., video, animation or interactive illustrations) (Mayer & Moreno, 2003).

The rationale for cognitive theory of multimedia is that learners learn more effectively from words and graphics rather than words alone. The words comprises printed words or narration; graphics comprises static graphics or dynamic graphics such as animation or video (Mayer, 2005).

When the multimedia presented, learners' information processing system processes both words and pictures together rather than words alone. The design of multimedia environments should be compatible with how learners learn (Mayer, 2001). Cognitive theory of multimedia learning in presenting learners' information processing system is stated by Mayer as displayed in Figure 1.

Figure 1 represents the theory as a series of boxes that is arranged into two rows and five columns, with arrows connecting them. The columns represent memory stores including sensor memory, working memory and long-term memory. The two rows represent the two information-processing channels, with the auditory/verbal channel on top and the visual/pictorial channel at the bottom.

The presented multimedia enters learner's sensory memory through their eyes and ears. Learners select the relevant visual and auditory information from multimedia presentation. When the information is aurally presented, it enters the auditory channel via the ears; on the other hand, when the information is presented visually, it enters the auditory visually via the eyes and is then processed in the visual channel. The sensor memory allows for the spoken words to be held as exact auditory memory for a brief time period in the auditory sensor memory, while pictures and text to be held as exact visual images for brief time period in the visual sensor memory. The arrow from pictures to eyes represents pictures being registered at the eyes; the arrow from words to ears represents spoken words being registered at the ears.

Long-term memory is the learner's storehouse of knowledge. Unlike working memory, long-term memory can hold large amount of knowledge over a long time period. In order for

learners to recall and actively use the information being stored in long-term memory, the information must be transferred into working memory.

Active cognitive processing in multimedia learning requires five cognitive processes that are represented by labelled arrows: selecting words, selecting images, organizing words, organizing images, and integrating (Mayer & Moreno, 2003). Selection is a mental process where the learner pays attention to relevant information verbally or non-verbally presented to him. Mayer (1984) defines selection as choosing the right information and adding it to the working memory.

The central work of multimedia learning occurs in the working memory. According to Mayer (2001), "the working memory is used for temporarily holding and manipulating knowledge in active consciousness." The learner will process the visual information in pictorial model while auditory information in verbal model in working memory, which is "the centre of cognition" (Clark & Mayer, 2003). In the working memory, the incoming information is stored as unprocessed information. Based on the visual and auditory sensory modalities, the learner organizes verbal and pictorial information into coherent mental representations. The selected words are organized into verbal model, while the selected pictures into pictorial model. Then, a meaningful connection is constructed between the verbal and its corresponding pictorial representations in working memory. Finally, learner makes the integration between the new information and relevant prior knowledge from long-term memory. This newly integrated knowledge is continually stored in long-term memory, leading to meaningful learning (Dacosta, 2008; Lan, 2010). Hence, the meaningful learning occurs when the learner engages in all cognitive processes. In other words, learner must carry out each of these cognitive processes: selecting relevant words and images, organizing them into coherent verbal and visual representations, and integrating corresponding verbal and visual representations (Mayer & Moreno, 2002b). The construction of connection between verbal and visual representations is more likely for situations in which the learner can simultaneously hold corresponding visual and verbal representations in memory at the same time. Thus, instructional materials should be designed to maximize the chances for such crucial cognitive processing.

2. Modality and learning

The modality principle emanates from limited capacity in working memory assumption of the cognitive theory of multimedia learning (Mayer, 2001). The modality effect happens when the presented information in dual combined mode (i.e., auditory and visual) is more effective than the same information presented in a single mode (Zheng, 2009). In other word, when the information, such as words, is presented in an auditory mode rather than graphic mode, at the same time this auditory information is integrated with graphics information such as video or animation (Tindall-Ford, Chandler, & Sweller, 1997).

According to cognitive theory, it is better to use auditory mode rather than printed mode when presenting words along with videos in a multimedia presentation (McNeill, Doolittle, & Hicks, 2009; Mayer, 2005b; Clark & Mayer, 2003;). In the same context, Mousavi, Low and Sweller

(1995) pointed out that the concurrent graphics geometry diagram and narrated explanation mode enhances the learning as compared with the on-screen text presentation mode. Thus, the designer of multimedia presentation courses should avoid using all words in printed mode rather than spoken mode in order to reduce the chances of overloading the learners' visual/pictorial channel.

