Reexamining the modality effect from the perspective of Baddeley's working memory model

Ying-Hua Guan
Graduate Program: Task-oriented Communication
Department of Linguistics and Literary Science
University of Bielefeld
P.O. Box: 10 01 31
33501 Bielefeld, Germany
yinghua.guan@uni-bielefeld.de

Abstract: Cognitive load theory is widely accepted by instructional designers, since it provides a theoretical foundation of designing guidelines for constructing learning material in a way that enhances learning. According to this theory, learning will be impaired if the learning material causes a cognitive overload. Since the capacity of working memory is very limited, the theory assumes that presenting different sources of information in the same modality (e. g. only visually) easily results in a *split-attention effect*, which leads to poor learning performance. To avoid this, a method suggested by cognitive load theory is to present information in different modalities (e. g. auditory text plus visual displays). In this article, I would like to address the issue of the effects of presentation-modality on information processing from the perspective of Baddeley's working memory model. I shall shortly report a study which yields evidence supporting my argument that the modality effect proposed by cognitive load theory should be revised or at least, needs further modification.

Introduction

In contrast to printed media, computer-based multimedia presentations are capable of presenting different types of information in different modalities at the same time. For example, text can be presented either visually or auditorily together with static or moving pictures. Due to this advantage, multimedia presentations are widely used for displaying learning materials. However, some studies show that multimedia learning materials do not necessarily lead to better learning results. In some cases, using multimedia presentations could even impair learning. The cognitive load theory proposed by Sweller and his coworkers (Sweller et al., 1990; Chandler and Sweller, 1991; Sweller, 1993, 1994) provides explanations why multimedia presentations fail to facilitate learning. The theory is principally based on the assumption that learning is ineffective if the learning material or instruction imposes heavy cognitive load on learners, that is, causing an overload of working memory. In light of the limited capacity of working memory, instruction should be constructed in a way that the information to be learned can be processed within the capacity of working memory.

Cognitive load theory predicts that the inappropriate design of instruction such as presentations that require learners to split their attention between text and diagrams (the split-attention effect) will overburden the limited capacity of working memory because learners have to carry out extensive visual search between the text and the diagrams in order to understand the instruction, which in turn inhibits learning. To circumvent this problem, it is suggested that text and diagrams should be physically integrated to eliminate the unnecessary visual search. Another way to avoid split-attention effect, according to the cognitive load theory, is to present information in different modalities rather than in a single modality (the modality effect). Consequently, it is better to present text auditorily if it is combined with graphics (Mousavi et al., 1995; Chandler & Sweller, 1992; Kalyuga et al., 1999).

This view that dual-modality presentation (DMP) is better than single modality presentation (SMP) is also supported by a series of studies conducted by Mayer and his colleagues (Mayer and Anderson, 1991, 1992; Mayer and

Moreno, 1998, 2002; Mayer et al., 1999). Based on their findings, subjects who received animation (showing the process of lightning formation or how a car's braking system works) with concurrent narration outperformed the subjects who received animation with concurrent on-screen text in recalling the content, in matching named elements in the illustration, and in problem-solving test. Moreover, presenting narration before or after the animation was proved to impair learning if the narration as well as the animation contain much information. If the narration and the animation were divided into small portions, the effect of successive presentation did not differ significantly from that of the concurrent presentation. Mayer et al. therefore concluded that presenting animation with on-screen text produces a split-attention effect, which results in poor learning performance. Animation and text are better displayed together rather than separately, whereby text should be presented auditorily so that the verbal and the pictorial information is held in working memory simultaneously, which makes the construction of referential connections between text and pictures easier.

The modality effect and working memory

The modality effect postulated by the cognitive load theory is based on the assumption that working memory, as suggested by Baddeley (1986, 1992, 1997), comprises at least three components: *a central executive* (CE) and two subsidiary slave systems, namely *the phonological loop* and *the visuo-spatial sketchpad.* The phonological loop is responsible for processing verbal information, while the visuo-spatial sketchpad is responsible for processing visual information. According to the cognitive load theory, presenting text auditorily together with pictures might increase effective working memory capacity, since both auditory and visual channels are used to process the information. This argument appears to be plausible at the first glance, and it might be valid under certain circumstances. However, if we reconsider Baddeley's working memory model, the modality effect is not necessarily true. Before I start discussing this point, I would like to give some detailed information of Baddeley's model.

