THE MODALITY OF TEXT IN MULTIMEDIA INSTRUCTIONS

refining the design guidelines

Huib K. Tabbers

Educational Technology Expertise Centre
Open University of the Netherlands
Heerlen

ISBN: 90-9016045-0

CONTENTS

CHAPTER 1 - General introduction

People learn. They learn when they read a novel, when they watch their favourite television show, when they are idly surfing the Internet and maybe even when they are sleeping with their homework under their pillow. Sometimes, however, we want this learning to be effective and efficient. Especially in education. And it is the challenging task of educational science to provide guidelines on how to stimulate this effective and efficient learning. One of the issues that educational science is dealing with is the question how to present information in instructional materials in such a way that learning is optimised. Recent developments in multimedia techniques have increased the interest in the effect that different presentation modes have on learning. Nowadays, instructional designers have a whole array of presentation modes at their disposal like text, pictures, audio, video and animations, which can all be brought into action. As a result, the learner is bombarded with multimedia instructions that are bursting with colour, sound and movement. Learning has become fun again, some instructional designers proudly proclaim. That might be true. Or not. In any case, the fun will only be complete if multimedia instructions are also effective.

Unfortunately, educational science has not yet been able to provide a clear and coherent collection of guidelines for designing effective multimedia instructions. In most cases, designers' intuitions on what might work and aesthetic considerations are the main driving forces behind the development of these kinds of learning materials. Although these viewpoints are important, there is a lack of unambiguous guidelines that derive from a theory of multimedia learning and that are supported with ample empirical evidence. A first step in the right direction, however, comes from two recent lines of research that have developed several guidelines for instructional design, based on what is known about human cognitive architecture. Both cognitive load theory (Sweller, 1988, 1994, 1999; Sweller, Van Merriënboer & Paas, 1998) and Mayer's generative theory of multimedia learning (Mayer, 1997, 1999, 2001) claim that information should be presented in such a way that the learner's limited working memory resources are employed as efficiently as possible. Especially with multimedia instructions, where learners have to integrate different information sources like text, pictures and spoken word, cognitive overload can be a serious threat to learning. Therefore, both theories have yielded guidelines that aim at a more efficient use of the learner's cognitive resources. The results of a large number of experiments have shown that application of these guidelines produced less cognitive load, a more effective learning process and higher learning outcomes.

One of the guidelines for multimedia learning is that text accompanying a picture or animation should be presented as spoken text, rather than visual text. Presenting information in two sensory modalities rather than one leads to a more efficient use of memory resources because the modality-specific subsystems of working memory are utilised optimally. This is called *the modality* *effect* (Sweller, 1999) or *modality principle* (Mayer, 2001), and has been demonstrated in a number of studies. The main aim of this thesis is to take a closer look at the modality effect in multimedia learning and refine further the practical guidelines that result from it. Five studies are presented in which the generalisability of the modality effect to different *content domains* is tested, in which the relationship of the modality effect with the *use of visual cues*, *time-on-task* and the *pacing of instructions* is investigated and in which *eye movements* in multimedia learning are studied.

Before the experimental studies, however, the remainder of this chapter will provide an introduction to the aspects of working memory that play an important role in multimedia learning. Furthermore, previous research on multimedia learning is discussed, focusing in more detail on the guidelines resulting from cognitive load theory and Mayer's generative theory of multimedia learning. Finally, the main research questions are introduced and an overview is given of the other chapters of the thesis.

Working memory and multimedia learning

According to Mayer (1992) " …meaningful learning occurs when the learner selects relevant information, organises that information in a coherent whole, and integrates that information with appropriate existing knowledge." (p.408). Working memory plays an essential role in these processes of selecting, organising and integrating information as it is the active part of the human cognitive system where controlled processing takes place.

The concept of working memory was introduced by Baddeley and Hitch (1974), and is strongly related to the modal theory of Atkinson and Shiffrin (1968). According to this theory, human memory consists of a temporary storage system of limited capacity called short-term memory, and a permanent storage system of seemingly unlimited capacity called long-term memory. Through a process of active rehearsal, information in the short-term store can be transferred to the long-term store from which it can be retrieved later on. Working memory is a refinement of the concept of short-term memory and generally refers to the system that is responsible for the maintenance of information relevant for performing a cognitive task. Simply stated, it is the gateway between the external world and the existing cognitive structures in long-term memory. Although present-day theories of working memory differ significantly on the underlying mechanisms (for an overview, see Miyake & Shah, 1999), consensus exists on two aspects that are relevant to multimedia learning. First, most theorists agree that working memory resources are limited, and second, in most models of working memory there are two or more separate modality-specific subsystems apart from a central regulation system.

The issue of limitations in memory has a history that can be traced back to a widely cited article by Miller (1956), in which the author discussed the limited processing capacity of the human mind. Miller distinguished between two mechanisms, immediate memory and absolute judgement, that have different attributes. Immediate memory is a predecessor of short-term memory as its only function is retaining information, and its capacity span is a fixed number of *chunks* (about seven). These chunks, however, can contain an unlimited amount of information assembled through recoding procedures. The second capacity limitation Miller discerned is the span of absolute judgements. He referred to this as the amount of information that an observer can process when trying to tell different stimuli apart, and it is measured in information units called *bits*. This span has several perceptual dimensions, so that adding different features like colour, spatial location and size can increase the information processing capacity of the observer.

Through the years, numerous studies have shown the influence of these limitations in information processing on performance in cognitive tasks (e.g., Anderson, Reder & Lebiere, 1996; Just & Carpenter, 1992; Norman & Bobrow, 1975). Although current working memory models differ in the number of chunks that can be kept active simultaneously, the nature of the limitations makes working memory the bottleneck of the human cognitive system. One way to overcome this problem is by the development of cognitive structures that can be processed as single chunks. Research into expertise has shown that the essential difference between a novice and an expert in a certain domain is in the quantity and organisational quality of available knowledge (Chi, Glaser & Rees, 1982). When confronted with a new problem, an expert will activate a schema that categorises the problem on its deep, structural properties as a single chunk in working memory, and follow the appropriate path to a solution. Novices on the other hand do not possess these schemata and fall back on weak problem-solving strategies like means-ends analysis, which leads to a high cognitive load (Sweller, 1988).

With multimedia learning, limitations in working memory capacity play an important role. If learners have to integrate different information elements in the instructions, like an animation and an explanatory text, the memory load can become high. A mental representation of one element has to be kept active in working memory while searching for the corresponding element. Especially when prior knowledge is low and no schemata exist to guide the search process, cognitive overload is a serious threat to learning (Sweller et al., 1998). The second aspect of working memory that is relevant for multimedia learning is the existence of separate memory resources for different input modalities. In the multiple components theory of Baddeley (1992, 1997), working memory consists of a *central executive* and two slave systems, the *visuospatial sketchpad* and the *phonological loop*. The visuospatial sketchpad is dedicated to processing visual and spatial information and the phonological loop is specialised in acoustic and verbal information. The main function of the central executive is the coordination of information in the two slave systems.

The working of the phonological loop can explain the modality effect found in verbal recall studies (for an overview, see Penney, 1989). These studies showed that recall is better when words are presented in two sensory modalities rather than one. This is a consequence of the working of the *phonological store*, a subsystem of the phonological loop that can preserve an acoustic memory code for at least one minute. According to Penney's separate streams hypothesis, written words are represented in a phonological or visual code, while spoken words are represented in an acoustic as well as a phonological code. So different processing streams exist in working memory for visual and spoken text.

In multimedia learning, the learner often receives information in different modalities, like on-screen text and audio. This has consequences for the way the instructions are processed because different slave systems are addressed. So the choice for a particular presentation modality in instructions influences the way that the available working memory resources are used.

In sum, a limited capacity and modality-specific slave systems are two aspects of working memory relevant for multimedia learning. Design guidelines should aim at an efficient use of memory resources so that cognitive overload is prevented, and try to use the existing modality-specific slave systems as optimally as possible.

Previous research on multimedia learning

For a long time, the search for effective guidelines for multimedia learning was overshadowed by the debate on so-called media effects. This debate centred upon the question whether media make a unique contribution to learning, in such a way that the choice for a certain medium will influence the effectiveness of instructions. This led to a large number of studies in which a lesson in one medium was compared to the same lesson in another medium, for example by comparing classroom lectures with video-based lessons. In a review, R. E. Clark (1983) showed that most of these studies were methodologically confounded, because not only did the medium vary but so did the instructional method (see also R. E. Clark, 1994; R. E. Clark & Craig, 1992). Moreover, hardly any consistent findings were reported, which led Clark to the conclusion that a medium is nothing but the grocery truck delivering the message, and that the effect on learning is not the result of the choice of medium, but of the instructional method applied. Other researchers like Kozma have challenged Clark's conclusions, by claiming that each medium has its own specific attributes that influence the effectiveness of an instructional method (Kozma, 1991, 1994). Nevertheless, the discussion on media effects has more or less subsided, not in the least because current multimedia computers enable the simultaneous use of different media formats. Thus, the somewhat unproductive search for media effects has been replaced by a search for guidelines for designing effective multimedia learning environments.

During the last ten years, several guidelines for multimedia instructions have been proposed (for example, Najjar, 1998; I. Park & Hannafin, 1994; O. Park, 1998). However, a coherent theoretical framework is typically lacking, so that technical developments rather than theoretical considerations have pushed the search for guidelines. Moreover, the empirical support for these guidelines is often inconclusive or even contradictory. Two research lines that seem to be more promising in that respect are cognitive load theory (Sweller, 1988, 1994, 1999; Sweller et al., 1998) and Mayer's generative theory of multimedia learning (Mayer, 1997, 1999, 2001). The guidelines for multimedia learning that both theories have produced are not based on what certain technologies might do, but rather on how people learn with multimedia instructions and how the characteristics of working memory influence the learning process.

Cognitive load theory has mainly been developed by Sweller and his colleagues at the University of New South Wales in Sydney. The central idea of the theory is that the working memory load of instructions should be one of the principal concerns for instructional designers. The available cognitive resources of the learners should be directed to the learning process itself and not to irrelevant features of the instructional materials. The theory distinguishes between *intrinsic* and *extrinsic* load of instructions. The intrinsic load is caused by the complexity of the learning task, and it is the basic amount of processing required to understand the instructions. The extrinsic load on the other hand is the processing that is related to the way that the instructions are presented. That means that intrinsic load is only dependent on the learning content and the learner's expertise, but that extrinsic load can be manipulated by the instructional designer. A further distinction is made between extrinsic load that promotes the construction of cognitive schemata (*germane load*), and extrinsic load that does not contribute to learning (*extraneous load*). Especially when there is a risk of cognitive overload, the extraneous load of instructions should be minimised. In multimedia learning, the necessary mental integration of information elements leads to a high cognitive load, so guidelines are required that keep the extraneous load as low as possible.

Mayer's generative theory of multimedia learning is the result of a research programme that he and his colleagues at the University of California, Santa Barbara, have pursued for several years. This programme focused on the use of multimedia instructional messages that showed how lightning storms develop, how car braking systems work and how bicycle tyre pumps work. In the theory, three main assumptions are made on the way people process these kinds of instructions. First of all, learners engage in active processing of the material. A coherent mental representation is created because the learner selects information, organises the selected information and integrates it with existing knowledge structures. Second, humans have separate processing channels for auditory and visual information. Mayer relates this dual-channel assumption not only to the phonological loop and the visuospatial sketchpad of Baddeley's working memory model (Baddeley, 1992, 1997), but also to dual coding theory (J. M. Clark & Paivio, 1991; Paivio, 1986). In this theory, separate representation systems exist for verbal and non-verbal information. That implies that visual words and spoken words are initially processed in different channels, but are subsequently represented in the same verbal system. The final assumption of Mayer's model is that the capacity of each processing channel is severely limited, and that the extraneous load of instructions should thus be minimised.

So both cognitive load theory and Mayer's generative theory of multimedia learning take the processes in working memory as the starting-point for developing instructional design guidelines. This has resulted in three main guidelines for multimedia learning that can be summarised as follows:

1. Get rid of redundant information
2. Prevent split attention
3. Use spoken text, not visual text

These three design principles will be briefly discussed, together with an overview of the empirical support from experimental studies.

Guideline 1: Get rid of redundant information

When designing instructional multimedia messages, leave out any redundant information. This design principle is called the *redundancy effect* in cognitive load theory and the *coherence* or *redundancy principle* in Mayer's theory. The argument is that any redundant information in multimedia instructions will increase the extraneous cognitive load, because part of the learner's working memory capacity is used for the processing of unnecessary information that does not contribute to learning. A number of experiments have shown that removing superfluous information from multimedia instructions indeed resulted in more effective learning. These studies can be subdivided in three categories related to the kind of information that is redundant.

First of all, information that is *irrelevant to learning* but only meant to make the multimedia instructions more fun is redundant. These information elements are often added to spice up the learning materials and keep the students motivated. However, these "seductive details" as Mayer calls them, seem to do more harm than good to learning, as the results of several studies have shown. Adding text and pictures or video clips that illustrate the subject of the learning material to make it more interesting (Harp & Mayer, 1997, 1998; Mayer, Heiser & Lonn, 2001, experiments 1, 3 and 4), or adding entertaining background music and sounds (Moreno & Mayer, 2000) all resulted in a decrease in transfer performance.

Second, information that can also be derived from *other information elements* is redundant. In a lot of cases, instructional designers assume that presenting information in multiple forms or extended form will enrich the knowledge construction and in the worst case only have a neutral effect on learning. This is a false assumption, as the results of experiments on the redundancy effect have shown. Presenting a text accompanying an animation or a picture both on-screen and as a narration (Kalyuga, Chandler & Sweller, 1999, experiment 1, 2000, experiment 1; Mayer et al., 2001, experiments 1 and 2; Mousavi, Low & Sweller, 1995, experiments 1 and 2), adding a explanatory text to a diagram that could be understood on its own (Chandler & Sweller, 1991, experiments 3, 4 and 5), or adding the full text to a summary of a science text (Mayer, Bove, Bryman, Mars & Tapangco, 1996) all had a negative effect on learning.

Third, information that the learner is *already familiar* with is redundant. As a learner develops expertise in a learning domain, information that was at first necessary to understand the multimedia instructions might become superfluous. In a number of experiments, Kalyuga et al. (1998, 2000) showed that over time, adding an explanatory text to a diagram was helpful at the start of a learning trajectory, but at a later stage led to worse performance compared to diagram-only instructions.

In sum, the existing evidence clearly shows that multimedia instructions should only present information that contributes to the learning process and omit all redundant information. It also implies that the other guidelines for multimedia learning are only relevant if the information elements used in the instructions like text and picture are both necessary for understanding and have to be integrated. In all other cases, leaving out one of the information elements might be the preferred option.

Guideline 2: Prevent split attention

In multimedia instructions, present information elements that refer to each other as close together as possible, so that learners do not have to split their attention between the different information sources. This principle is known in cognitive load theory as the *split-attention effect* and in Mayer's theory as the *contiguity principle*. The explanation for the effect is that the integration of different elements like a picture and a text will be much easier when these elements are presented next to each other. Unnecessary visual search is prevented and the time to keep information elements actively represented is shortened, so that working memory resources are used more efficiently.

Several studies have shown that learning is improved if split attention is prevented either in space or in time. First, placing text elements next to the corresponding parts of a picture or animation led to better learning results (Chandler & Sweller, 1991, experiments 1 and 6, 1992, experiment 1; Mayer, 1989; Mayer, Steinhoff, Bower & Mars, 1995; Moreno & Mayer, 1999, experiment 1; Sweller, Chandler, Tierney & Cooper, 1990, experiments 1, 2 and 3; Tarmizi & Sweller, 1988, experiments 4 and 5; Tindall-Ford, Chandler & Sweller, 1997, experiment 1). Also presenting text and picture simultaneously instead of sequentially improved problem solving transfer (Mayer & Anderson, 1991, 1992; Mayer & Sims, 1994). It must be noted that this temporal split-attention effect was not found when text fragments were very short (Mayer, Moreno, Boire & Vagge, 1999; Moreno & Mayer, 1999, experiment 2; Mousavi et al., 1995, experiments 3 and 4).

The results of the empirical studies show that preventing split attention is especially effective when two information elements like a text and a picture are presented at the same time, which is mostly the case in multimedia learning. The extraneous load caused by the visual search and the mental effort needed to integrate text and picture can be minimised by placing the text inside the picture next to the part it is referring to.

Guideline 3: Use spoken text, not visual text

Whenever a picture or an animation is accompanied by a textual explanation, present the text as a narration rather than as visual text. This *modality effect* (cognitive load theory) or *modality principle* (Mayer's theory) is accounted for by considering the modality-specific slave systems in working memory. With spoken text, the phonological loop is directly addressed and the visual channel is not overloaded. As a result, extraneous load is decreased compared to the situation in which text is presented visually.

In support of this guideline, a number of studies have found superior learning results when visual text in multimedia instructions was replaced with spoken text (Kalyuga et al., 1999, experiment 1, 2000, experiment 1; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi et al., 1995; Tindall-Ford et al., 1997). Further evidence for a decrease in extraneous load comes from Tindall-Ford et al. (1997), who showed that learners in the audio conditions invested less mental effort on the instructions. In one study, the modality effect was not achieved when pictures were high in complexity (Jeung, Chandler & Sweller, 1997). However, by adding visual cues to the pictures that related the spoken text to the right parts of the picture, the modality effect was recovered.

Overall, the evidence for the modality effect has been such as to lead to the general guideline, in both cognitive load theory and Mayer's theory, that in multimedia instructions consisting of verbal and pictorial information, the text should be presented auditorily rather than visually.