The rationale for this avoidance is that the learners may undergo an overload of their pictorial/visual channel when they ought to simultaneously process the printed and words graphics in the memory system. When the printed words and graphics (i.e., video) are presented simultaneously at a rapid pace and the learners need to follow the printed words, the result is that the learners cannot fully attend to the video or graphics (Clark & Mayer, 2003).

When the words and graphics are presented concurrently, Mayer (2005b) suggested the use of spoken words rather than printed words as a way of reducing the demands on visual processing.

Clark and Mayer (2003) recognized that in some cases it may not be realistic to apply the modality principle due to some technical disabilities in the learning environment related to the computer.

Rationale Psychological for modality principle is that the information presented simultaneously both on-screen text and graphics can conflict with the way the learner's mind works (Clark & Mayer, 2003), and this might cause cognitive overload (Mayer & Moreno, 2002). According to the cognitive theory, learners have two separate information processing channels: visual and verbal channels (Mayer, 2005b).

When the learner receives information simultaneously on-screen text and graphics, both will be processed through the eyes in the visual/pictorial channel. Given that the capacity of each channel is limited, the concurrent processing of on-screen text and graphics burdens the visual channel and as a result, it becomes overload. In addition, the learner may not probably have much cognitive capacity residual to create connections between on-screen text and graphics, as shown in Figure 2 (Clark & Mayer, 2003; Mayer, & Moreno, 2002).

On the other hand, when the information is presented to the learner via graphics and narration, the load on visual channel will be minimized. Thus, no channels are overloaded (refer to Figure 3). The learner uses auditory channel to process the spoken words while visual channel to process the graphics (Clark & Mayer, 2003), and more probably learners are able to create connection between the identical on-screen text and graphics (Mayer & Moreno, 2002).

4. Research hypotheses

There is no significant difference in the post-test score (PTS) among learners using Video cum Narration (VN) mode and learners using Video cum Text (VT) modes.

5. Methodology

5.1 Participants

The participants of this study consisted of 202 school teachers selected from Gaza schools, from all subjects, and of both genders (male and female).

5.2 Materials

The current study focused on learners achievement of the ICDL course. This course is essential for increasing teachers' ability to use computer in teaching and learning process as well as in their life. This course contained seven modules, as stated by ECDL Foundation (2007): 1 – Concepts of Information and Communication Technology (ICT), 2 – Using Computer and Managing Files, 3 – Word Processing, 4 – Spreadsheets, 5 – Using Databases, 6 – Presentation, 7- Web Browsing and Communication.

5.3 Design

The study utilized a quasi-experimental method to measure the effects of the two different treatment modes (i.e., VN and VT) on learners' achievement in the ICDL course. The independent variable in this study was the modes of video presentation – Video cum Narration (VN) and Video cum Text (VT). The dependent variable is the learners' achievement, as measured by their post-test scores.

6. Result

The statistical results indicated that the learners using the VN mode attained significantly higher post-test scores than the learners using the VT mode.

A probability level of 0.05 was used to test statistical significance. An ANCOVA procedure was carried out to determine whether the difference between the two groups of learners is significant. The results showed that there was a significant difference found in the achievement of learners between the VN and VT groups, $F_{(1, 199)} = 35.802$, $p = 0.000$.

A follow-up post-hoc pairwise comparison via LSD procedure was conducted to further determine where the significant differences in the mean had resided.

Table 3 showed the summary of the post hoc pairwise comparisons result. It can be seen that the mean difference found in the achievement of learners between the VN group and the VT group was significant with p= 0.000. The VN group attained significantly higher post-test mean score (mean = 88.05) than that of the VT group (mean = 76.85).