In the model originally proposed by Baddeley and Hitch (1974), it is assumed that the phonological loop consists of two components: a phonological store for holding speech-based information for about 2 seconds, and an articualtory control process which can transform printed material into phonological codes and refresh the information held in the phonological store by subvocal rehearsal. The visuo-spatial sketchpad is capable of setting up and manipulating mental images. It is assumed that the visual and the spatial information is probably processed as separable components in this system. The CE is a capacity-limited attentional system, which supervises and coordinates the visuo-spatial sketchpad and the phonological loop. The functions of the CE, according to Baddeley (1996, 2001), involve 1) coordinating two tasks, 2) switching retrieval strategies used for the random generation task, 3) selectively focusing attention on one stimulus and inhibiting the disrupting effect of others, which is based on the assumption that “anything that limited attentional capacity would impair performance” (Baddeley, 2001: 856), and 4) holding and manipulating information from long-term memory. Whether the CE should be regarded as “a single coordinated system that serves multiple functions, a true executive, or a largely autonomous control processes□ an executive committee“ (Baddeley, 1996: 26) remains an open question.

There are, however, some problems in Baddeley's model. Firstly, the CE is supposed to be capable of combining the information from working memory with that from long-term memory. Yet, the functions of the CE described so far have not yielded any concrete information about how the CE and long-term memory interact with each other. Secondly, the model does not explain how and where the CE combine the verbal and visual information from the two subsystems. Recently, Baddeley (2000) proposed a fourth component to the working memory model: the episodic buffer (see figure 1).

The episodic buffer is supposed to be the place where information from the subsystems of working memory and that from long-term memory is integrated. The buffer is able to chunk information and hold coherent chunks of information. The integration of different codes is explained by assuming that the buffer uses a kind of common code. Moreover, the episodic buffer is supposed to depend heavily on the CE because there is no direct link between the buffer and the phonological loop as well as the visuo-spatial sketchpad. The integration of information from the two subsystems and that from long-term memory is still mainly controlled by the CE. In other words, the role of the episodic buffer is to provide a workspace where the integration of information as well as the influence of long-term memory in the process of information processing (e.g. chunking) take place and offer a temporary storage with limited capacity for the integrated information.

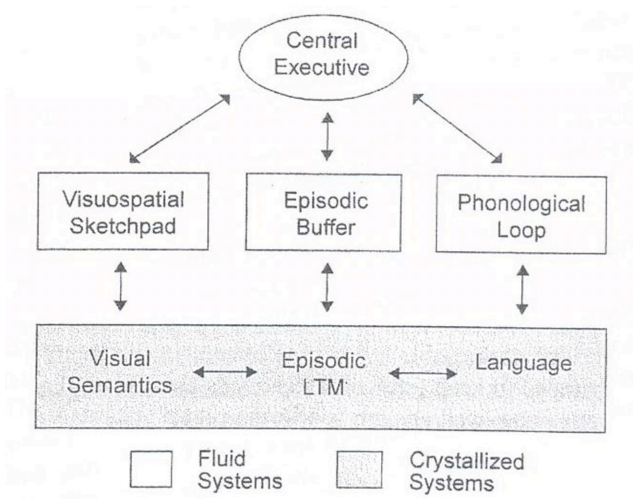

Figure 1. The revised working memory model (taken from Baddeley, 2000).

In relation to the problem of the modality effect mentioned above, I would like to discuss the processes how verbal and pictorial information presented in different modalities are processed or integrated by working memory. Consider Baddeley's working memory model, it is to bear in mind that no matter how the verbal and pictorial information is displayed, the integration of both information is carried out by the episodic buffer which is controlled by the capacity-limited CE. That is, working memory could be overloaded if the amount of information to be coordinated or integrated at a time exceeds the capacity of the CE. The cognitive load theory says that the instruction presented in single-modality, where text and pictures are both shown visually, would require learners to "split their attention" among text and pictures in order to comprehend the instruction. Using DMP can circumvent this problem, which is similar to the physical integration of text and pictures (Kalyuga et al., 1999). In my view, it is not plausible to assume that DMP does not cause any split-attention effect when learners are required to process visual and auditory information at the same time, since both information is not only processed by its corresponding system in working memory, but also has to be integrated by the CE, which is crucial for comprehending the instruction. With the SMP, the textual and the pictorial information can only get into working memory successively because people can not read the text and inspect the pictures simultaneously. In this case, the load imposed on the CE might not be as heavy as that emanating from the DMP. When auditory and visual information must be received and processed at the same time, it is, in my view, analogous to the typical dual-task paradigm employed in the psychological experiments for testing subjects' ability to coordinate two tasks. Since the capacity of the CE is limited, the attentional resource must be divided when performing the dual task, which results in a trade-off between the two tasks. In contrast, there is no such trade-off if the two tasks are performed successively.