A closer look at the modality effect

Although the evidence on the modality effect in multimedia learning seems convincing, the question can be asked whether the guideline to use spoken text in multimedia instructions is as generally applicable as both cognitive load theory and Mayer's theory of multimedia learning seem to suggest. This question is especially relevant from a practical viewpoint, because the production of audio is expensive, and delivering audio puts higher demands on the equipment that is used for presenting the instructions. For example, headphones are needed to prevent learners in groups from disturbing each other. Therefore, the designer of multimedia instructions would like to be sure that the use of spoken text indeed results in more effective learning. The aim of this thesis is to provide more refined guidelines for the use of spoken text in multimedia instructions by taking a closer look at some research issues that can be raised given the evidence so far on the modality effect in multimedia learning.

First, the instructions used in previous studies on the modality effect all dealt with subjects from the exact sciences, like geometry and electrical engineering. Also, the length of the instructions was at most five minutes, so only short instructional messages were presented. Moreover, most of the experiments took place in a laboratory setting, under strictly controlled circumstances. That raises the question whether the modality effect also applies with instructional materials from another content domain, which take more than just a few minutes to present, and whether the effect can be replicated outside the laboratory walls. If not, the guideline to use spoken text will only have limited practical value. So the first important research question is whether the modality effect can be generalised to longer instructions from a different content domain and whether it can also be obtained in an ecologically valid classroom setting.

Second, the instructions used in previous research were either system-paced based on the pace of the narration or, with paper-based instructions, based on the total time of the narration. Thus, the relationship of the modality effect with time-on-task and the pacing of instructions is yet unclear. What will happen, for example, if learners are given control over the pacing of the instructions? It might be argued that learner-pacing is especially effective in instructions with visual text, because learners can take more time to integrate the text and the picture or animation. So the second research question is whether the modality effect also occurs if time-on-task is prolonged by changing the pace of instructions.

Overview of the thesis

The following chapters present five empirical studies on the modality effect in multimedia learning, in which a closer look is taken at the generalisability of the effect and at the relationship with the pacing of instructions.

Chapter 2 focuses on the generalisability of the modality effect in multimedia learning. The main question is whether the modality effect can be replicated with learner-paced instructions on a non-technical subject (instructional design) which take more than just a few minutes to study, and which are administered in an ecologically more valid classroom setting. Moreover, the role of preventing visual search is taken into account by looking at the effect of adding visual cues to the pictures. Four versions of the multimedia instructions are compared, varying in modality (visual text versus spoken text) and the use of visual cues (cues in picture versus no cues in picture). A reverse modality effect obtains, with superior learning results in the visual-text conditions. This is attributed to the fact that the instructions used in the experiment are learner-paced, contrary to the system-paced instructions used in previous studies.

Chapter 3 presents two studies . The first is a replication of the study in the previous chapter, only this time the instructions are system-paced instead of learner-paced. Two versions of the instructions are compared, varying in modality (visual text versus spoken text), and a modality effect shows up in terms of less mental effort in the spoken-text condition. The role of pacing is further investigated in the second study of this chapter. The main question is whether the modality effect can only be obtained with system-paced instructions and not with learner-paced instructions. Four versions of the instructions are compared, varying in modality (visual text versus spoken text) and in pacing (system-paced versus learner-paced). Indeed, the superiority of spoken text over visual text only holds in the system-paced conditions.

Chapter 4 takes a closer look at the interaction of modality and pacing, by taking the role of time-on-task into account. The main question is whether the interaction of modality with pacing is due to a difference in time-on-task or also to some other aspect of pacing. Four versions of the instructions are compared, varying in modality (visual text versus spoken text) and in pacing (system-paced, system-paced with extended time-on-task, and learner-paced). With extended time-on-task, no modality effect ensues, whereas learner-pacing even leads to a reverse modality effect.

In Chapter 5, GazeTracker™ is discussed, a tool for studying eye-movements in dynamic multimedia environments. This tool is demonstrated in a study in which different patterns of looking in multimedia learning are investigated related to modality and pacing.

Chapter 6, the final chapter of the thesis, presents a general discussion of the studies. A review of the results is given, followed by a discussion of the implications for theories of multimedia learning. Furthermore, a set of refined guidelines for multimedia design is presented and suggestions for further research are made.

References

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30,* 221-256.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

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

Baddeley, A. D. (1997). *Human memory: theory and practice* (revised ed.). Hove, United Kingdom: Psychology Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

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

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

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7-75). Hillsdale, NJ: Erlbaum.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3,* 149-210.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53,* 445-459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology, Research & Development, 42,* 21-29.

Clark, R. E., & Craig, T. G. (1992). Research and theory on multi-media learning effects. In M. Giardina (Ed.), *Interactive multimedia learning environments:*

Human factors and technical considerations on design issues (pp. 19-30). New York: Springer-Verlag.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional and cognitive interest. *Journal of Educational Psychology, 89,* 92-102.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90,* 414-434.

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

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40,* 1-17.

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

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92,* 126-136.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61,* 179-211.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology, Research & Development, 42,* 7-19.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81,* 240-246.

Mayer, R. E. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology, 84,* 405-412.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32,* 1-19.

Mayer, R. E. (1999). Research-based principles for the design of instructional messages. *Document Design, 1,* 7-20.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83,* 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,* 444-452.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88,* 64-73.

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,* 187-198.

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,* 312-320.

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

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,* 389-401.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology, Research & Development, 43,* 31-44.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92,* 117-125.

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

Najjar, L. J. (1998). Principles of educational multimedia user interface design. *Human Factors, 40,* 311-323.

Norman, D. A., & Bobrow, D. C. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44-64.

Paivio, A. (1986). *Mental representations: A dual coding approach.* New York: Oxford University Press.

Park, I., & Hannafin, M. J. (1994). Empirically-based guidelines for the design of interactive multimedia. *Educational Technology, Research & Development, 41,* 66-85.

Park, O. (1998). Visual displays and contextual presentations in computer-based instruction. *Educational Technology, Research & Development, 46,* 37-50.

Penney, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17,* 398-422.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

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

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

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

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-296.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80,* 424-436.

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

CHAPTER 2 - Multimedia instructions and cognitive load theory: Effects of modality and cueing*

Abstract

According to cognitive load theory (Sweller, 1999) and Mayer's theory of multimedia learning (Mayer, 2001), replacing visual text with spoken text (*the modality effect*) and adding visual cues relating elements of a picture to the text (*the cueing effect*) increase the effectiveness of multimedia instructions. The aim of this study was to test the generalisability of both effects in a classroom setting. The participants were 111 second-year students of educational science (age between 19 and 25 years), who studied a web-based multimedia lesson on instructional design for about one hour and completed a retention and a transfer test. During instructions and tests, self-report measures of mental effort were administered. Adding visual cues to the pictures resulted in higher retention scores, whereas replacing visual text with spoken text resulted in lower retention and transfer scores. A possible explanation for the reversed modality effect is that the instructions were learner-paced, and not system-paced like in earlier research.

The use of multimedia computers in education has led to the development of all sorts of instructional material in which verbal and non-verbal presentation modes are combined. Unfortunately, educational research has not yet identified how to design effective multimedia instructions. However, two recent lines of research that have yielded some promising results are the work on cognitive load theory by Sweller and others (for an overview, see Sweller, 1999) and the experiments on multimedia learning carried out by Mayer and colleagues (for an overview, see Mayer, 2001). Both researchers base their instructional design principles on human cognitive architecture and the way in which the multimedia material is processed. In his theory of multimedia learning, Mayer (2001) describes how the learner builds mental representations of multimedia instructions. One important step in this process is the integration of both verbal and visual information in working memory. For example, when instructions consist of a picture and an explanatory text, the learner has to switch back and forth between the two to integrate them mentally. This process is cognitively demanding, at the expense of mental resources that could otherwise be allocated to the learning process.

---

* based on: Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (in press). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology.*

Cognitive load theory calls the unnecessary memory load caused by the presentation format of instructions *extraneous* load (Sweller, Van Merriënboer & Paas, 1998). Changing the presentation format can lower this extraneous load and increase the effectiveness of instructions. For example, Sweller and others have shown that the physical integration of verbal and visual information resulted in improved test scores (Chandler & Sweller, 1991, 1992; Kalyuga, Chandler & Sweller, 1998; Sweller, Chandler, Tierney & Cooper, 1990; Tarmizi & Sweller, 1988). When a textbox is placed right next to the part of the picture the text is referring to, the need to mentally integrate text and picture is reduced, which lowers the extraneous load and facilitates the learning process. Sweller et al. (1998) call this the *split-attention effect*. A similar effect has been demonstrated by Mayer and colleagues in a series of experiments in which they showed that multimedia instructions were more effective when verbal and visual information were presented close to each other rather than spatially separated (Mayer, 1989; Mayer, Steinhoff, Bower & Mars, 1995; Moreno & Mayer, 1999). Mayer (2001) calls it the *contiguity principle*.

A more recent finding is that multimedia instructions can be more effective when the verbal information is presented auditorily instead of visually. This is called the *modality effect* (Sweller et al., 1998) or *modality principle* (Mayer, 2001). A number of experiments have demonstrated that replacing written or on-screen text with spoken text improved the learning process in different ways: lower mental effort during instruction and higher test scores (Tindall-Ford, Chandler & Sweller, 1997), less time on subsequent problem solving (Jeung, Chandler & Sweller, 1997; Mousavi, Low & Sweller, 1995), and improved scores on retention, transfer and matching tests (Kalyuga, Chandler & Sweller, 1999, 2000; Mayer & Moreno, 1998; Moreno & Mayer, 1999). The authors explain these diverse results by referring to the working memory model of Baddeley (1992). In this model, working memory has two modality-specific slave systems, one for visual and spatial information and one for acoustic information. When information is presented in two sensory modalities (visual and auditory) rather than one, both slave systems are addressed and total working memory capacity is used more efficiently. So, relative to the available resources, the extraneous load of the multimedia instructions is reduced.

Both strategies, physically integrating text and picture and replacing written or on-screen text with spoken text, reduce the extraneous load of multimedia instructions and thus increase the effectiveness of the learning process. In both cases, this reduction in cognitive load can partly be accounted for by the reduction in the amount of visual search needed to integrate text and picture. The effect of reducing visual search has been explicitly demonstrated in two studies by Jeung et al. (1997) and by Kalyuga et al. (1999). Jeung et al. showed that replacing visual text with spoken text does not always improve the effectiveness of multimedia instructions, especially when pictures with a high visual complexity are used. Only when they added to the pictures in the bimodal condition visual cues that related the right elements in the picture to the accompanying spoken text, did they recover the modality effect in terms of shorter time on subsequent problem solving. Kalyuga et al. found the same *cueing effect* with visual-only instructions. In one experiment they used colour-coding to link on-screen text with corresponding parts of the picture. This resulted in better test scores when compared to instructions without any visual cues.

In the studies in which the modality effect was demonstrated, the authors claimed that the reduction in extraneous load of the multimedia instructions resulted from a more efficient use of the available memory resources. Nonetheless, the results obtained in the experiments could also largely be explained in terms of a reduction in visual search. For example, Jeung et al. (1997), Mousavi et al. (1995) and Tindall-Ford et al. (1997) used visual-only instructions in which *all* explanatory text was printed next to the diagram and compared it to instructions in which the students saw only the picture and could listen to the explanation. That means that they not only replaced visual text with spoken text, but also drastically reduced the visual search necessary to link the right parts of the text with the right parts of the diagram. So in their experiments, the difference in effectiveness between bimodal and visual-only instructions could be largely attributed to the difference in visual complexity.

Mayer and Moreno (1998; Moreno & Mayer, 1999) and Kalyuga et al. (1999, 2000) on the other hand cut their explanatory texts in smaller pieces, reducing the visual search to a minimum. However, in their experiments the instructions were presented as system-paced animations. The time a student could study a picture and its accompanying texts was determined by the pace of the narration in the bimodal condition. The learners in the bimodal condition could use this limited period of time more effectively because they could look at the picture and listen to the text at the same time, while the learners in the visual-only condition had to spend part of their time in a process of visual search as they had to skip back and forth between text and picture. To adjust for this unwanted effect, Moreno and Mayer (1999) in one experiment used instructions in which the animation and the accompanying text were presented sequentially instead of simultaneously. Despite the temporal detachment of text and picture, bimodal instructions still proved to be superior to visual-only instructions. According to the authors, this result showed that the modality effect is at least partly the result of a more efficient use of working memory resources.

Based on the results obtained in their empirical work, both Sweller and Mayer claim that multimedia instructions will be more effective when the verbal information is presented auditorily instead of visually. However, some reservations can be made on the generalisability of their findings. First, the studies conducted thus far were all tightly controlled laboratory experiments. Moreover, almost all multimedia instructions used in the above-mentioned studies taught subjects from technical domains like geometry (Jeung et al., 1997; Mousavi et al., 1995), scientific explanations of how lightning develops (Mayer & Moreno, 1998; Moreno & Mayer, 1999), electrical engineering (Kalyuga et al., 1999; Tindall-Ford et al., 1997), and reading a technical diagram (Kalyuga, Chandler & Sweller, 2000). Finally, students had only a few minutes to study the instructional material, and the maximum study time was always based on the time needed to hear the narration.

The aim of the present study was to test the generalisability of the modality effect. Therefore, the set-up of the current experiment differed from earlier experiments in a number of ways. First, the multimedia material discussed a non-technical subject matter, namely instructional design. Furthermore, the instruction time was more than an hour, and the instructions were learner-paced instead of system-paced. Finally, the experiment took place in an ecologically more valid classroom setting. To see if a reduction in visual search could partly account for the modality effect, the cueing effect was included in the study as well.

Cognitive load theory would predict that presenting texts accompanying a picture as spoken text will decrease the extraneous load and increase the effectiveness of the instructions (the modality effect), and that adding visual cues to a picture that relate the relevant elements of the picture to the text will prevent visual search and also increase the effectiveness of the instructions (the cueing effect). To study both the modality and the cueing effect, four different versions of the multimedia instructions were created, differing in the modality of text and the use of visual cues. To determine the effectiveness of the instructions, we looked at the extent to which students could recall elements of the instructional design model that had been studied in a retention test, and at the extent to which they could apply the model in a new situation in a transfer test. Furthermore, to draw conclusions not only about the effectiveness but also about the efficiency of the different presentation modes, we used a self-report measure of mental effort during both the instructions and the tests, and recorded the total time spent on the instructions.

Method

Participants

The participants were 111 second-year students from the Department of Education of the University of Gent in Belgium (age between 19 and 25 years; 16 males and 95 females). Originally, 114 students participated, but the results of three students were removed from the sample because they had not completed the instructions in the maximum time. The experiment was part of a regular course on instructional design, but at the time of the experiment the students had not received any lessons yet. Before this course, they had not been taught any instructional models, so the subject matter was completely new for them. Furthermore, all students were accustomed to working with multimedia computers in their studies, and the experiment took place in the classroom in which they normally had their computer-based classes. The students were randomly divided over the experimental groups, with 30 students in the VN group (visual text, no cues in diagram), 26 in the VC group (visual text, cues in diagram), 27 in the AN group (audio, no cues in diagram), and 28 in the AC group (audio, cues in diagram).

Materials

Multimedia instructions
For this study we developed multimedia instructions on the four component instructional design (4C/ID) model of Van Merriënboer (1997). This model describes a design strategy for the training of complex cognitive skills. The instructions focused on how to develop a blueprint for a training programme based on the skills hierarchy of a complex skill. The material was constructed as a website with a linear structure that offered two worked-out examples followed by a general explanation of the design strategy.

The instructions started with two pages containing a short textual introduction to the 4C/ID model. Afterwards, a series of six diagrams followed, representing skill hierarchies and sequences of learning tasks. Together, these six diagrams formed the first worked-out example showing the different stages in developing a blueprint for the training of the complex skill *doing experimental research*. The second worked-out example consisted of three diagrams showing the same process of developing a blueprint for the training of the complex skill *designing a house*, and finally the strategy of the 4C/ID model was summarised in two general diagrams. All 11 diagrams were accompanied by a textual explanation on how the model was applied in the specific situation.

Four different versions of the instructions were created that differed in the way the text accompanying the diagrams was presented, and in the use of visual cues in the diagrams. In the two audio versions, students could listen to the text that accompanied the diagrams through a headphone, while in the two visual versions exactly the same text was depicted right above the diagrams. All explanatory texts were split into smaller fragments of only one or two sentences long, such that each fragment referred to a specific part of a diagram. In the two cued versions, these parts were coloured bright red so that students would know where to look in a diagram, whereas in the two non-cued versions no colour-coding was used. Furthermore, in the audio versions it was possible to replay each text fragment by clicking on a small *play* button. In all versions students could click on a forward arrow to advance to the next text fragment, and a backward arrow to return to the previous fragment (if there was one). The diagrams stayed the same; only the accompanying text fragments and the visual cues (if any) changed. Figure 1 shows a screen example of the visual version of the instructions, and Figure 2 shows the audio version.

Mental effort measure
The self-report measure of mental effort used in this study was a 9-point rating scale ranging from *very, very low mental effort* to *very, very high mental effort*, and was developed as a non-intrusive measure of cognitive load by Paas (1992). The average score on the eleven mental effort rating scales used in the instructions (one for each diagram) was taken as a measure of mental effort during instructions (Cronbach's alpha = .92).