7. Discussion and Conclusion

The current study results found that learners using VN mode attained better post-test scores compared to learners using VT mode. The results of the current study agreed with the results reported by many studies which have confirmed the effectiveness of the modality principle in learning. Hii (2012) found that the graphic novel and narrated mode had a significant positive effect on learning as compared to graphic novel and text mode; Aldalalah (2010) reported that pupils using audio and image gained significantly higher post-test scores than pupils using text and image. These results indicated that a modality principle was effective to improve learning. Mousavi, Low and Sweller (1995) pointed out that concurrent graphics and narration explanation mode could better enhance learning as compared to on-screen text presentation mode. Likewise, Aldalalah, Fong and Ababneh (2010) reported the positive effects of the modality principle on learning, that is, students understand more deeply using narration or audio with graphics than on-screen-text with graphics. In the same context, many other studies asserted the modality principle effect and they mentioned that the combination of narration and graphics presentation mode is more effective than written words and graphics presentation mode (Mayer, Dow & Mayer, 2003; Kalyuga, Chandler & Sweller, 1999; Mayer & Moreno, 1998; Jeung, Chandler & Sweller 1997; Mousavi et al., 1995).

This result can be explained as follows: The study found that learners using the VN mode attained significant better achievement than the learners using the VT mode. This result confirms the modality principle effect mentioned by Clark & Mayer (2008), that is, learners using the VN mode are likely to learn better and more deeply than learners using the VT mode. According to Mayer (2001), the information that is presented in both visual and auditory modes imposes lesser cognitive demands on learners' working memory. In other words, while processing visual and auditory information, both the channels will be utilized without overloading any one of them. More working memory capacity remains available for information processing and this may improve the learning outcomes.

According to Mayer (2001), "The working memory is used for temporarily holding and manipulating knowledge in active consciousness." Baddeley (1986) mentioned that there are two channels in the working memory, pictorial and verbal model. Learner processes visual information in the visual channel, while auditory information is processed in the verbal channel (Clark & Mayer, 2003). Based on the visual and auditory sensory modalities, the learner organizes verbal and pictorial information into coherent mental representations, where the selected narration words are organized into verbal model, and the selected pictures into pictorial model. Then, a meaningful connection between the verbal model and the pictorial model is established. Finally, the verbal and pictorial information are integrated with relevant prior knowledge from long-term memory.

According to cognitive theory of multimedia learning, learners have limited capacity to process information in the working memory at any one time (Clark & Mayer, 2003; Mayer, 2001). Thus, a balance occurs when the information is processed through the two channels and there is no

cognitive overload on both of the two channels. The video enters through the eyes and is then processed in the visual channel, while the narration enters through the ears and is then processed in the auditory channel. On the other hand, cognitive overload occurs when the video is presented synchronized with on-screen text in VT mode. As a result, the learners in the VT group are more likely to experience cognitive overload. This might adversely affect their information processing abilities and learning is thus inhibited.

One plausible explanation for the unsatisfactory achievement of the VT group is that learners may experience an overload in their visual channel when they simultaneously process the video and the on-screen text. When viewing the VT mode, both video and on-screen textual information enter through the learners' eyes at the same time, then being transferred and processed in the visual channel. Due to the limited working memory capacity, cognitive overload may occur at the visual channel. In contrast, the VN mode which combines video with narrated words instead of on-screen text allows the information to be processed separately in both visual and auditory channel. Thus, the amount of information being processed in the visual channel and auditory channel is not affected by cognitive overload.

Furthermore, while interacting with the VT mode, learners are required to split their attention between the two different sources of visual information – videos and on-screen text (Mayer, 2008). That is, they need to select relevant words from the on-screen text and then manually integrate them with the corresponding portion of the video to make sense of the information presented. In fact, this creates an extraneous cognitive load on learners' working memory which inhibits information processing and learning, as supported by Sweller's (1999) cognitive load theory and split-attention effect. Additionally, while reading the on-screen text, learners in the VT group may not be able to fully attend to the video or graphics (Clark & Mayer, 2003) and so fail to construct an accurate mental representation.

To present a multimedia presentation consisting of words and graphics, Mayer (2005b) suggested the use of spoken words rather than printed words to reduce the demands on visual processing. In this study, the significantly high achievement of the VN group implied that the use of narrated words and video decreases the consumption of the limited working capacity. When viewing the VN mode, the narrated words are registered at learners' ears and then being processed in their auditory channel, while the videos are registered at learners' eyes and then being processed in their visual channel. That is, unlike that of the VTN mode, the VN mode allows learner to utilize both visual and auditory channel without overloading any of them. Thus, more working memory capacities are thus devoted to make connection between the new pieces of information as well as to integrate them with prior knowledge. Learning is thus facilitated.