There was an interesting study conducted by Jeung, Chandler and Sweller (1997), in which the modality effect was modified. They conducted experiments to examine under what conditions DMP is superior to SMP. The results suggest that when a diagram presented together with (auditory or visual) text was visually high-demanding, DMP was not superior unless a visual aid (flashing) was used to guide learners' visual attention. On the contrary, when the diagram was visually easy to process, DMP was beneficial. The results of this study are still questionable in my view. Firstly, if the diagram is very complex, DMP, of course, is no better than the SMP given that both presentations would overburden working memory anyway. However, another possibility that should not be ruled out is that DMP is more likely to overload working memory than does the SMP because the former actually imposes heavier load on the CE by forcing it to process visual and auditory information simultaneously. Secondly, if the diagram is very easy, DMP is not necessarily be superior to the SMP, since the total visual demand in both conditions is low. Therefore, the learning performance of both conditions should be about the same. The question is, what about the diagram with medium complexity? In my opinion, DMP might be beneficial in this case provided that the total amount of information to be processed at a time does not exceed the capacity of the CE. In attempt to test these conjectures, I conducted an experiment to compare learning performance of different presentation conditions, under which diagrams with different complexity-levels were presented together either with visual or auditory text.

The instruction employed in this experiment explained the movement rules of the Chinese Chess pieces. There were altogether seven different Chinese Chess pieces, each of which was depicted by a single web page. On the page, there was a static diagram combined either with a visual or an auditory text. The links to other pages were put in a table at the bottom of the each page. The complexity of the diagram was varied in three levels-simple, medium, and complex (examples see figure 2, figure 3, and figure 4)- by adding visual distractors and using visually more demanding small drawings instead of simple dots. In addition to that, I animated the complex diagrams to investigate whether animation can successfully guide subjects' visual attention to the moving chess pieces and therefore improve learning. The animation was presented either with visual or auditory text (see figure 5) as well. Subjects had either no or very little experience in playing chess. They were randomly assigned to the eight experimental conditions to learn the instruction as long as they wanted until they feel that they can remember all the