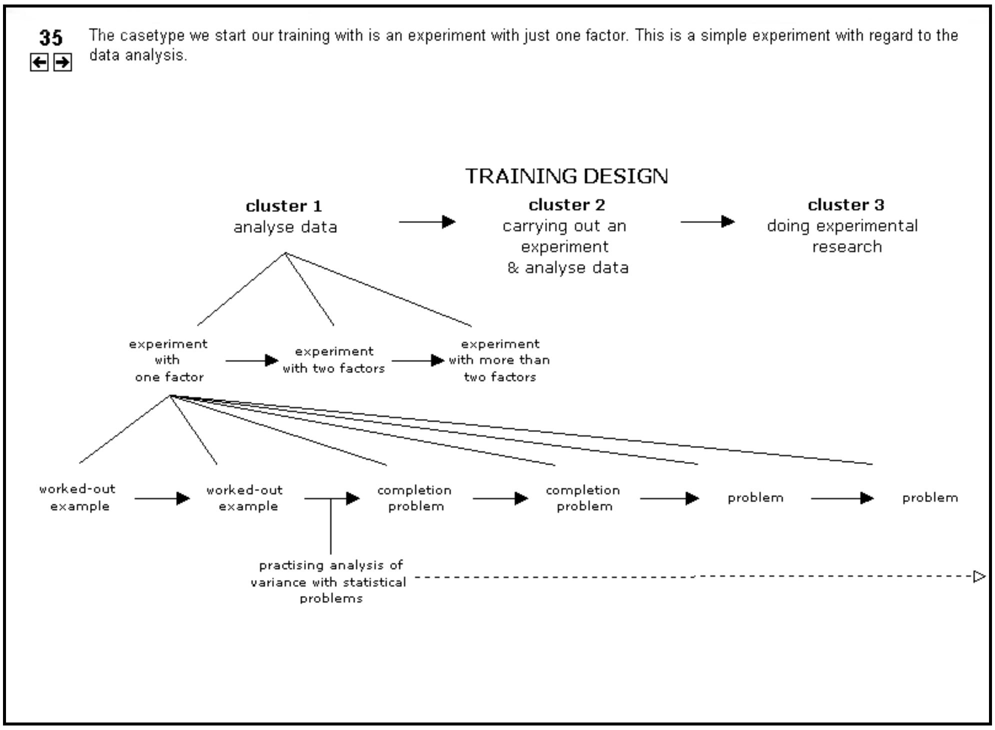


Figure 1
Screen-example of the VN-version (visual text, no cues in picture), showing a diagram of a learning sequence accompanied by a fragment of explanatory text at the top.

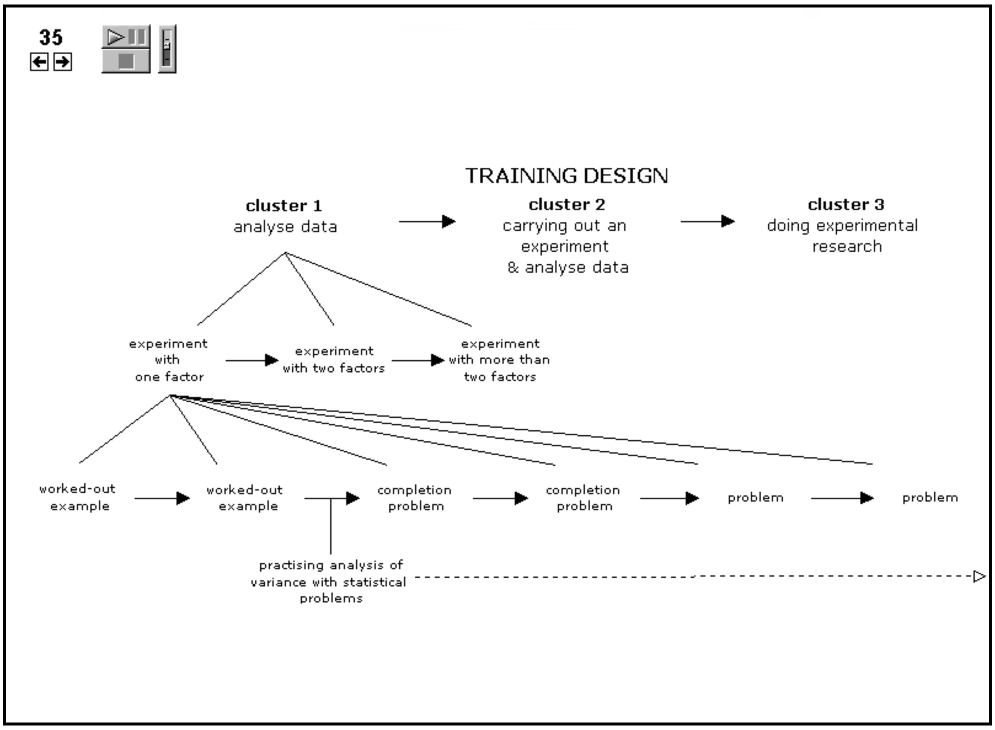


Figure 2
Screen-example of the AN-version (audio, no cues in picture). When a student opened a page, the audio fragment started playing automatically.

Retention test
The retention test originally consisted of two paper-and-pencil tests, one of 30 and one of 20 multiple-choice items. The 30-item test contained only verbal statements, while the 20-item test combined the verbal statements with small parts of diagrams. All items were statements about the 4C/ID model like “A macro-sequence in the 4C/ID model is a series of subskills in a cluster”, or “According to the 4C/ID model, the same subskills can be trained in more than one learning task”, and the students could choose between *correct*, *incorrect* or *I don't know*. Each right answer yielded one point. Together, the sum of the scores on all 50 items formed one total retention score (Cronbach's alpha = .67).

Transfer test
The transfer test was also a paper-and pencil test and consisted of a short description of the skills that an expert researcher applies when searching for literature, in combination with the assignment to design a blueprint for the training of this complex skill on a blank answering form. To score the results of the transfer test, a scoring form was developed consisting of 40 *yes*/*no*-questions that checked to what extent and how accurately the strategy prescribed by the 4C/ID model had been applied in the transfer task. Every *yes* scored one point, and the sum score ranged from zero (no steps from the model taken) to 40 (all steps taken accurately). After the experiment, three independent raters scored the transfer tests using the form. The tests of twenty-six students were scored by all three raters, showing an interrater agreement of .88 (calculated as a single measure intraclass correlation, see McGraw & Wong, 1996).

Procedure

The experiment was carried out in three sessions, and in each session between 35 and 40 students were tested simultaneously. These sessions took place in a classroom that had 40 multimedia computers connected to the Internet through the university network, with 10 computers for each experimental condition. The computers that delivered bimodal instructions had headphones attached to them. The headphones used in the experiment were 'open', so that surround noise was still audible. When the students entered the room, they were randomly assigned to one of the computers. Each computer showed a browser-window (without any of the menu options visible) set on a web page with some general information on how to navigate in the instructional material and how to complete the mental effort scales that were administered during the instructions. In the two audio conditions, students were reminded that they had to wear the headphones during the instructions. Furthermore, it was announced that the students would be tested after the instructions.

All students started at the same time and were given a maximum of 70 minutes to study the instructional material. If they finished earlier, they could do something for themselves in silence, but they were not allowed to leave the classroom or talk to other students. The server on which the instructional website ran kept record of the time spent on the learning task and of the mental effort scores of each participant. After each diagram in the instructions, a separate page followed with a subjective rating scale on which the students could rate the mental effort they had spent. When a student clicked on one of the options of the rating scale, the program automatically continued with the next diagram.

After the instruction phase the three paper-and-pencil tests were administered. Maximum time for each retention test was 10 minutes, and in the transfer test the students got a maximum of 30 minutes to design a training programme. After each test the students rated their mental effort on a 9-point scale similar to the ones used in the instructions. At the end, the students were asked to complete a questionnaire to evaluate how they had experienced the experiment and whether there had been any problems with either the computer or the instructional material. Each session took about two-and-a-half hours.

Results

The variables under analysis were training time, mental effort spent on instruction and on the tests, retention score and transfer score. All scores were analysed with two-factor analyses of variance (ANOVAs), with modality (visual vs. spoken text) and cueing (no cues vs. cues in the diagram) as the between-subjects factors. For all statistical tests, a significance level of .05 was applied. Table 1 shows the average scores on the dependent measures for all four conditions.

Training time was not equal for all conditions, with participants in the visual conditions (*M* = 47.9 minutes, *SD* = 7.2) needing significantly less time than participants in the audio conditions (*M* = 57.3 minutes, *SD* = 7.0), $F(1,107) = 47.27$, MSE = 51.25, $p < .001$. Nevertheless, the results of the evaluation questionnaire showed that the slower downloading of the audio files over the Internet accounted for at least part of the difference.

The average mental effort score during instructions was 4.0 (*SD* = 1.0), which represents a *rather low mental effort*. The students in the visual conditions (*M* = 4.2, *SD* = 1.0) spent a little more effort on the instructions than their colleagues in the audio conditions (*M* = 3.8, *SD* = 1.0). However, this difference did not reach statistical significance, $F(1, 107) = 3.16$, MSE = 0.93, $p = .08$. The mental effort score for the retention test did show a significant effect for the modality of instructions, $F(1, 107) = 11.84$, MSE = 1.12, $p < .01$, because students in the visual conditions (*M* = 6.8, *SD* = 1.0) reported more effort than their colleagues in the audio conditions (*M* = 6.1, *SD* = 1.1). No significant differences were found on the mental effort scores for the transfer test.

A significant effect for the modality of text was found in the retention test, with the visual conditions (*M* = 32.8, *SD* = 5.2) scoring significantly higher than the audio conditions (*M* = 29.4, *SD* = 5.0), $F(1, 107) = 13.13$, MSE = 25.72, $p < .01$. The effect of adding cues to the diagram also reached statistical significance, $F\ (1, 107) = 4.02$, $p < .05$, with a higher score for the cued conditions (*M* = 32.0, *SD* = 5.3) than for the no-cues conditions (*M* = 30.3, *SD* = 5.4).

Table 1
Means and standard deviations of dependent measures for all conditions

	Condition VN (n = 30)		Condition VC (n = 26)		Condition AN (n = 27)		Condition AC (n = 28)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Training Time (Minutes)	47.8	7.5	48.1	7.1	56.6	7.0	58.0	7.0
Mental Effort during Instructions (1-9)	4.1	1.1	4.2	0.8	3.7	0.8	3.9	1.1
Mental Effort on Retention Tests (1-9)	6.7	0.9	6.9	1.0	6.2	1.1	6.0	1.2
Mental Effort on Transfer Test (1-9)	6.4	1.4	6.9	1.4	6.3	1.4	6.4	1.5
Retention Score (0-50)	32.2	4.8	33.5	5.7	28.1	5.2	30.6	4.6
Transfer Score (0-40)	21.6	6.2	22.2	6.7	17.9	5.6	20.0	6.1

The scores on the transfer task showed a significant effect for the modality of the text, $F(1,107) = 6.49$, MSE = 37.62, $p < .05$, in the same direction as in the retention test (visual text: M= 21.9, SD = 6.4; audio: M = 19.0, SD = 5.9) but no effect for cueing.

Discussion

It is clear from the results that the modality and the cueing effects demonstrated in earlier experiments on multimedia instructions have not been replicated in this study. First, the presumed positive effect of adding visual cues to the diagrams is only noticeable in the results of the retention test but not in the transfer test or in any of the mental effort measures. Second, regarding the hypothesised superiority of spoken text over visual text, the results on the tests are even contrary to the expectations. Students in the visual conditions perform better than students in the audio conditions on both retention and transfer tests. In terms of the efficiency of the different presentation formats, the results are somewhat mixed. The mental effort that students spend in the audio conditions is marginally less than the effort that is spent in the visual conditions, which seems to be in line with what Tindall-Ford et al. (1997) have found. Moreover, in the retention test the students in the visual conditions report higher mental effort scores. So the fact that they have obtained higher retention scores can at least partly be explained as a result of investing more mental effort. Conversely, on the transfer test no differences in mental effort appear that can explain the difference in transfer score. Finally, students in the visual conditions have spent significantly less time on the instructions, which only strengthens the conclusion that they have really outperformed their colleagues in the audio conditions.

Why do our results differ so significantly from earlier research on multimedia learning and cognitive load, especially regarding the modality effect? First of all, this study was designed to test the generalisability of the modality effect in a more ecologically valid classroom setting. That may have introduced confounding factors that were excluded in the earlier experiments in controlled laboratory settings. One could easily think of factors like the simultaneous testing of thirty or more students or the use of a flexible but somewhat unstable delivery medium as the Internet. Moreover, in the audio conditions, downloading the fragments took some time, which might have resulted in loss of motivation in the students. Finally, students spent more than one hour studying the instructional material, which contrasts sharply with the few-minutes instructions used in earlier research on the modality effect. Differences in extraneous load that have an influence in short learning tasks may lose their influence as more time-related factors become dominant in the learning process, such as concentration and span of attention. Listening might be even more tiresome or boring than reading, resulting in less motivation. However, the mental effort measures during instruction do not indicate large differences between the conditions, and also the results from the evaluation questionnaires give no indications of differences in motivation or concentration during the instructions. So it does not explain why we find a reverse modality effect.

Another explanation might be that the multimedia instructions used in the present study differ in two ways from the instructions used in earlier studies, both in subject matter and in pacing. First of all, it can be argued that instructional design strategies are more procedural and less descriptive in nature than for example the scientific explanations used by Mayer and Moreno (1998; Moreno & Mayer, 1999). Visual text might be more suitable for presenting procedural information than spoken text, as the learner has more time to reflect on the information. On the other hand, students in the audio conditions had the opportunity to listen to a piece of text as many times as they wished, giving them ample opportunity to elaborate on the information.

The pacing of the instructions might be a more plausible factor explaining why bimodal instructions were not superior to visual-only instructions. This could explain why students in the audio conditions have achieved lower test scores. In the studies by Mayer and Moreno (1998; Moreno & Mayer, 1999) and Kalyuga et al. (1999, 2000), the multimedia instructions were presented as system-paced animations. In the present study, however, the learners could scroll through the explanatory texts at their own pace. Possibly bimodal instructions are only advantageous when animations are system-paced, whereas visual-only instructions are more effective when the learner can set the pace. The advantage of bimodal instructions is that the picture and the text can be perceived simultaneously, resulting in a lower extraneous load than in visual-only instructions where the learner has to skip between text and picture in a limited time. In learner-paced instructions, however, this advantage disappears because the learner with the visual-only instructions has more time to relate the text to the picture. Moreover, with visual texts it is much easier to jump back and forth through the text than with spoken texts that are linear by nature and are much less easy to skim through. So learner-pacing could make visual-only instructions more effective than bimodal instructions and reverse the modality effect. In future research, this hypothesis should be investigated by comparing system-paced with learner-paced bimodal instructions.

Taken together, the results of this study show that the design principles that adding visual cues to pictures and replacing visual text with spoken text will increase the effectiveness of the instructions in multimedia instructions are simply not generally applicable. Although we did find a small positive effect of cueing in our experiment, we could not replicate the modality effect found in earlier studies. Replacing visual text with spoken text even had a negative effect on learning, contrary to what both cognitive load theory and Mayer’s theory of multimedia learning would predict. It seems that a bimodal presentation is only advantageous when the system sets the pace of the instructions, whereas visual-only instructions are the preferred format if the learner is in control. Further research into the conditions under which the modality and cueing effects occur might produce more specific design principles for multimedia instructions that can successfully be applied in real-life educational settings.

References

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

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

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

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40,* 1-17.

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

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92,* 126-136.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81,* 240-246.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

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,* 312-320.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology, Research & Development, 43,* 31-44.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

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

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84,* 429-434.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

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

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-296.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80,* 424-436.

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

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training.* Englewood Cliffs, NJ: Educational Technology Publications.

CHAPTER 3 - A closer look at the modality effect in multimedia learning*

Abstract

Both Mayer (2001) and Sweller (1999) claim that replacing visual text with spoken text in multimedia instructions decreases working memory load and improves learning. This modality effect was tested within the domain of instructional design (Experiment 1) and with learner-pacing (Experiment 2). In Experiment 1, the participants studied an audio or a visual-text version of a system-paced multimedia lesson. They rated their mental effort and made retention and transfer tests. The audio group reported less effort than the visual text group. In Experiment 2, two extra learner-paced versions were created. The audio group showed higher test scores with system-paced instructions, but not with learner-paced instructions. Thus, the modality effect in multimedia only applies with system-paced instructions.

Watching an animation on how lightning develops, studying the functioning of an electrical circuit from a colourful picture, or looking at an array of geometrical figures to learn about angles: learning with multimedia can be a joyful and effective experience. At least, if the guidelines are applied that follow from the empirical research by Mayer and his colleagues on multimedia learning (Mayer, 2001) and from the research by Sweller and others on cognitive load theory (Sweller, 1988, 1999; Sweller, Van Merriënboer & Paas, 1998). One common finding by both Mayer and Sweller is that learning with multimedia can be more effective if the text that accompanies an animation or a picture is presented as spoken word, rather than visual text.

The explanation both Mayer and Sweller give for the superiority of spoken text over visual text in multimedia learning is mainly based on the assumption that working memory has a limited capacity and has two modality-specific subsystems. The issue of limited memory resources has a long history that traces back to the influential article by Miller (1956) on the processing span of the human mind. Recent theories of working memory still stress the limited capacity view (Miyake & Shah, 1999). Furthermore, most contemporary theories presuppose separate working memory resources for different input modalities. In the multiple-components theory developed by Baddeley (1992), working memory consists of a central executive and two modality-specific slave systems, the *visuospatial sketchpad* and the *phonological loop*. The visuospatial

* based on: Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2001). The modality effect in multimedia instructions. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 1024-1029). Mahwah, NJ: Erlbaum.

sketchpad is the main entrance for visual and spatial information, whereas acoustic and verbal information is processed in the phonological loop.