References

Aldalalah, O., Fong, S. F., & Ababneh, Z. (2010). Effects of modality principle on the learning of Arabic language among first grade pupils in Jordan. *Journal of Educational Sciences*.

Aldalalah, O. (2010). *Effect of modality and redundancy principles on the learning of music theory and attitude among primary pupils of different aptitudes in Jordan.* Unpublished Ph.D Thesis, Unversiti Sains Malaysia, Penang.

Baddeley, A. D. (1999). *Human memory*. Boston: Allyn & Bacon.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Clark, R. C., & Mayer, R. E. (2003). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. San Francisco: Jossey-Bass/Pfeiffer.

Dacosta, B. (2008). *The effect of cognitive aging on multimedia learning.* University of Central Florida, Florida.

ECDL_Foundation. (2007). European computer driving licence/ International computer driving licence syllabus version 5.0. *The European Computer Driving Licence Foundation Ltd*,

Hodes, L. (1994). Processing visual information: Implications of the dual code theory. *Journal of Instructional Psychology, 21*(1), 36-43.

Hii, C. S. (2012). *Effects of multimedia graphic novel on the performance and perceived motivation among students of different learning approaches.* USM.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

Lan, T. (2010). *The design, development and evaluation of educational games-based learning environment: its efficacy in learning the concept of water cycle at the primary school level.* Universiti Sains Malaysia.

Lowe, R., & Schnotz, W. (2008). Learning with animation: Research implications for design Cambridge: Cambridge University Press.

Mayer, R. (2005). *Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity.* New York: Cambridge University Press.

Mayer, R. (2005b). *The Cambridge Handbook of Multimedia learning*: Cambridge University Press.

Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43 – 52.

Mayer, R. E. (2001). *Multimedia learning*. United Kingdom: Cambridge University Press.

Mayer, R., & Sims, V. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389-401.

Moreno, R., & Mayer, R. (2002b). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*, 151 – 163.

Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319-334.

McNeill, A., Doolittle, P., & Hicks, D. (2009). The Effects of Training, Modality, and Redundancy on the Development of a Historical Inquiry Strategy in a Multimedia Learning Environment. *Journal of Interactive Online Learning, 8*, 255-269.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*(4), 257-287.

Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*, 329-343.

Zheng, R. (2009). *Cognitive Effects of Multimedia Learning*. New York: Information Science Reference (an imprint of IGI Global).

Table 1 Mean and Std. Deviation of post-test scores of (VN & VT)

Group	Number	Mean	Std. Deviation
VN	100	88.05	17.9
VT	102	76.85	16.7
Total	202	78.14	17.3

Table 2 ANCOVA of the Post-Test Scores of Learners Using Different Modes of Video (VN, VT)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	13490.459a	2	6745.230	39.472	.000
Intercept	521952.090	1	521952.090	3.054E3	.000
Treatment	6117.944	1	6117.944	35.802	.000
Pre	4892.576	1	4892.576	28.631	.000
Error	34006.061	199	170.885		
Total	1419061.000	202			
Corrected Total	47496.520	201			

a. R Squared = .284 (Adjusted R Squared = .277)

Table 3 Summary of the Post Hoc Pairwise Comparisons

(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference Lower Bound	Upper Bound
VN	VT	*11.200	1.872	.000	7.509	14.891
VT	VN	*-11.200	1.872	.000	-14.891	-7.509