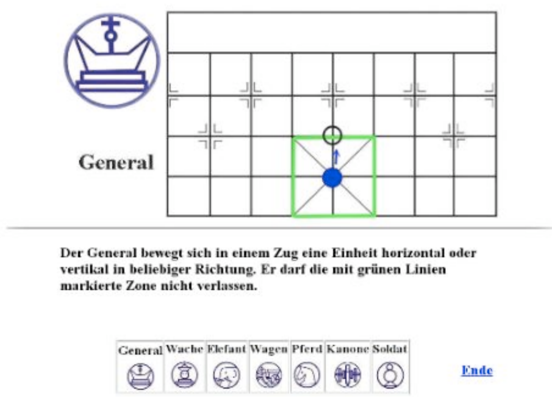

Figure 2. Simple diagram + visual text

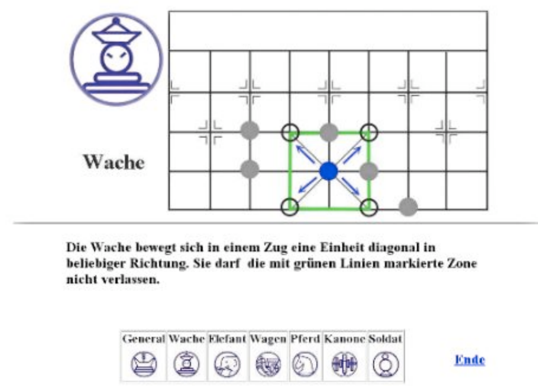

Figure 3. Medium diagram + visual text

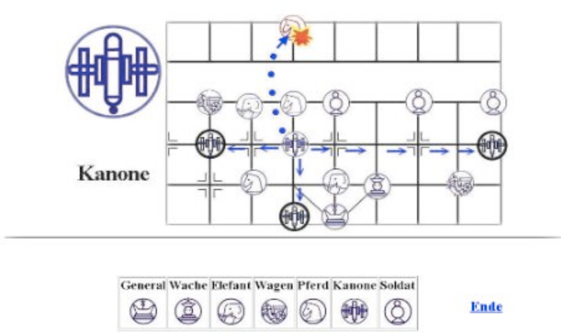

Figure 4. Complex diagram + auditory text

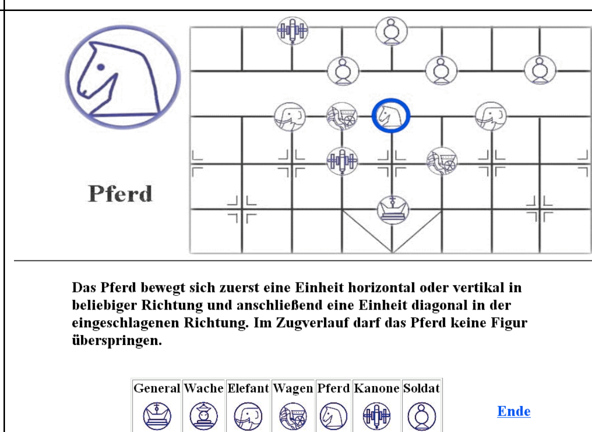

Figure 5. Animation + visual text

rules, but they were encouraged to learn as soon as possible. While subjects were learning the instruction, their eye movements were measured by an eye-tracker. After they finished learning, they were asked to take a test to assess their learning performance. The results show that there were no significant difference in error rate between the DMP-condition and the SMP-condition when the diagrams were simple. With medium picture complexity, the DMP tended to be slightly better than the SMP, but the difference was not significant. However, when the diagrams were complex, the error rate of the DMP-condition was significantly higher than that of the SMP-condition. When animated complex diagrams were presented, the DMP-condition outperformed the SMP-condition. It is to note that the performance of the subjects who received static complex diagram and visual text did not substantially differ from that of the subjects who received simple or medium diagrams (combined with visual or auditory text) and the subjects who received animation and auditory text. All in all, whether the diagrams were static or animated did not affect the learning performance.

Discussion

The results of the Chinese Chess Experiment confirm the assumption that the DMP is not superior to the SMP when the diagrams were easy to process. Nevertheless, when the diagrams were visually demanding, the DMP did impose heavier load on the CE than did the SMP, which points out that the amount of information to be processed at the same time determines whether working memory is overloaded. The entire results of this experiment, however, do not show any positive effect of the DMP except when animation was involved. The disadvantage of presenting animation with visual text lies in the structural interference of the task. Since subjects could not view the animation and read the text at the same time, they could only do it successively or try to move their eyes frequently and rapidly between the animation and the text. In the second case, subjects can hardly coordinate the information of animation and visual text because they certainly miss some information from the animation when they move their eyes to read the text. Based on the observation of eye-movement behavior, subjects did use speedy eye movements as the strategy to coordinate the animation and the visual text. As the result showed, subjects' performance was the worst even though they were allowed to repeat the instruction as often as they would like. Animation synchronized with auditory text, in contrast, is easy to coordinate and to comprehend because it avoids the structural interference. Besides, subjects' eye movements showed that animation indeed guided their visual attention to the moving chess pieces on the complex diagrams, and therefore, eliminated unnecessary visual search. Nonetheless, whether the performance of subjects was simply improved by the visual-guiding function of animation or eventually by directly seeing how the chess pieces move can not be clearly explained by the given data. Furthermore, the learning performance did not differ whether subjects received animated or static diagrams in the instruction. I assume that the static diagrams shown in this experiment are already clear enough in demonstrating how the chess pieces move. Therefore, it is not necessary to depict the rules by showing them in a real motion.

Conclusion

The study I reported above yields evidence indicating that the modality effect proposed by the cognitive load theory needs to be modified. The so-called split-attention effect caused by the SMP might not be as strong as that elicited by the DMP. This argument is based on the assumption that the CE (or the episodic buffer) is the crucial component of working memory that is responsible for coordinating and integrating verbal and pictorial information, and that the DMP imposes heavier load on the CE than does the SMP if the information to be processed is complex. Consequently, we should be cautious in using DMP for instruction, and we should not underestimate the usage of SMP. Whether animation is better than static pictures depends on many factors. It appears that animation could only be beneficial when static pictures fail to present the instruction appropriately. When combined with text, animation should be better synchronized with auditory text.

References

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Baddeley, A. D. (1992). Working Memory. *Science*, 255, 556-559.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology,* 49A(1), 5-28.

Baddeley, A. D. (1997). *Human Memory: Theory and Practice.* (Revised Addition). UK: Psychology Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences,* 4(11), 417-423.

Baddeley, A. D. (2001). Is Working Memory Still Working? *American Psychologist,* Vol. 56, No. 11, 851-864.

Baddeley, A. D. and Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation,* Vol. 8. New York: Academic Press.

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

Chandler, P. and Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology,* 62, 233-246.

Jeung, H. J., Chandler, P. and Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology,* 17(3), 329-343.

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

Mayer, R.E. & Anderson, R.B. (1991). Animation needs narrations: An experiment test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83(4), 484-490.

Mayer, R.E. & Anderson, R.B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84(4), 444-452.

Mayer, R.E. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing Sytem in Working Memory. *Journal of Educational Psychology*, 90(2), 312-320.

Mayer, R.E., Moreno, R., Boire, M., and Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology*, 91(4), 638-643.

Mayer, R.E. & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, 12, 107-119.

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

Sweller, J. (1993). Some cognitive processes and their consequences for the organisation and presentation of information. *Australian Journal of Psychology*, *45*, 1-8.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295-312.

Sweller, J., Chandler, P., Tierney, P., and Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, 119(2), 176-192.

Sweller, J. and Chandler, P. (1994). Why some material is difficult to learn? *Cognition and Instruction*, 12(3), 185-233.