These two characteristics of working memory, a limited capacity and modality-specific subsystems, have some important implications for multimedia learning, according to Mayer and Sweller. First, both researchers claim that multimedia instructions consisting of verbal and pictorial information, like for example a picture of a machine and a text about its functioning, place a high demand on the limited memory resources, because the learner has to switch attention between text and picture in order to integrate them mentally. Second, when information is presented in two sensory modalities (visual and auditory) rather than one, this memory load will be reduced. According to Sweller (1999), both slave systems are addressed instead of only one, so that total working memory capacity is used more efficiently. Relative to the available resources, the memory load of the multimedia instructions is reduced, leaving more space for the actual learning process. Mayer (2001) employs quite a similar explanation, called the dual-channel assumption, and claims that visual information is processed separately from auditory information. When words are presented as visual text, the visual channel will be easily overloaded as it is not only used for processing the picture but also, at least initially, for processing the text. This overload can be prevented by presenting the text as a narration so that both the visual and the auditory channel are used. So both Mayer and Sweller conclude that replacing visual text with spoken text in multimedia instructions will reduce the working memory load and improve learning.

Sweller and Mayer have demonstrated the modality effect in multimedia learning in a number of experiments. Jeung, Chandler and Sweller (1997) and Mousavi, Low and Sweller (1995) showed that, compared to students receiving visual-text instructions, students receiving multimedia instructions with spoken text spent less time on subsequent problem solving. Furthermore, in studies by Mayer and Moreno (1998; Moreno & Mayer, 1999) and by Kalyuga, Chandler and Sweller (1999, 2000), students receiving spoken text instructions had higher scores on various retention and transfer tests, and in experiments by Tindall-Ford, Chandler and Sweller (1997) students not only obtained higher test scores but also reported less mental effort during the instructions. Together, these results strongly support the design guideline to use spoken text in multimedia instructions.

However, one limitation of the above-mentioned studies is that they all used short multimedia instructions of only a few minutes long, on subjects from the exact sciences like geometry and electrical engineering. What influence does this have on the strength of the modality effect in multimedia learning? Will the effect still be obtained if the subject matter is from another domain, where learners have to study the multimedia material for a longer period of time? Theoretically, the modality effect should also apply in these circumstances, but empirically, this has not been tested yet. This issue is dealt with in Experiment 1, in which the modality effect is tested with a multimedia lesson on instructional design.

A more fundamental issue that can be raised given the evidence so far is the question whether the superiority of spoken text over visual text is really the result of a more efficient use of working memory capacity. In the working memory model of Baddeley (1992), all verbal material is processed in the phonological loop. The only difference is that spoken text will produce an extra acoustic memory trace that can improve the recall of auditorily presented words (Penney, 1989). So based on Baddeley's model, replacing visual text with spoken text in multimedia instructions will not necessarily increase the available working memory resources.

Moreover, the superiority of spoken text over visual text in multimedia learning found in previous studies can partly be explained as the result of a reduction in visual search needed to relate text and picture, and partly as the result of a more efficient use of time. Jeung et al. (1997), Mousavi et al. (1995) and Tindall-Ford et al. (1997) used multimedia instructions in which the complete explanatory text was printed next to a picture and compared it to instructions in which the students saw the picture and could listen to the explanation simultaneously. This way, they not only replaced visual text with spoken text, but also reduced the visual search needed to link the right parts of the text with the right parts of the picture. So in their experiments, the superiority of spoken text over visual-text instructions can also be attributed to the reduction in visual search.

Mayer and Moreno (1998; Moreno & Mayer, 1999) and Kalyuga et al. (1999, 2000), minimised the visual search in their instructions by presenting the text accompanying an animation in smaller parts. However, their instructions were system-paced and the time a student could read each part of the text that accompanied an animation was limited, based on the time of the narration. In the audio versions, learners could use this time more efficiently because they could look at the animation and listen to the narration at the same time. So the superiority of spoken text over visual text found in the experiments by Mayer and Moreno and Kalyuga et al. can also be explained as a more efficient use of time by using two senses rather than one.

Moreno and Mayer (1999) tried to rule out this explanation by setting up an experiment in which the animation and the text were presented sequentially. Still, spoken text proved to be more effective than visual text, which was a replication of the results of two similar experiments by Mousavi et al. (1995). Moreno and Mayer concluded that time does not seem to play an essential role in the modality effect in multimedia learning. However, we think the results obtained with sequential presentation of text and animation cannot be generalised that easily to the situation in which text and animation are presented at the same time. The superiority of spoken text over visual text in sequential presentation might also be the result of the better recall of spoken text due to the preservation of the extra acoustical memory trace. On the other hand, if spoken text and animation are presented simultaneously, integration can take place immediately and better recall of the text is not necessarily beneficial.

If the modality effect in system-paced instructions is merely the result of a more efficient use of time, spoken text will only be superior to visual text if the pacing of the instructions is based on the time of the narration. That implies that if more time is given to the learners, for example by giving them control over the pacing of the instructions, the superiority of spoken text over visual text will be less strong or even disappear completely. The beneficial effects of introducing learner-pacing in multimedia learning have already been demonstrated by Mayer and Chandler (2001), who showed that adding some simple user interaction to a multimedia animation did improve transfer performance. We think that learner-pacing will be especially helpful with a visual-text version of multimedia instructions, because the students will have more time to switch between text and picture to integrate them. This issue is dealt with in Experiment 2, in which the modality effect is tested with either system-paced or learner-paced instructions.

Experiment 1

The aim of the first experiment was to replicate the modality effect in multimedia learning using longer multimedia instructions on a subject not related to the exact sciences. For this purpose we developed a system-paced multimedia task about an instructional design model. The material mainly consisted of worked-out examples of training development sequences that were presented as a series of diagrams with explanatory texts. We developed two versions of our instructions, one with the explanatory texts presented on-screen (*visual text version*) and one with the texts presented as spoken word (*audio version*). Furthermore, we added a colour coding to the diagrams to relate each piece of explanatory text to the right part of the diagrams, to minimise visual search.

The hypothesis that follows from Sweller's cognitive load theory and Mayer's work on multimedia learning is that presenting the texts accompanying the diagrams as spoken text will decrease the memory load of the instructions and lead to a better performance. To measure this improvement in performance we administered a retention test and a transfer test after the instructions. Furthermore, to estimate the difference in memory load of the instructions, we used a subjective measure of mental effort, developed by Paas (1992). This measure was also used in previous research on the modality effect by Kalyuga et al. (1999, 2000) and Tindall-Ford et al. (1997). We also measured the mental effort spent on the tests, to relate the performance on the tests not only to the effort spent on the instructions but also to the effort spent on the tests. Paas and Van Merriënboer (1994) argued that mental effort is just one dimension of memory load that is not only influenced by task-characteristics but also by subject characteristics like prior knowledge and subject-task interactions like motivation. These effects were excluded by randomisation of our participants over the experimental groups.

Method

Participants

The participants were 41 students from a Teacher Training College for Primary Education in Heerlen, the Netherlands (20 second-years and 21 third-years; age between 18 and 24; 11 males and 30 females). They had applied on a voluntary base and were paid forty Dutch guilders for their participation. The students had not received any lessons on instructional design models yet, so they did not have any relevant prior knowledge. Twenty students were randomly assigned to the visual-text group that received the visual-text version of the instructions and 21 to the audio group that received the audio version of the instructions.

Materials

The multimedia instructions used in the experiment were developed as a web-based application, and dealt with the Four Component Instructional Design (4C/ID) model of Van Merriënboer (1997). This model describes a design strategy for the training of complex cognitive skills. The instructions focused on the question how to develop a blueprint for a training programme based on the skills hierarchy of a complex skill. The instructions started with a very short textual introduction to the model. Subsequently, the design strategy of the 4C/ID model was demonstrated in a series of 11 diagrams representing skills hierarchies and elaborated sequences of learning tasks. Together, these diagrams formed two worked-out examples and a general explanation of the strategy. The first example consisted of six diagrams that showed the different stages in developing a blueprint of a training programme for the complex skill *doing experimental research*. The second worked-out example consisted of three diagrams showing the same process for the complex skill *designing a house*, and finally the general strategy of the 4C/ID model was explained in the last two diagrams. All diagrams were accompanied by a textual explanation on how the model was applied in the specific situation. These explanatory texts were split into smaller pieces of only one or two sentences, in such a way that each piece of text referred to a specific part of the diagram. These parts in the diagram were coloured bright red.

Two versions of the instructions were created that differed in the way the texts accompanying the diagrams were presented (see Figure 1 for screen captions). In the audio version, the text that accompanied a diagram was presented as spoken text. Three seconds after an audio fragment stopped playing, the colour coding in the diagram changed and the next piece of audio started playing automatically. In the visual text version the text fragments were depicted right above the diagrams. After exactly the same period of time as in the audio version, the colour coding in the diagram changed and a new piece of text appeared above the diagram. So only the accompanying text and the colour coding changed, not the diagram itself. The total presentation time of all 11 diagrams was 26.2 minutes.

The self-report measure of memory load was a 9-point rating scale on which the students could rate the mental effort they had spent on the instructions, ranging from *very, very low mental effort* to *very, very high mental effort*. The average score on the rating scales presented after each diagram was taken as a measure of mental effort during instructions (Cronbach's alpha = .90).

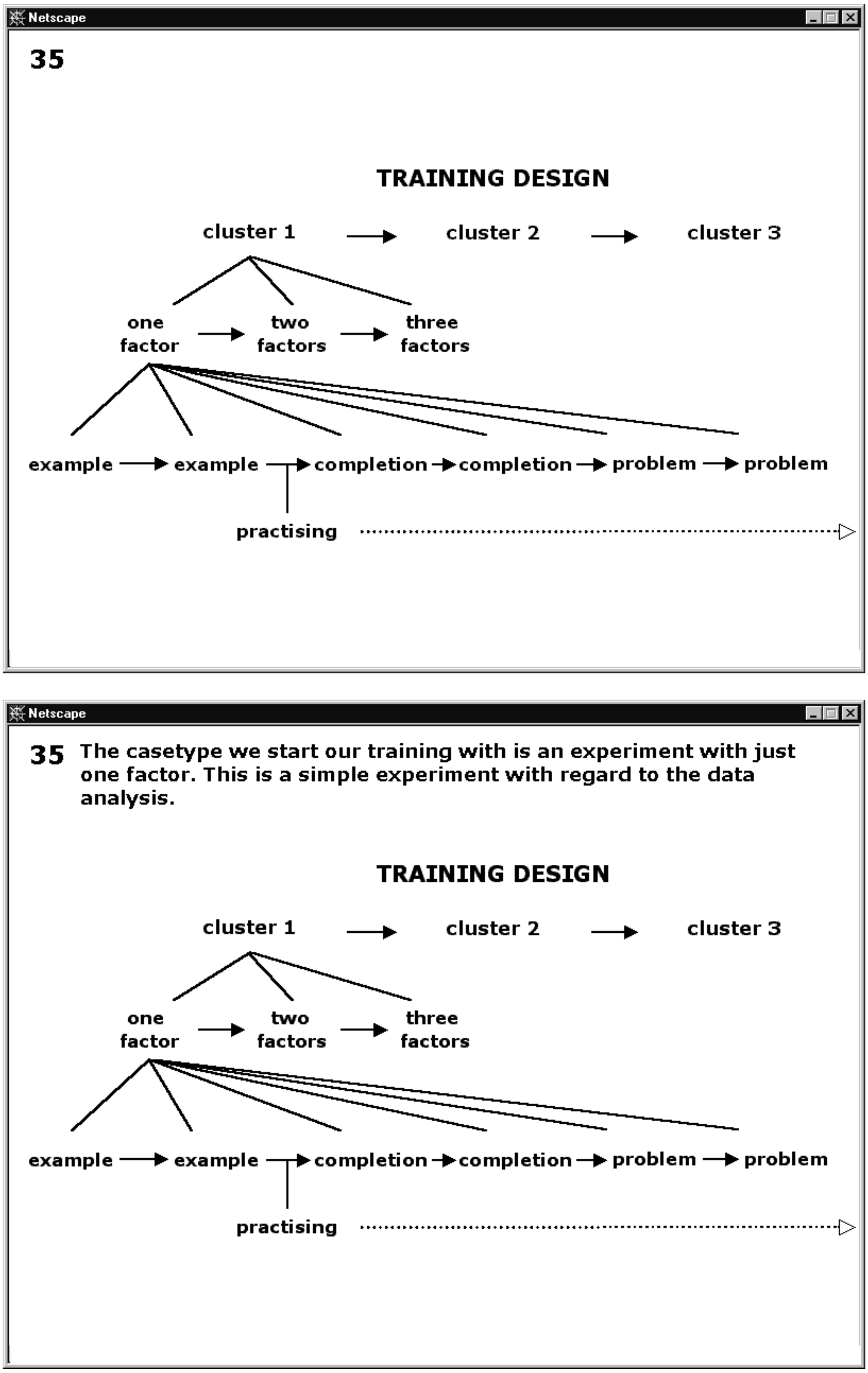


Figure 1

Screen examples of the multimedia instructions used in Experiment 1 (translated from the original Dutch version). In the audio version (on top) the diagram is accompanied by an audio fragment of the explanatory text that starts playing automatically.

The retention test consisted of two paper-and-pencil tests, one of 30 and one of 20 multiple-choice items. The 30-item test contained only verbal statements, while the 20-item test combined verbal statements with small parts of diagrams. All items were statements about the 4C/ID model like “A macro-sequence in the 4C/ID model is a series of subskills in a cluster”, or “According to the 4C/ID model, the same subskills can be trained in more than one learning task”, and the students could choose between *correct*, *incorrect* or *I don't know*. The retention score was calculated by taking the sum of right answers on all 50 items (Cronbach's alpha = .74).

The transfer test was a paper-and pencil test that contained a short description of the skills expert researchers apply when doing a literature search. The assignment was to design a blueprint of a training programme for this complex skill following the strategy of the 4C/ID model. The students had to construct a skills hierarchy, and design a sequence of learning tasks based on this hierarchy. To be able to score the results of the transfer test, a separate scoring form was developed consisting of 28 *yes/no*-questions that checked to what extent and how accurately the strategy prescribed by the model had been applied in the transfer task. The transfer score ranged from zero (no steps from the model taken) to 28 (all steps taken accurately). After the experiment, two independent raters scored the transfer tests using the form, showing an inter-rater agreement on the sum score of .95 (intraclass correlation coefficient). The average score of both raters was taken as the transfer score.

Procedure

The experiment was carried out at the Open University in Heerlen, the Netherlands, in eight sessions of two hours. In each session, between one and seven students were tested simultaneously. These sessions took place in a room that had seven computers connected to the server on which the instructional web site was installed. Three computers had headphones attached to them and were connected to the audio version of the instructions, whereas the other four computers were connected to the visual-text version. When the students entered the room they were randomly assigned to one of the computers. Each computer showed a browser-window, without any of the menu options visible, that was set on the introduction page with some general information about the experiment. All students started at the same time and studied the instructions by themselves. After each diagram, a separate page followed with the mental effort scale. When a student clicked on one of the nine options, the program automatically continued with the next diagram. The web server recorded the mental effort scores of each participant.

After the instruction phase the three paper-and-pencil tests were administered. First, the students received the 30-item multiple-choice test, which had to be completed in ten minutes, followed by the 20-item test, which also had to be completed in ten minutes, and finally they received the transfer test that had to be completed within 30 minutes. Together with the transfer test, the students got a blank answering form on which they could draw their design and write their comments. After each test a rating scale had to be completed as a measure of the mental effort spent on the tests.

Results

The variables under analysis were retention score, transfer score, mental effort spent on the instructions, on the retention test, and on the transfer test. All scores were analysed with one-tailed t-tests, and for all statistical tests, a significance level of .05 was applied. Table 1 shows the average scores on the dependent measures for both experimental groups (both retention and transfer score are reported as percentages).

Table 1

Mean Scores on Dependent Measures for both groups in Experiment 1

	Audio		Visual-Text	
	M	*SD*	*M*	*SD*
Retention Score (%)	63	12	60	11
Transfer Score (%)	34	22	37	19
Mental Effort Instructions	4.3	0.8	4.9	0.9
Mental Effort Retention Tests	6.2	0.8	6.4	1.2
Mental Effort Transfer Test	6.4	1.4	7.1	1.1

The main hypothesis was that the students in the audio group would obtain higher retention and transfer scores and report less mental effort. Although the audio group did a little better than the visual-text group on the retention test, this effect was not statistically significant, $t(39) = 0.88$, $p > .10$. Also no significant difference was found between the groups on the transfer score, $t(39) = -0.40$, $p > .10$.

The mean score on mental effort during instructions was 4.6 (on a scale from 1 to 9) and showed a significant effect for the modality of text, $t(39) = 2.19$, $p = .02$. Students in the audio group reported less effort on the instructions than their colleagues in the visual-text group. On the mental effort spent on the retention test no significant differences were found between the two groups, $t(39) = 0.53$, $p > .10$. The mental effort scores in the transfer test did show a significant difference between groups, $t(39) = 1.85$, $p = .04$, with the students in the audio group spending again less effort than their colleagues in the visual-text groups.

Discussion

The results of Experiment 1 show that the modality effect can at least be partly replicated with a longer multimedia lesson in a domain like instructional design. Students in the audio group report lower mental effort scores than students from the visual-text group, indicating that replacing visual text with spoken text indeed results in a decrease in memory load. Both groups score equally well on the retention and transfer test, so obtaining the same learning results has cost the students in the audio group less mental effort.