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

HR_MARS_
Exploring Intellectual Capital

Figure 1

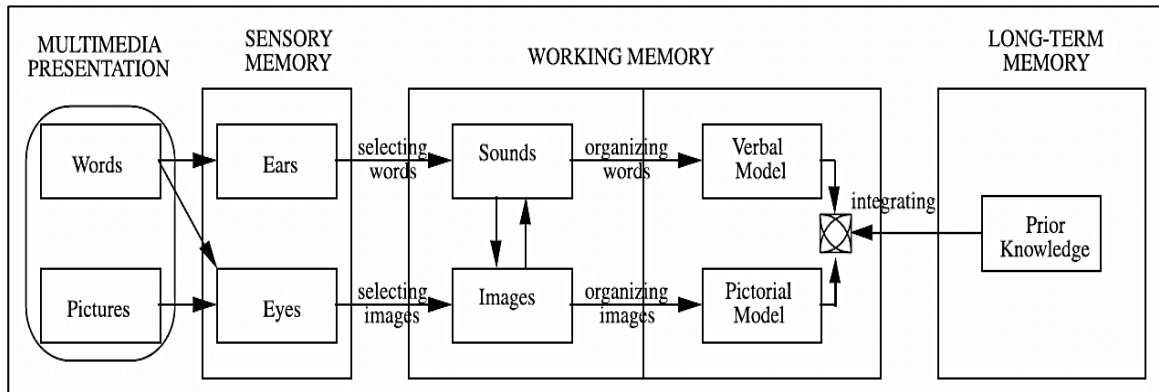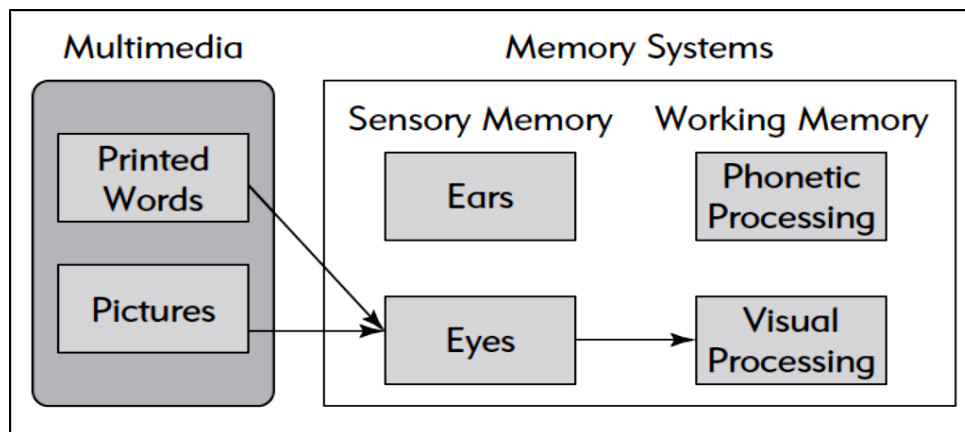**Cognitive Theory of Multimedia Learning (Mayer, 2001)**

Figure 2 Overloading of Visual Channel with Presentation of Printed Words and Pictures
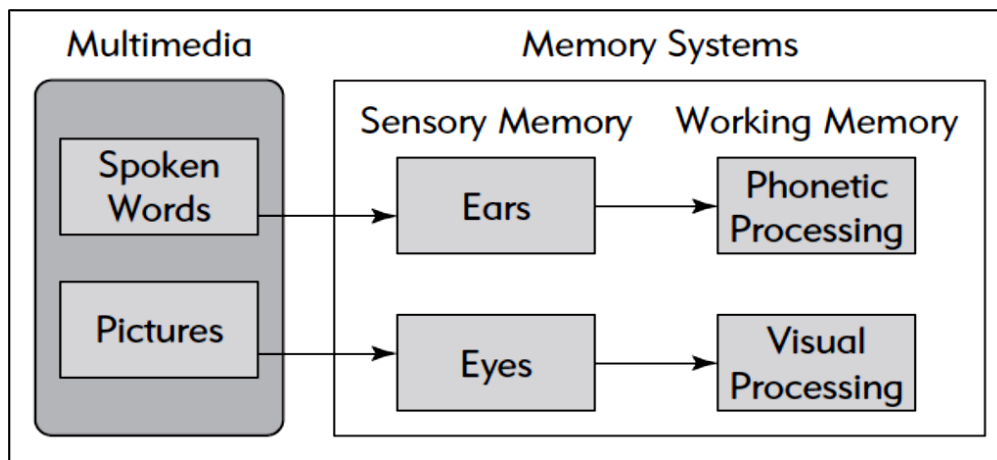

Figure 3 Balanced Content Across Visual and Auditory Channels with Presentation of Narration and Graphics.