The modality effect is only found in terms of mental efficiency (getting the same result with less effort) and not in terms of effectiveness (getting a better learning result), whereas previous researchers like Tindall-Ford et al. (1997) found both lower mental effort scores and better test results for the audio instructions. This lack of a modality effect in learning outcomes in Experiment 1 might be explained by the low scores on especially the transfer test, indicating that this test was too difficult for our participants. Nevertheless, the fact that students in the audio group obtained the same test results as the students in the visual-text group with less mental effort still points at a superiority of spoken text over visual text for system-paced multimedia instructions, even with instructions of a longer duration on the subject of instructional design.

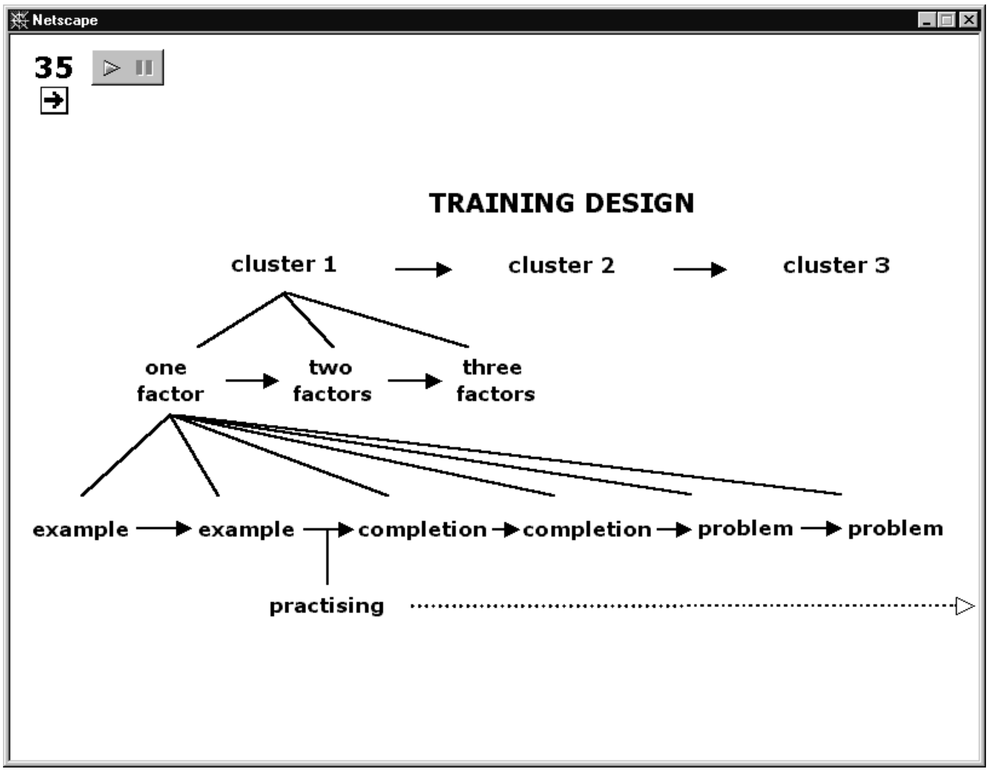
Experiment 2

In the second experiment we wanted to investigate if the modality effect in multimedia learning can still be found if students have the control over the pacing of the instructions. Therefore we not only varied the modality of the text (spoken text versus visual text), but also the pacing of the instructions (system-paced versus learner-paced). The system-paced versions of the instructions were identical to the two versions used in Experiment 1, whereas two extra learner-paced versions of the instructions were created in which students could set the pace of the instructions for themselves. This way we compared four experimental conditions: *audio-system*, *visual-system*, *audio-learner*, and *visual-learner*. If we assume that the modality effect in multimedia learning is mainly the result of a more efficient use of time in the audio condition, the effect will disappear when the learners are given control over the pacing of the instructions. So our main hypothesis is that spoken text will be superior to visual text in the system-paced groups, but that this difference will disappear in the learner-paced groups. Thus, an interaction between modality and pacing is expected.

Method

Participants

The participants were 130 second-year students from the Department of Education of the University of Gent in Belgium (age between 18 and 31 years; 12 males and 118 females). The experiment was part of a regular course on instructional design, but at the time of the experiment the students had not received any lessons on instructional design models yet. Thirty participants were randomly assigned to the audio-learner group, another 29 to the audio-system group, 37 to the visual-learner group, and 34 to the visual-system group.



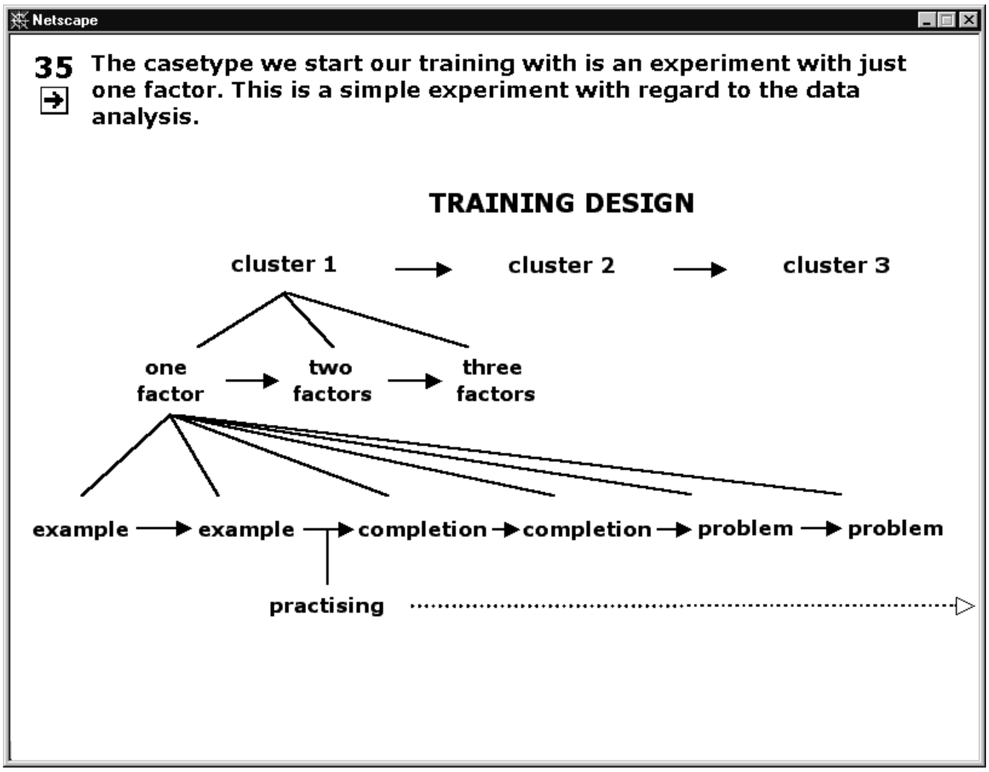


Figure 2
Screen examples of the learner-paced multimedia instructions used in Experiment 2 (translated from the original Dutch version). In the audio-learner version (on top) the diagram is accompanied by an audio fragment of the explanatory text. The audio starts playing automatically, and can be replayed by clicking on the little play-button in the upper-left corner.

Materials
The multimedia instructions were the same as in the Experiment 1, except that this time two extra learner-paced versions were created (see Figure 2 for screen examples). In these versions, a forward arrow was added that could be clicked to get to the next text fragment. In the audio-learner version, each text fragment could be replayed by clicking on a small play-button.

The measurements were mostly the same as in Experiment 1. The training time in the learner-paced conditions was calculated by taking the total time spent on the diagrams in the instructions (so not on the mental effort scales or on the introductory text). The average score on the 11 mental effort scales was again taken as a measure of mental effort during instructions (Cronbach's alpha = .92). The retention test was computer-based this time, and consisted of 40 items taken from the retention test of Experiment 1 (Cronbach's alpha = .64). The transfer test was still paper-based. After the experiment, two independent raters scored the transfer tests using the same scoring form as in Experiment 1, showing an inter-rater agreement of .90. The average score of both raters was taken as the transfer score.

Procedure
The experiment was carried out in seven sessions of about two-and-a-half hours, and in each session between 15 and 24 students were tested simultaneously. These sessions took place in a classroom that had 24 multimedia computers connected to the server on which the instructional web site was installed, with six computers for each experimental group. The procedure was identical to Experiment 1. However, this time the students could continue with the retention test on the computer screen whenever they had finished the instructions. The server kept record of the time spent on the instructions (in minutes) and of the mental effort scores of each participant.

Results

Unfortunately, in the first three sessions, the server loggings of the time-on-task and mental effort scores were lost, so that we can only report the effort and time data of 81 students (18 students in the audio-learner group, 18 in the audio-system group, 24 in the visual-learner group, and 21 in the visual-system group).

The variables under analysis were training time, mental effort spent on the instructions, on the retention test and on the transfer test (N = 81), and retention score and transfer score (N = 130). Except for training time, all scores were analysed with two-factor analyses of variance (ANOVAs), with modality (spoken text vs. visual text) and pacing of the instructions (system pacing vs. learner pacing) as the between-subjects factors. For all statistical tests, a significance level of .05 was applied. Table 2 shows the average scores on the dependent measures for all four groups.

Table 2
Mean Scores on Dependent Measures for all groups in Experiment 2

	Audio-Learner		Audio-System		Visual-Learner		Visual-System	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Training time (minutes)	28.3	3.3	26.2	0	30.9	5.0	26.2	0
Mental Effort Instructions (1-9)	4.3	1.0	4.1	0.7	4.0	1.0	4.2	1.0
Mental Effort Retention Test (1-9)	6.5	1.3	6.7	1.0	6.5	1.3	6.7	1.0
Mental Effort Transfer Test (1-9)	7.3	1.1	7.3	1.5	7.5	1.1	7.1	1.5
Retention Score (%)	68	12	72	8	72	8	67	13
Transfer Score (%)	63	14	63	14	64	18	51	18

With regard to training time, only the two learner-paced groups were compared, because in the system-paced groups the time spent on studying the diagrams was equal for all students. On average, the learner-paced groups spent a few more minutes on the diagrams than the system-paced groups (29.8 and 26.2 minutes, respectively). The difference between the students in the visual-learner group and the audio-learner group did not reach statistical significance ($t(40) = 1.85$, $p > .05$, two-tailed).

Comparing all four groups, no main effect of modality or pacing on mental effort during instructions was found, and the interaction was also not significant, $MSE = 0.93$, all Fs < 1. The same goes for the mental effort spent on the retention test, $MSE = 1.36$, all Fs < 1, and for the mental effort spent on the transfer test, $MSE = 1.72$, all Fs < 1.

There were also no main effects of modality or pacing on retention score, $MSE = 18.17$, all Fs < 1. However, the interaction of modality and pacing was significant, $F(1,126) = 6.76$, $p = .01$. In the two system-paced groups, audio did better than visual text, whereas in the learner-paced groups this effect was turned around, with the visual-learner group outperforming the audio-learner group (see Figure 3). Pairwise comparisons within each level of pacing showed only a significant effect of modality in the system-paced groups, $F(1,126) = 4.10$, $p < .05$.

Although students in the audio groups did a little better on the transfer test than the students in the visual-text groups (63% versus 58%), this difference did not reach statistical significance, $F(1,126) = 3.39$, $MSE = 21.00$, $p = .07$. The effect of pacing, however, was significant, $F(1,126) = 5.02$, $p = .03$, with the students in the learner-paced groups doing better than the students in the system-paced groups (64% versus 57%). Moreover, the interaction of modality and pacing was significant, $F(1,126) = 4.95$, $p = .03$. Visual inspection of the separate group means shows a superiority of audio over visual text in the system-paced groups, but not in the learner-paced groups (see Figure 3). A pairwise comparison within each level of pacing showed a significant effect of modality in the system-paced groups, $F(1,126) = 8.04$, $p < .01$, but not in the learner-paced groups.

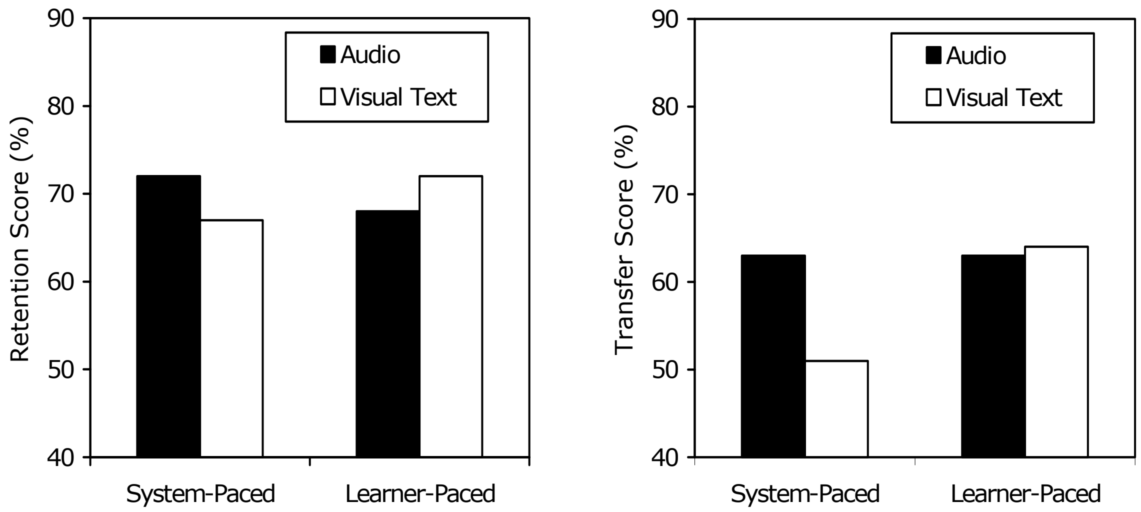


Figure 3
Mean retention and transfer scores (as percentages) for all groups in Experiment 2.

Discussion

The results show that we find a modality effect in the two system-paced groups. Students receiving the audio instructions have higher retention and transfer scores than students receiving the visual-text instructions. However, in the two learner-paced groups in which the students set the pace of the instruction, the superiority of spoken text over visual text completely disappears. Not only do students in the visual-learner group perform just as well as the students in the audio groups on the transfer test, on the retention test they even slightly outperform the students in the audio-learner group.

On average, the students in both learner-paced groups have taken a few more minutes to study the multimedia instructions, which confirms our hypothesis that the modality effect in the system-paced groups is mainly the result of a lack of time to relate the text to the diagrams in the visual-system group. When the students in the visual text group can set the pace themselves, they have more time to relate the text to the diagram and, as a result, perform equally well as the students in the audio groups. The introduction of learner-pacing in the audio group in our experiment does not seem to be beneficial at all, contrary to what Mayer and Chandler (2001) found. However, in their experiment the multimedia instructions were presented twice and students could not replay the audio, which makes it difficult to compare their results with our findings. Nevertheless, this issue needs to be investigated in further detail.

General discussion

The results of both experiments show that a superiority of spoken text over visual text in multimedia learning is found when the student has no control over the pacing of the instruction and the pacing is set by the time of the narration. In that case we find either lower mental effort scores or better test results, also with longer instructions on a subject matter like instructional design. However, if students have the opportunity to determine the pace of the instructions themselves, visual-text instructions are just as effective as audio instructions.

It should be noted that the modality effect obtained in the system-paced groups of Experiment 2 in terms of higher retention and transfer scores with equal effort differs from the results in Experiment 1 in which the students in the audio group spend less mental effort but do not have better test scores. This might be related to the differences between the samples. In Experiment 1, the students from a Teacher Training College applied on a voluntary base and were paid for their participation, whereas the students of educational science took part in Experiment 2 as part a regular course. The students in the first experiment found the tests very difficult, resulting in low overall scores and no differences between the conditions. The educational students on the other hand were more able and maybe even more motivated to invest effort in the learning process itself, resulting in higher test scores with equal effort, comparable to the results of Kalyuga et al. (1999, 2000).

From a theoretical point of view, the results indicate that the modality effect in multimedia learning as demonstrated in earlier experiments can be accounted for in other terms than an increase in available memory resources. Earlier research on verbal recall has shown that presenting the text as spoken text or as visual text influences the way the information is processed in working memory, as it results in slightly different memory traces (Penney, 1989). The main reason, however, why spoken text is superior to visual text in multimedia learning does not seem to be the extra acoustic memory trace created in the phonological store, but the fact that the use of two senses makes it possible to perceive both text and picture at the same time. This way the most optimal temporal contiguity is achieved. Text and picture can be integrated immediately without the necessity to keep one of the elements active in working memory while relating it to the corresponding element. It seems that the extra time needed to switch between verbal and pictorial information is mainly responsible for the detrimental effect of presenting text visually instead of auditorily. After all, our results show that giving students control over the pacing of the instructions can undo this negative effect and make visual text just as effective as spoken text.

From a practical point of view it is interesting to see that the superiority of spoken text vanishes into thin air when the multimedia instructions are not system-paced anymore. With learner-pacing it does not seem to matter much if spoken text or visual text is used, which has some important consequences for multimedia design. In most cases, spoken text will not be the first choice of the designer when presenting text. First of all, spoken text is more expensive to produce than printed or on-screen text. Moreover, presenting spoken text requires extra resources for the learner. Equipment is needed that can handle audio, and headphones or speakers. Finally, visual text can be more easily reread or skipped through. Especially when the designer has no clear indication of the level of prior knowledge of the learner, it might be wiser to present the text visually, as redundant visual text is easier to neglect than redundant spoken text. Kalyuga et al. (2000) already showed that with increasing expertise, the audio version of multimedia instructions became less effective. This implies that with learner-paced multimedia instructions, a choice for printed or on-screen text seems to be the most sensible one.

Taken together, the results from our study indicate that on the one hand the modality effect in multimedia learning does apply to other subject areas and longer instructions, but on the other hand it is restricted to system-paced multimedia instructions. It also shows that introducing user-interaction has an influence on the applicability of design guidelines for multimedia learning. This means that more research is needed that investigates how different kinds of user-interaction interact with the way people process multimedia material. Only in that way will we be able to extend the guidelines for designing multimedia to a wider range of educational applications than just simple system-paced instructions.

References

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

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

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

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92,* 126-136.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93,* 390-397.

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,* 312-320.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Miyake, A., & Shah, P. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake & P. Shah (Eds.), *Models of working* *memory: mechanisms of active maintenance and executive control* (pp. 442-481). New York: Cambridge University Press.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

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

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84,* 429-434.

Paas, F., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6,* 351-371.

Penney, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17,* 398-422.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-296.

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

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training.* Englewood Cliffs, NJ: Educational Technology Publications.

CHAPTER 4 - The interaction of modality with pacing in multimedia learning*

Abstract

Six versions of a multimedia lesson on instructional design were compared differing in modality (visual text vs. spoken text) and pacing (system-paced, system-paced with extended time-on-task, learner-paced). Lower mental effort was expected in all conditions with spoken text; higher transfer was expected only for spoken text in the system-paced conditions. Ninety-four second-year students of educational science got the lesson, made a retention and transfer test and rated their mental effort. As hypothesised, the spoken-text groups reported less effort than the visual-text groups. A significant interaction between modality and pacing was found on transfer, indicating an advantage of spoken text over visual text in the system-paced conditions, no differences in the extended conditions, and a reverse pattern in the learner-paced conditions.

When a student watches an animation on a computer screen and at the same time tries to read the accompanying on-screen comment, multimedia learning is taking place. The distinguishing features of multimedia learning are that the learner receives information in more than one presentation mode, and that the information elements have to be mentally integrated in order to be understood (Mayer, 2001). This integration process takes place in working memory, which is also the place where new knowledge is constructed on the basis of both the presented information and the existing knowledge structures in long-term memory. Not surprisingly, any prior knowledge on the topic of study in the form of well-developed schemata will decrease the burden on working memory and facilitate the learning process. However, if information is new and complex, the effort needed for the integration process can easily overload the limited resources of working memory.

This, at least, is one of the central ideas in the work of Sweller on cognitive load theory (Sweller, 1999; Sweller, Van Merriënboer & Paas, 1998). Based on the generally accepted notion of a working memory with limited capacity, Sweller and his co-workers have developed several guidelines for instructional design that for a large part aim at preventing cognitive overload. One of these guidelines is especially applicable to multimedia instructions and is based on the so-called *modality effect*, which is the repeatedly found superiority of spoken text over visual text in multimedia learning. In several experiments, the use of spoken text instead of on-screen or written text has led to faster problem solving (Jeung, Chandler & Sweller, 1997; Mousavi, Low & Sweller, 1995), higher test

* based on: Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2002). *The interaction of modality with pacing in multimedia learning*. Manuscript submitted for publication.

scores (Kalyuga, Chandler & Sweller, 1999, 2000; Mayer & Moreno, 1998; Moreno & Mayer, 1999) and less mental effort reported by the learners (Tabbers, Martens & Van Merriënboer, 2001; Tindall-Ford, Chandler & Sweller, 1997; Van Gerven, Paas, Van Merriënboer & Schmidt, 2002).

In this article, we question the theoretical rationale that is usually given for the modality effect in multimedia learning and argue that it can better be accounted for as the result of preventing split attention. Moreover, the lack of empirical evidence for a high cognitive load in multimedia learning is discussed and an alternative explanation for the repeatedly found superiority of spoken text is suggested by taking into account that the instructions used in previous studies were mostly system-paced. Finally, an experiment is reported in which we test our assumptions on the interaction of modality with pacing in multimedia learning.

The modality effect as a result of preventing split attention

The explanation Sweller et al. (1998) give for the modality effect in multimedia learning is mainly based on the multimodal working memory model of Baddeley (1992). In this model, working memory not only has a central executive part but also two modality-specific slave systems, the *visuo-spatial sketchpad* and the *phonological loop*. It is the latter that is responsible for processing all speech-based information. Sweller argues that presenting information in two modalities rather than one increases the effective memory resources, because the phonological loop is used directly for processing the auditory information. Mayer (2001) extends Sweller's argument by claiming that auditorily and visually presented information take different routes or *channels* into working memory, and that the combination of visual text and picture will easily overload the visual channel. Replacing visual text with spoken text will reduce the burden on the visual channel and thus improve learning.

So the combined rationale for the modality effect in multimedia learning is that spoken text takes a different route into working memory and is processed in a separate subsystem. This way, the available working memory resources are used more efficiently, so that cognitive overload is prevented and learning can be improved. This explanation leads to the guideline that an instructional designer should always use spoken text rather than visual text in multimedia instructions, especially when information is new or complex. Although this guideline is supported with ample empirical evidence, we question the assumptions underlying the explanation of the modality effect and would like to argue that the modality effect in multimedia learning is not the result of a more efficient use of working memory capacity preventing cognitive overload, but rather the result of preventing split attention.

First of all, the assumption that the use of spoken text instead of visual text in multimedia instructions will increase working memory resources is not in line with the working memory model of Baddeley (1992, 1997). In his model, visually presented words that enter working memory are converted into a phonological code through silent articulation and are processed in the phonological loop, just like spoken words. Only when this articulation process is disrupted, by either overt or covert articulation of an irrelevant item, will no phonological code be produced. So replacing visual text with audio in multimedia instructions will not necessarily increase working memory resources, because in both cases the phonological loop is addressed. The only difference between spoken text and visual text is that the former will produce an extra acoustic memory trace in the *phonological store*, a subsystem of the phonological loop, which can explain the modality effect found in verbal recall studies (Penney, 1989). It is questionable if this weak acoustic trace can also account for the modality effect in multimedia learning.

Second, the assumption that the combination of visual text and picture will easily overload the visual channel can be refuted on the same grounds. In multimedia learning, spoken text and picture can be perceived simultaneously, because attention can be directed to text and picture at the same time. Visual text and picture, however, cannot be perceived simultaneously, because learners have to split their attention between the two information sources. Eye movement research on the integration of pictures and text has shown that in most cases learners first read (part of) the text and then switch to the picture to integrate the verbal and the pictorial information (Hegarty, 1992; Rayner, Rotello, Stewart, Keir & Duffy, 2001). Because the visual text will be converted into phonological code almost immediately, the subsequent processing of the picture in the visual channel will not be hindered. So in that case, replacing visual text with spoken text will not decrease the load in the visual channel.

To sum up, we do not think it very likely that the modality effect in multimedia learning is the result of either an increase in working memory resources or a capacity conflict in the visual channel. That means an alternative account is needed because previous studies did find lower mental effort scores and better learning results when replacing written or on-screen text with spoken text in multimedia instructions. This explanation can be found in the work of Sweller and Mayer on the *split-attention effect* or *spatial contiguity effect* in multimedia learning, which is the finding that presenting picture and text near each other resulted in less cognitive load and better learning results than presenting picture and text far from each other (e.g., Chandler & Sweller, 1991; Mayer, Steinhoff, Bower & Mars, 1995; Sweller, Chandler, Tierney & Cooper, 1990). The explanation for this effect is that learners do not have to search for the text and the corresponding part of the picture when they are physically integrated (Mayer, 2001), and that as a result, learners need less mental effort to integrate both sources (Sweller, 1999).

We claim that the modality effect in multimedia learning can be based on the same rationale as the split-attention effect. Replacing visual text with spoken text results in a more efficient integration process because learners look at a picture or animation and listen to a narration at the same time and do not have to split their attention between both information sources. That makes the superiority of spoken text over visual text in multimedia learning the result of preventing split attention, rather than the result of an increase in working memory capacity. This idea is supported by the results of two studies. First, Tindall-Ford et al. (1997) showed in one experiment that a physically integrated visual-text format was just as effective as a spoken-text format, which is consistent with the view that both are based on the same principle of enabling a smooth integration of text and picture. Second, Jeung et al. (1997) did not obtain a modality effect in their study when the picture accompanying the spoken text was complex and visual search was high. Only by adding visual cues to the picture that related the spoken text to the appropriate parts of the picture, did they recover the modality effect. So if learners had to search a picture while listening to a narration, the advantage of spoken text over visual text disappeared. This is another indication that the superiority of spoken text in multimedia instructions might be mainly the result of an optimal integration of text and picture.

The explanation of the modality effect in multimedia learning in terms of preventing split attention can very well explain the results found in previous studies. With spoken text, the learner watches a picture or animation and listens to the narration at the same time. The text does not have to be actively represented in working memory in order to be integrated with the picture, assuming of course that the text is not too complex and that integration can take place immediately. With visual text, however, the learner will read the text and has to keep its representation active in working memory while searching for the relevant parts of the picture, increasing the working memory load. So the use of visual text instead of spoken text will lead to a higher cognitive load and inferior learning results, which is exactly what has been found in previous studies on the modality effect in multimedia learning.

The interaction of modality with pacing

Although explaining the modality effect in terms of preventing split-attention can account for the findings of previous studies, yet another explanation is needed for the superior learning results found with spoken text. One of the main ideas behind cognitive load theory is that a more efficient use of working memory resources is needed when there is a risk of overload (Sweller, 1999). So the use of visual text in multimedia instructions will only be harmful for learning when cognitive load is high. If not, the learner can compensate the increase in memory load by spending more mental effort and get the same learning outcomes as with spoken text. However, in a number of studies related to the modality effect in multimedia learning, researchers used subjective rating scales to estimate the mental effort invested by the learners and reported scores that were average or even below average (Kalyuga et al., 1999, 2000; Mayer & Chandler, 2001; Tabbers et al., 2001; Tindall-Ford et al., 1997; Van Gerven et al., 2002). Although these self-report scales cannot be seen as absolute measurements of working memory load, the assumed overload in the visual text conditions was certainly not reflected in the scores on the cognitive load measures. So the assumption that the modality effect in multimedia learning only applies with cognitive overload is not supported with empirical evidence.

That raises the question why the studies that did not find any cognitive overload still found better learning results with spoken text. We think an alternative explanation can be found by taking the pacing of the instructions into account. In most of the modality studies the time-on-task was limited because the pacing of the multimedia instructions was based on the pace of the narration, both in the visual-text and the spoken-text version. Because visual text and picture cannot be perceived simultaneously, the learners have to split their attention between the two information sources. That makes this presentation mode less efficient in terms of time-on-task than spoken text and picture that can be processed simultaneously. So the fact that the instructions in previous studies were system-paced, based on the pace of the narration, might have been the main reason why superior learning results were obtained with spoken text. It also implies that when learners have more time to integrate text and picture, visual text is just as effective as spoken text and the modality effect disappears, at least as long as no cognitive overload occurs.

This interaction of modality with pacing in multimedia learning was first demonstrated in a study in which system-paced instructions with a pacing based on the narration were compared with learner-paced instructions (Tabbers et al., 2001). The superiority of spoken text over visual text in terms of better learning results was only obtained with system-paced instructions and not with learner-paced instructions. On average, study time was longer in the learner-paced conditions, which seems to support our idea that time-on-task is a relevant factor in the modality effect. However, because of the individual differences in study time in the learner-paced conditions, the disappearance of the modality effect could not be unequivocally attributed to the extended time-on-task. For a better test of the interaction of modality and pacing, a more direct manipulation of time-on-task like slowing down the pacing of the instruction is needed.

We set up an experiment to test our hypothesis on the interaction of modality with pacing, using the same instructional material as Tabbers et al. (2001). This time, we not only compared system-paced instructions with a pacing based on the narration with learner-paced instructions, but also included a system-paced condition in which the pacing was slowed down altogether. This was accomplished by doubling the time of the instructions, either by playing each audio fragment twice or by showing each visual text fragment twice as long. This way, six different conditions were compared: system-paced audio (SA), system-paced visual-text (SV), double system-paced audio (2SA), double system-paced visual-text (2SV), learner-paced audio (LA), and learner-paced visual-text (LV). Our first hypothesis is that replacing visual text with spoken text in multimedia instructions will prevent split-attention and make the integration of pictures and text easier, which will lower the mental effort. Our second hypothesis is that replacing visual text with spoken text will make multimedia instructions more effective in terms of higher transfer of learning only if the time to integrate picture and text is limited and based on the pacing of the narration. Thus, an interaction of modality and pacing is expected.

Method

Participants

The participants were 94 second-year students from the Department of Education of the University of Gent in Belgium (age between 18 and 23 years; 8 males and 86 females). The experiment was part of a regular course on instructional design, but at the time of the experiment the students had not received any lessons on instructional design models yet. The participants were randomly assigned to one of the six conditions, so that each experimental group contained 16 students, except for the 2SA and the LA group that both contained 15 students.

Materials

Instructions

The multimedia lesson used in the experiment was developed as a web-based application, and dealt with the Four Component Instructional Design (4C/ID) model of Van Merriënboer (1997). This model describes an instructional design strategy for the training of complex cognitive skills. In the multimedia lesson an explanation was given of the procedure to be followed when developing a blueprint for a training programme based on the skills hierarchy of a complex skill. This procedure was first demonstrated with two worked-out examples and then summarised. The first worked-out example consisted of three diagrams representing skills hierarchies and elaborated sequences of learning tasks that showed the different stages in developing a training programme for the complex skill *doing experimental research*. The second worked-out example consisted of three diagrams showing the same procedure for the complex skill *designing a house*. In the end, the whole procedure was summarised in two diagrams. All diagrams were accompanied by a textual explanation on how the procedure was applied in the specific situation. These explanatory texts were split into smaller pieces of only one or two sentences, in such a way that each piece of text referred to a specific part of the diagram. Moreover, these parts in the diagram were coloured bright red to prevent any unnecessary visual search.

Six versions of the instructions were created that differed in the modality of the text accompanying the diagrams and in the pacing of the instructions. In the system-paced audio version, students could listen to the pieces of explanatory text that accompanied a diagram through a headphone. When the audio had finished playing, the colour-coding in the diagram changed and the next piece of audio started. In the system-paced visual-text version the pieces of explanatory text were depicted right above the diagrams (see Figure 1 for a screen example). After exactly the same period of time as in the AS1 condition, the colour-coding in the diagram changed and a new piece of text appeared above the diagram. So only the accompanying text and the colour coding changed, and not the diagram itself. The presentation of the eight diagrams took 19.3 minutes.

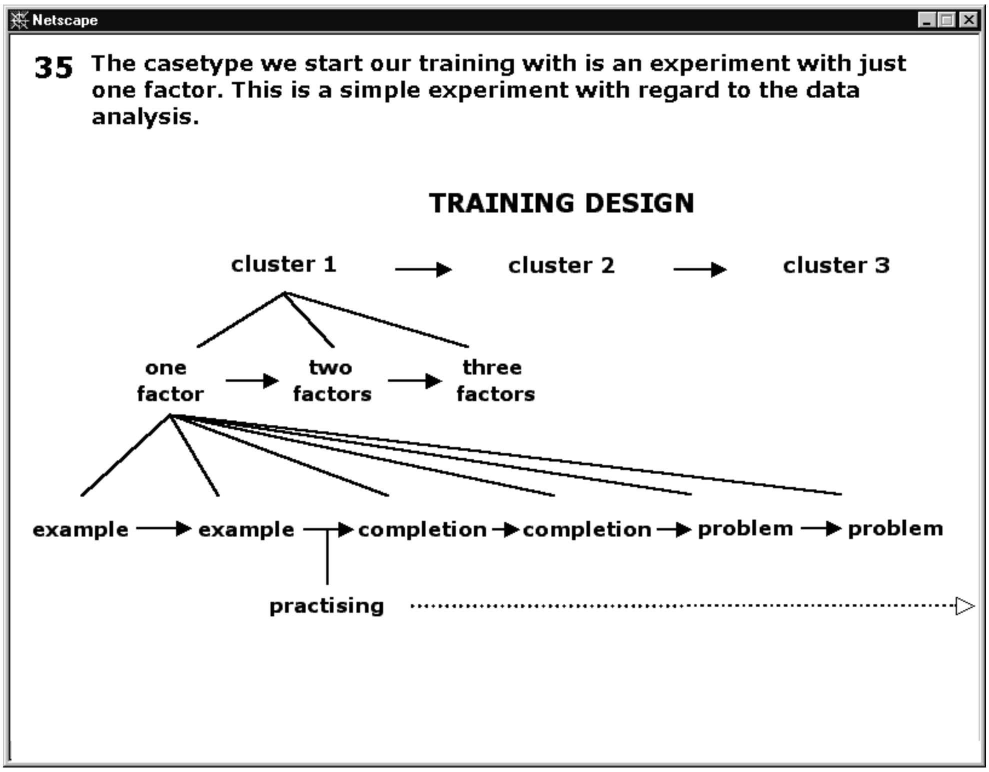


Figure 1
Screen example of the system-paced visual-text version of the multimedia instructions (translated from the original Dutch version). In the audio version the text accompanying the diagram was presented as spoken text and started playing automatically.

The two double system-paced versions of the instructions were identical to the normal system-paced versions, only this time each piece of audio was played twice in the audio variant and the time each piece of text was displayed was twice as long in the visual text variant. Thus, the total presentation time was doubled to 38.6 minutes.

In the learner-paced audio version, students were able to replay the sentences they had just heard by clicking on a small play-button, whereas in the learner-paced visual-text version students could reread the text as many times as they wanted to. To continue with the next piece of text students in both versions had to click on a forward arrow, so the presentation time of the eight diagrams was variable.

Mental effort scale

A 9-point rating scale was used on which the students could indicate the mental effort they had spent, ranging from *very, very low mental effort* to *very, very high mental effort*. This self-report measure was originally developed by Paas (1992), based on a measure of perceived task difficulty of Borg, Bratfisch, and Dornic (1971), and was also used in previous studies of multimedia learning (e.g., Kalyuga et al., 1999; Mayer & Chandler, 2001). The average score of the rating scales following the diagrams in the instructions was taken as a measure of mental effort during instructions (Cronbach's alpha = .90).

Retention test

The retention test consisted of 30 items about the 4C/ID model. Twenty-one items were statements on the general strategy of the model like “The training of each skill cluster starts with the most complex case type". The other nine items were statements about the two worked examples that were studied, sometimes in combination with a small part of the diagram, like "According to this skills hierarchy, an expert researcher will first formulate a hypothesis and then identify the relevant variables". The students had to evaluate the statements with either *correct*, *incorrect* or *I don't know*. Each right answer was awarded one point, a wrong answer minus one point and *I don’t know* zero points. After item analysis 7 items were removed from the test. The retention score was calculated by taking the sum score of the 23 remaining items, Cronbach's alpha = .65 (14 items about the general model and 9 items about the worked examples).

Transfer test

The transfer test was a paper-and pencil test that contained a short description of the skills expert researchers apply doing a literature search. The assignment was to design a blueprint for the training of this complex skill following the procedure of the 4C/ID model. The students had to construct a skills hierarchy, and design a sequence of learning tasks based on this hierarchy.

To score the results of the transfer test, a separate scoring form was developed consisting of 35 questions that checked to what extent and how accurately the procedure prescribed by the 4C/ID model had been applied in the transfer task. After the experiment, two independent raters scored the transfer tests using the scoring form. The interrater agreement was calculated, and only the items with a Cohen's kappa larger than .60 were used to calculate the total transfer score. This resulted in a transfer score based on the average score of both raters on 25 items that ranged from zero (none of the steps prescribed by the model taken) to 25 (all steps taken accurately).

Procedure

The experiment was carried out at the University of Gent in Belgium in six sessions of about two-and-a-half hours. In each session, between 11 and 18 students were tested simultaneously. The sessions took place in a classroom that had 18 multimedia computers connected to the server on which the instructional web site was installed. Nine computers had headphones attached to them and were connected to the audio versions of the instructions, whereas the other nine computers were connected to the visual-text versions. In two sessions, only the double system-paced versions of the instructions were presented, whereas in the other four sessions, the system-paced and the learner-paced versions were presented. When the students entered the classroom they were randomly assigned to one of the computers. Each computer showed a browser-window, without any of the menu options visible, that was set on the introduction page presenting some general information about the experiment. All students started at the same time and studied the instructions by themselves. After each diagram, a separate page followed with the mental effort scale. When a student clicked on one of the nine options, the program automatically continued with the next diagram.

Whenever the students had finished studying the instructions, they could start with the retention test that was also presented on the computer screen. This retention test was followed by a single mental effort scale. After completing this scale, the paper-and-pencil transfer test was handed out. The students received a blank answering form on which they could draw their design and write their comments. During the transfer test the computer screen indicated to the individual students how much time they had left, and after thirty minutes the display went red and the test was collected again. After the transfer test, again a mental effort scale had to be completed. Finally the students filled in a questionnaire to evaluate how they had experienced the experiment and whether there had been any problems with either the computer or the instructional material.

Results

The variables under analysis were time on diagrams; mental effort spent on the instructions, on the retention test and on the transfer test; retention score and transfer score. Except for time on diagrams, all scores were analysed with a 2x3 analysis of variance (ANOVA), with modality (audio vs. visual text) and pacing of the instructions (system-paced, double system-paced, and learner-paced) as the between-subjects factors. For all statistical tests, a significance level of .05 was applied. Table 1 shows the means on the dependent measures for all six groups.

With regard to time on diagrams, only the two learner-paced groups were compared, because in the system-paced and the double system-paced groups the time spent on studying the diagrams was fixed. The average time spent on the diagrams in the learner-paced conditions was 24.1 minutes (SD = 3.1). Furthermore, only one student had spent less than 19.3 minutes on the instructions (the time of the system-paced conditions), and none of the students needed as much as 38.6 minutes (the time of the double system-paced conditions). Audio and visual text did not differ significantly in time on diagrams, $t(29) = 1.13$, $p > .10$ (two-tailed).

The mean mental effort score during instruction was 4.2 (SD = 1.0), which is below average on the mental effort scale (which goes from 1 to 9). Comparing the different groups, a main effect of modality was obtained, $F(1, 88) = 6.35$, $MSE = 0.88$, $p = .01$. Students in the audio groups had spent less mental effort on the instructions (M = 3.9, SD = 1.0) than students in the visual groups (M = 4.4, SD = 0.9). The effect of pacing, however, was not significant, $F(2, 88) = 1.24$, and neither was the interaction of modality and pacing, $F(2, 88) = 0.32$. Furthermore, there were no significant differences in mental effort spent on the retention test $MSE = 1.43$, all Fs < 1, or in mental effort spent on the transfer test, $MSE = 1.38$, all Fs ≤ 1.

Table 1
Mean Scores on the Dependent Measures for Each Experimental Group (Standard Deviations Between Brackets)

	System-Paced		Double System-Paced		Learner-Paced	
	Audio	Visual Text	Audio	Visual Text	Audio	Visual Text
Time on Diagrams (minutes)	19.3 (0)	19.3 (0)	38.6 (0)	38.6 (0)	23.5 (2.5)	24.7 (3.5)
Mental Effort on Instructions (1-9)	4.1 (0.8)	4.6 (0.9)	3.8 (1.1)	4.1 (0.9)	3.9 (1.0)	4.5 (0.9)
Mental Effort on Retention Test (1-9)	6.1 (1.0)	5.9 (1.4)	5.9 (1.4)	5.8 (1.1)	5.9 (1.1)	6.0 (1.2)
Mental Effort on Transfer Test (1-9)	6.5 (1.2)	6.8 (1.1)	7.0 (1.4)	6.8 (0.9)	6.8 (1.3)	7.4 (1.2)
Retention Score (-23 to +23)	14.0 (4.7)	13.9 (4.4)	16.9 (4.9)	16.9 (4.7)	13.9 (4.7)	17.9 (5.1)
Transfer Score (0-25)	18.3 (3.4)	16.2 (4.5)	18.2 (4.5)	17.8 (3.7)	16.7 (4.6)	19.8 (4.5)

The results of the retention test showed no effect of modality, $MSE = 22.87$, $F(1, 88) = 1.78$, but a significant effect of pacing, $F(2, 88) = 3.07$, $p = .05$. A post-hoc analysis showed that the students in the double system-paced conditions ($M = 16.9$, $SD = 4.7$) scored significantly higher than their colleagues in the system-paced conditions ($M = 14.0$, $SD = 4.5$), $p = .04$ (Tukey's HSD). The interaction of modality and pacing was not significant, $F(2, 88) = 1.76$. A pairwise comparison within each level of pacing showed a reverse modality effect in the learner-paced groups, $F(1,88) = 4.75$, $p = .03$, with the visual-text group ($M = 17.9$, $SD = 5.1$) outperforming the audio group ($M = 13.9$, $SD = 4.7$).

In the transfer test, no main effects were found ($MSE = 20.43$; both Fs < 1), but the interaction of modality and pacing was significant, $F(2, 88) = 3.56$, $p = .03$. Figure 2 shows that visual text did worse than audio in the system-paced groups, that there were no differences in the double system-paced groups and that visual text outperformed audio in the learner-paced groups. Pairwise comparisons of the effect of modality within each level of pacing showed a significant difference between visual text ($M = 19.8$, $SD = 4.5$) and audio ($M = 16.7$, $SD = 4.6$) in the learner-paced groups, $F(1,88) = 4.75$, $p = .03$, but not in the system-paced or double system-paced groups.

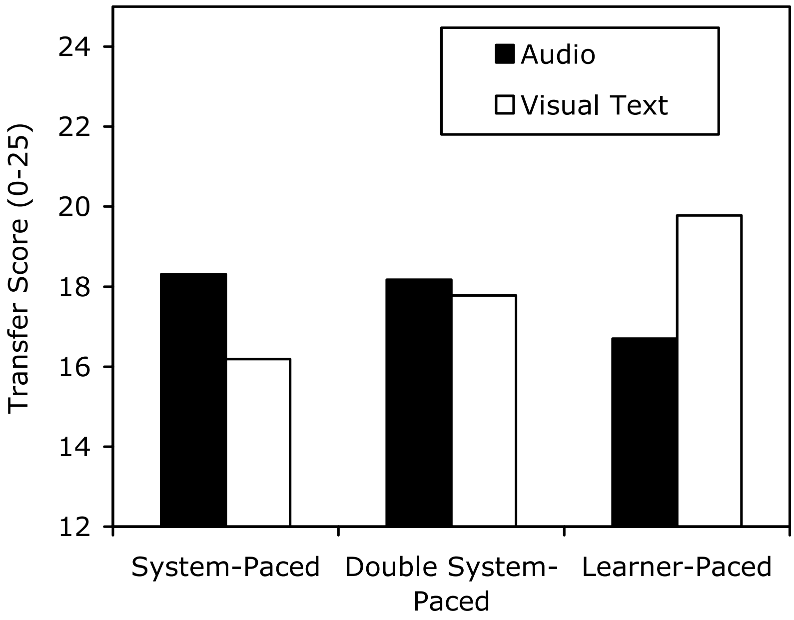


Figure 2
Mean transfer scores for all groups.

Discussion

Our first hypothesis that replacing visual text with spoken text in multimedia instructions will reduce the effort needed to integrate text and picture as a result of preventing split attention is strongly supported by our results. The perceived mental effort in the audio conditions is lower than in the visual-text conditions, no matter how much time people have to integrate text and picture. These results are clearly in line with earlier studies (Tabbers et al., 2001; Tindall-Ford et al., 1997; Van Gerven et al., 2002), and show that the modality effect in terms of mental effort also applies with prolonged time-on-task, either by slowing down the pacing or by introducing learner pacing. There is no indication of cognitive overload during instructions, because the overall mental effort scores during instruction stay below average. Moreover, the difference in mental effort is found only during the instructions and not during the tests, indicating that the higher mental effort scores in the visual-text conditions are the result of the higher perceived load of the instructions and not of some other differences between the experimental groups.

Our second hypothesis that only system-paced multimedia instructions will result in a modality effect on transfer is not supported so clearly. Spoken text leads to better transfer performance than visual text in the system-paced conditions, although the evidence is not very strong. Doubling the time to integrate text and pictures, however, removes any existing differences in transfer performance between spoken text and visual text. So prolonging time-on-task by slowing the pacing indeed seems to undo the modality effect in terms of transfer performance like we predicted. Moreover, introducing learner-pacing even leads to a reverse modality effect, with a superior performance for visual text on both the retention and the transfer test.

It is remarkable that with visual text, learner-pacing seems to be especially availing, whereas with spoken text, control over the pacing seems to have a negative effect. That might be related to the fact that listening to a spoken text is a rather passive process, which makes it very suitable for linear presentations with little interaction. Reading visual text, on the other hand, is a more active process, with the opportunity to go through texts strategically, paying more detail to some parts of the text and skipping over irrelevant elements. So giving the learner control over the pacing of the instructions seems to match with reading behaviour better than it does with listening behaviour. This would be an interesting hypothesis to investigate in more detail, for example by studying eye movements in multimedia learning to see if there are differences in looking behaviour between system-paced and learner-paced instructions.

The interaction of modality and pacing in multimedia learning as demonstrated in this study has some interesting implications, both at a theoretical and a practical level. The theoretical rationale for the modality effect presented by cognitive load theory (Sweller, 1999) and Mayer's theory of multimedia learning (Mayer, 2001) cannot account for a disappearance or reversal of the modality effect. An explanation in terms of preventing overload by utilising the auditory store in working memory would predict superior learning results in the audio conditions, even with extended time-on-task. However, by redefining the modality effect as a result of preventing split-attention, which can also account for earlier results, and by assuming that no overload occurs, the disappearance of the modality effect with extended time-on-task can be predicted correctly. With a split format of visual text and picture, learners have to compensate the higher cognitive load by investing more mental effort, and need more time to integrate the textual and pictorial information. But as long as cognitive load is not very high and the pacing of instructions is such that learners have enough time to process them, visual text is at least as effective as spoken text.

From a practical instructional design viewpoint, our findings imply that the guideline to use spoken text can be restricted to situations in which time pressures are high and instructions are system-paced, based on the pace of the narration, and to situations in which there is a potential high cognitive load so that it is not easy to compensate by investing more mental effort. In all other cases, visual text seems the more sensible presentation mode, especially because it is cheaper to produce, easier to deliver, and in combination with learner-paced instructions even more effective in terms of transfer of learning.

Of course, some reservations can be made concerning the generalisability of our findings. First, the instructional material used in our experiment is rather deviant from the material used in previous experiments on the modality effect. For example, Mayer and Moreno used a short animated multimedia message on how lightning develops (1998; Moreno & Mayer, 1999). Another point that can be made is that the regular classroom setting in which we conducted our experiment may have confounded our results somewhat. Students were able to see each other and notice that some of the students were listening through headphones and other students were reading text from the screen. This could have affected the students' perception of the difficulty of the instructions and thus the experienced mental effort. Furthermore, to prevent students in the same session from being distracted by each other due to large differences in time-on-task, the students in the double system-paced conditions were tested in separate sessions. Therefore, a replication of the experiment with other instructional materials and in more laboratory-like settings in which students cannot see each other would be needed to further strengthen our conclusions.

Nevertheless, with system-paced instructions, we successfully replicated the modality effect in an ecologically more valid classroom setting. This not only emphasises the relevance of the modality effect in multimedia design, but also supports the idea that the reversal of the effect with learner-paced instructions is not an experimental artefact, but rather the result of the interaction of modality with pacing in multimedia learning.

In conclusion, our study shows that when the pacing of multimedia instructions is not based on the narration and as long as no cognitive overload occurs, spoken text and visual text are equally effective. Giving learners control over the pacing of the instructions even makes on-screen text the most effective presentation mode. That suggests that the original modality effect in multimedia learning only applies to a restricted category of system-paced multimedia instructions, and that the theoretical rationale behind the effect has to be reconsidered. Research is needed to further clarify the role of pacing in multimedia learning in relation to presentation mode, so that more refined instructional design guidelines can be developed.

References

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

Baddeley, A. D. (1997). *Human memory: theory and practice* (revised ed.). Hove, United Kingdom: Psychology Press.

Borg, G., Bratfisch, O., & Dornic, S. (1971). On the problem of perceived difficulty. *Scandinavian Journal of Psychology, 12,* 249-260.

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

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 1084-1102.

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

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

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92,* 126-136.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93,* 390-397.

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,* 312-320.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology, Research & Development, 43,* 31-44.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

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

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84,* 429-434.

Penney, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17,* 398-422.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7,* 219-226.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

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

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-296.

Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2001). The modality effect in multimedia instructions. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 1024-1029). Mahwah, NJ: Erlbaum.

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

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2002). *Modality and variability as factors in training the elderly.* Manuscript submitted for publication.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training.* Englewood Cliffs, NJ: Educational Technology Publications.

CHAPTER 5 - GazeTracker™: A tool for studying eye movements in dynamic multimedia environments*

Abstract

This paper discusses the usefulness of GazeTracker™ as a tool to study eye movements in dynamic multimedia environments. By integrating itself into the low-level functions of the Windows operating system, GazeTracker™ links eye movement data to information about the internal computer processes and automatically combines the two for further analyses. The functionality of the tool is illustrated with an experiment on the integration of text and pictures in a web-based multimedia lesson on instructional design. In this experiment GazeTracker™ is used to connect eye fixation data to the associated web pages and to the areas of interest like pictures and text. The difference in fixation patterns between several information presentation formats is investigated. It is concluded that GazeTracker™ is well suited for conducting eye movement research with dynamic interfaces like web browsers, to study the way people integrate text and pictures in these environments.

In the field of eye movement research there are numerous studies on reading behaviour and on scene perception (see Rayner, 1998, for an overview), but hardly any studies on the integration of text and pictures (Duffy, 1992). There are some notable exceptions like the work of d'Ydewalle and colleagues on television subtitles (for an overview, see d'Ydewalle & Gielen, 1992), studies by Hegarty on mental animation (1992a, 1992b), some work on the perception of cartoons by Carroll, Young, and Guertin (1992), and more recently a study on how people look at advertisements by Rayner, Rotello, Stewart, Keir and Duffy (2001). Still, in the context of the rapid development of enabling technologies for multimedia presentations it is remarkable that so little attention is paid to such an important cognitive activity as integrating text and pictures.

Currently, there is a lack of theoretical understanding of how information presented in different modalities and modes in multimedia applications like web browsers is actually processed. We consider eye tracking as a useful method to study the integration of verbal and pictorial information in these dynamic environments. However, apart from the subtitling research of d'Ydewalle, all studies mentioned before used static scenes that were presented one at a time. Even researchers in the area of human-computer interaction, that use eye movement measures to determine the usability of interfaces, present their

* based on Tabbers, H. K., Paas, F., Lankford, C., Martens, R. L., & Van Merriënboer, J. J. G. (2002). *Gazetracker™: A tool for studying eye movements in dynamic multimedia environments*. Manuscript submitted for publication.

subjects with computer screens that have the elements of interest on fixed positions (e.g., Goldberg & Kotval, 1999).

One of the reasons why so little eye movement research has been done with dynamic interfaces might be that the combination with eye tracking data can result in a very complex analysis process. Fortunately, at the moment some interesting analysis tools have become available that integrate eye movement data with the processes that simultaneously take place on the computer screen (e.g., Crowe & Narayanan, 2000; Lankford, 2000b). One of these tools, GazeTracker™, will be described in this article, because it is perfectly suited for analysing eye movement data in dynamic multimedia applications like web browsers. To illustrate the usefulness of the tool, we also report an experiment in which we applied GazeTracker™ to study eye movements in multimedia learning, a research area in which eye-tracking methods have not been used before.

The GazeTracker™ software

GazeTracker™ is an analysis tool that resulted from the work on the Eye-gaze Response Interface Computer Aid (ERICA) at the University of Virginia. The ERICA system was originally developed to help individuals with disabilities communicate via the computer, and takes the eye movements of the user as input to operate mouse and keyboard functions in software applications (Lankford, 2000a). To facilitate the analysis of eye-movement data, the GazeTracker™ software was developed. GazeTracker™ offers three modes of analysis – image, video, and application.

In Image Analysis, GazeTracker™ records the eye-gaze data of test subjects as they view sequences of static images shown on a computer display. In Video Analysis, GazeTracker™ synchronises a test subject's eye-gaze response to the video clips he or she watches on a computer during an experiment, simplifying eye-gaze analysis for dynamic video content. Similarly, in Application Analysis, the program combines the input from eye-tracking systems like ERICA, ASL or SMI with information about the activities of the user of a computer application, like keystrokes and mouse clicks. In this way, all activities on the screen can be related to gaze position data, which gives the opportunity to track eye movements in several applications simultaneously and even control for scrolling behaviour in web browsers. Moreover, specific areas of interest, called LookZones, can be defined for separate windows and for web pages. These LookZones provide metrics concerning how long and how often a test subject observed different areas of interest.

After recording, the data including the interactions of the user with the applications can be replayed, and can be displayed as a gaze trail, which depicts the scan path of a test subject, superimposed on an application window. The program also provides other graphical analysis methods, such as bar charts in Excel based on the LookZone data or three-dimensional views of the application window with the time duration of the eye-tracking data in different regions depicted in the z-dimension. GazeTracker™ also allows experimenters to export the data to text files or to Microsoft Excel for further statistical analysis in other statistical software packages (for more information about GazeTrackerTM, see Lankford, 2000b).

The most interesting feature of GazeTracker™ for the analysis of dynamic picture-text combinations is the possibility to track all operations people perform in a specific window and relate these to the gaze position on the computer screen. That way it becomes much easier to conduct eye movement research with dynamic interfaces like web browsers, and study the way people integrate textual and pictorial information in these environments. GazeTracker™ accomplishes this by integrating itself into the low-level functions of the Windows operating system. This allows GazeTracker™ to capture all mouse and keyboard events that occur on the computer. GazeTracker™ receives the eye-tracking data through a serial port and uses a global timer to synchronise the data it reads from the serial port with the mouse and keyboard data it intercepts from the operating system. GazeTracker™'s integration with Windows also allows it to track the web pages that each test subject visits and to compensate the recorded eye-gaze and mouse data with the current scroll bar position in Internet Explorer. This ensures that all captured data is associated with the proper content shown on the screen during the experiment. Moreover, the program can parse the HTML-code of web pages and automatically create LookZones for each hyperlink or image. This makes GazeTracker™ an ideal tool for doing research in any area related to studying cognitive activities when people are interacting with dynamic computer applications, for example in the area of multimedia learning.

A study of eye movements in multimedia learning

In educational psychology, there is much debate about how students learn when they have to integrate verbal and non-verbal information, mostly text and pictures. Recent theories on multimedia learning like Mayer's generative theory (2001) and cognitive load theory (Sweller, 1999; Sweller, Van Merriënboer & Paas, 1998) claim that the mental integration of pictures and text is a process that can easily overload the limited working memory resources and thus interfere with learning. This has produced a great amount of empirical research on the effectiveness of different presentation formats of multimedia instructions (for an overview, see Mayer, 2001; Sweller, 1999). However, as far as we know, no research in this area has used eye-tracking methods before to study looking behaviour. We think it might be very interesting to study eye movements in multimedia learning, especially because the existing theories are for a great deal based on assumptions about where people look when they are integrating text and pictures.

Our study focuses on the so-called *modality effect,* one of the main findings in the field of multimedia learning that replacing on-screen text with spoken text reduces the working memory load and that way improves learning. The superiority of spoken text over visual text in multimedia learning has been demonstrated in several studies in terms of faster problem solving (Jeung, Chandler & Sweller, 1997; Mousavi, Low & Sweller, 1995), higher scores on retention and transfer tests (Kalyuga, Chandler & Sweller, 1999, 2000; Mayer & Moreno, 1998; Moreno & Mayer, 1999) and less mental effort reported by the learners (Tabbers, Martens & Van Merriënboer, 2001; Tindall-Ford, Chandler & Sweller, 1997). The assumption behind the modality effect in multimedia learning is that the mental integration of spoken text and pictures is less demanding than the integration of visual text and pictures in terms of cognitive load, because the modality-specific resources in working memory are used more effectively.

However, most studies of the modality effect used system-paced instructions in which the time to integrate text and picture was limited, based on the time of the narration. That leaves open another explanation for the superiority of audio over visual text in multimedia learning, namely that students listening to a narration have more time to integrate text and pictures than students that have to divide their attention between visual text and picture. In one experiment, we tested this assumption by comparing system-paced multimedia instructions with learner-paced instructions (Tabbers et al., 2001). With system-paced instructions, spoken text led to superior results on a subsequent transfer test, whereas with learner-paced instructions no difference in effectiveness was found between spoken text and visual text. That confirmed our idea that the amount of time to integrate text and pictures plays an important role in multimedia learning.

To corroborate our findings we set up an experiment to study eye movements with the same multimedia material as in our 2001 study. We compared three conditions differing in the presentation format of the instructions: two system-paced conditions, with either audio or visual text, in which the pacing was based on the time of the narration, and one learner-paced condition with visual text. Our earlier study had shown that the system-paced visual-text condition resulted in the worst transfer performance. We explained this effect by stating that the students in this condition lack the time to integrate the text with the picture. Translated to eye movements that would imply that total fixation time in the pictures should be less in the system-paced visual-text condition than in the audio condition or the learner-paced condition.

Moreover, the question is how the fixations are divided over picture and text. In their work on the integration of picture and text, Carroll et al. (1992), Hegarty (1992a, 1992b) and Rayner (2001) found that subjects tended to read the text first and then look at the picture, without much switching. As study time was not limited in these studies, we would expect the same fixation pattern in the learner-paced condition. On the other hand, in the system-paced visual-text condition different patterns might occur as a result of the time constraints. Finally, we also looked at the original explanation of the modality effect in terms of less cognitive load. Therefore, we not only measured the mental effort our students had spent on the instructions but also looked at some possible indicators of mental workload like fixation frequency and fixation duration (Van Orden, Limbert, Makeig & Jung, 2001).

Method

Participants and Design
The participants were 12 students from a Teacher Training College for Primary Education in Heerlen, the Netherlands (age between 17 and 23; 1 male and 11 females). They had applied on a voluntary base and were paid 10 euro for their participation. Each participant studied the multimedia instructions in three parts and each part was presented in a different presentation format, so that the participant received all conditions. To prevent any sequencing effects in this within-subject design, the order of presentation formats was counterbalanced between the participants.

Apparatus
The eye movements were recorded with a 50Hz video-based remote eye-tracking device from SensoMotoric Instruments (SMI). The infrared camera was placed under the 21-inch display screen of the stimulus PC on which the multimedia instructions were presented. Special SMI-software to operate the camera and the calibration process ran on a separate PC that was connected to the stimulus PC. On the stimulus PC the GazeTracker™ program combined the input of eye movement data from the SMI-PC with data of the user interactions with the web browser. A chin and forehead rest was placed in front of the screen in such a way that the subject's eye was 70 centimetres from the computer screen and level with its centre.

Materials
Multimedia instructions. The instructions used in the experiment were developed with Microsoft FrontPage as a linear sequence of web pages. On these pages it was shown how to develop instructions for the training of complex cognitive skills, following the design procedure of the four-component instructional design model (4C/ID model) of Van Merriënboer (1997). Each page consisted of a diagram representing a skills hierarchy or an elaborated sequence of learning tasks and a textual explanation accompanying the diagram. The instructions were divided in three parts. The first part contained three diagrams that showed the different stages in developing a blueprint for the training of the complex skill *doing experimental research*, whereas the second part of the instructions presented the same process for the complex skill *designing a house*, again with three diagrams. In the third part the general strategy of the 4C/ID model was displayed in two diagrams. The textual explanation that accompanied each diagram was presented in smaller pieces of only one or two sentences long, in such a way that each piece of text referred to a specific part of a diagram. Moreover, the part of the diagram that the text referred to was coloured bright red. While scrolling through the explanation of a diagram, only the accompanying text and the colour-coding changed, not the diagram itself.

Furthermore, three versions of the instructions were created that differed in presentation format: a system-paced audio version, a system-paced visual-text version and a learner-paced visual-text version. In the system-paced audio version, students could listen to the explanation that accompanied a diagram through the speakers next to the computer screen. When an audio fragment of one or two sentences had finished playing, the colour-coding in the diagram changed and the next audio fragment started. In the system-paced visual-text version, the explanation was placed right above the diagram. After exactly the same period of time as in the audio version, the colour-coding in the diagram changed and a new piece of text appeared above the diagram. With the learner-paced visual-text version students could reread each piece of text as many times as they wanted to. To continue with the next piece of text students had to click on a forward button. Thus in the learner-paced visual text version the total time to study each diagram was variable.

Mental effort scale. After each diagram, a subjective measure of mental effort was administered. This was a 9-point scale on which the students could rate the mental effort they had spent on the diagram ranging from *very, very low mental effort* to *very, very high mental effort*. The scale was developed by Paas (1992), based on a measure of perceived task difficulty of Borg, Bratfisch, and Dornic (1971). The scale's reliability and sensitivity (Paas, Van Merriënboer & Adam, 1994) and its non-intrusive nature make this scale a useful measure of perceived working memory load, and it has been used extensively in studies of multimedia learning (e.g., Kalyuga et al., 1999; Mayer & Chandler, 2001; Tabbers et al., 2001).

Evaluation questionnaire. The evaluation questionnaire contained 12 items about the instructional material, each accompanied with a 5-point scale on which students could indicate how much they agreed with the content of each item. We used this questionnaire to get an idea if the students had understood the instructions, if they had experienced any problems and if they had worked with concentration. It also contained the question which presentation format student had liked best.

Procedure

The students were tested one at a time. They were seated in a solid chair that could not move and told to put their heads in the chin and forehead rest that was positioned in front of the computer screen. First they read some general information about the experiment without anything being recorded. Subsequently, their eye movements were calibrated after which they could start studying the first part of the instructions. After each diagram in the worked-out example, the students had to fill in a subjective mental effort scale that was presented on the screen. When a student clicked on one of the nine options, the program automatically continued with the next diagram. When the students had finished studying the first part of the instructions, their eyes were once again calibrated and the second part of the instructions was presented in a different presentation format. This procedure was repeated after the second part of the instructions. After finishing the third part, students could remove their heads from the chin and forehead rest and the eye movement recording was stopped. To conclude, the students completed the evaluation questionnaire on the computer screen. The whole procedure took about three-quarters of an hour.

Results and discussion

The GazeTracker™ program uses a dispersion-threshold identification algorithm to calculate fixations (see Salvucci & Goldberg, 2000). The dispersion threshold was set at 25 pixels, which corresponds to approximately three or four letter spaces in the instructional material or one degree of visual angle, and the duration threshold was set at 100 milliseconds. The main dependent variables in the experiment were total fixation time, number of fixations, fixation duration, fixation frequency and subjective mental effort. We conducted a repeated measures MANOVA, with presentation format as the within-subjects factor. For any post-hoc analyses we used paired t-tests. For all statistical tests, a significance level of .05 was applied. Table 1 shows the means and standard deviations for all dependent measures.

Table 1
Means and Standard Deviations of Dependent Measures

		System-Paced Audio		System-Paced Visual-Text		Learner-Paced Visual-Text	
		M	*SD*	*M*	*SD*	*M*	*SD*
Overall:	Number of Fixations	509	302	604	340	765	420
	Total Fixation Time (s)	158	97	139	82	174	100
	Average Fixation Duration (s)	0.31	0.05	0.22	0.02	0.22	0.02
	Fixation Frequency	2.26	0.17	2.79	0.33	2.89	0.20
	Mental effort score (1-9)	4.2	1.0	4.8	1.4	4.1	1.0
Diagram:	Number of Fixations	497	295	243	172	250	133
	Total Time Fixated (s)	156	96	66	51	69	40
	Average Fixation Duration (s)	0.31	0.05	0.25	0.05	0.26	0.03
Text:	Number of Fixations	-	-	341	226	488	305
	Total Time Fixated (s)	-	-	68	44	97	63
	Average Fixation Duration (s)	-	-	0.20	0.03	0.20	0.03

For the fixation time and number of fixations, we found an overall significant effect of presentation format, Wilks' lambda = 0.24, $F(4, 42) = 10.88$, $p < .01$, but no specific differences in the post-hoc tests. The effect of presentation format on the indicators of workload like fixation duration, fixation frequency and mental effort was also significant, Wilks' lambda = 0.16, $F(6, 40)$ = 10,88, $p < .01$. Post-hoc comparisons showed that the audio condition was lower in fixation frequency than both the system-paced visual-text condition, t = 4.85, $p < .01$, and the learner-paced visual text condition, t = 8.23, $p < .01$. Also the fixation duration was longer in the audio condition than in the system-paced visual-text condition, t = 6.73, $p < .01$, and the learner-paced visual-text condition, $t = 6.34$, $p < .01$. Although the system-paced visual-text condition produced a higher mental effort score than the other two conditions, this difference was not statistically significant.

In the audio condition students spent more than 98% of their fixation time in the diagrams, versus 44% in the system-paced and 38% in the learner-paced visual text version. Moreover, the average fixation duration in the diagram was significantly longer in the audio condition than in the system-paced visual-text condition, $t = 2.88$, $p < .05$, and in the learner-paced visual-text condition, $t = 2.98$, $p < .05$. Comparing the visual-text conditions on the fixations in the text, no statistically significant differences were found in the fixation data.Finally, the results of the evaluation questionnaire showed that two-thirds of the students had preferred the learner-paced visual-text version. Moreover, the students judged the part of the instructions presented in the learner-paced visual-text version as the easiest to comprehend.

Overall, the results do not show clear differences in fixation pattern between the presentation formats. Of course the pattern in the audio condition deviates because there is no text to fixate on, but apart from that there are no apparent differences in fixation number or fixation time. Moreover, the division of attention between picture and text in the visual text conditions seems to be quite identical, contrary to what we expected. Furthermore, looking at the possible workload indicators, it is interesting to see that the students in the audio condition fixate less frequently but with a longer duration. Even if we only look at the fixations in the diagram, average fixation duration is still longer in the audio condition. Primarily, this difference seems to reflect the calmness of the looking pattern in the audio condition, where students do not have to switch between text and picture. It does not, however, seem to indicate less cognitive load in the audio condition, because we do not find a similar difference between audio and both visual text conditions in the mental effort scores.

We hypothesised that the students in the learner-paced condition would spend extra time in the pictures, but we do not find it in the results. So the difference in effectiveness found in other studies between system-paced and learner-paced multimedia instructions with visual text does not seem to derive from an overall difference in fixation pattern. Nevertheless, students report a higher mental effort in the system-paced condition, and generally prefer the learner-paced visual-text version. It might be that the superiority of learner-paced over system-paced visual-text is not so much the result of a general difference in fixation patterns, but because students can control the division of attention between pictures and text more easily and adapt it to their individual needs. To test this hypothesis, an approach beyond the scope of the present study is needed that directly links the eye movement data to a cognitive model, for example with the technique of tracing (Salvucci & Anderson, 2001).

General discussion

Our study shows the usefulness of GazeTracker™ as a tool for analysing eye movement data with a dynamic presentation of text and pictures in a web browser. Despite the dynamic nature of the presented material and the large number of different web pages that the students had to study, the analysis could be done relatively easily, because GazeTracker™ kept track of the events on the computer screen and the interactions of the student with the computer and linked these data to the eye movement data. Also the areas of interest in our study, the text and pictures, were loaded automatically as LookZones. Finally, the GazeTracker™ program simplified the subsequent data analysis by offering the opportunity to indicate which subjects, which web pages and which LookZones should or should not be included in the analysis.

In conclusion, the use of a program like GazeTracker™ gives the eye movement research community the means to expand the area of interest to dynamic computer applications like web browsers, and study the process of integrating different information elements that are presented at different locations and at different times. This kind of research will be of great interest for any study of human-computer interactions, including the area of multimedia learning, because it enables fine-grained analyses of the cognitive processes that take place when people are working with a dynamic computer application.

References

Borg, G., Bratfisch, O., & Dornic, S. (1971). On the problem of perceived difficulty. *Scandinavian Journal of Psychology, 12,* 249-260.

Carroll, P. J., Young, J. R., & Guertin, M. S. (1992). Visual analysis of cartoons: A view from the far side. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 444-461). New York: Springer-Verlag.

Crowe, E. C., & Narayanan, N. H. (2000). Comparing interfaces based on what users watch and do. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 29-36). New York: ACM Press.

d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 415-427). New York: Springer-Verlag.

Duffy, S. A. (1992). Eye movements and complex comprehension processes. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 462-471). New York: Springer-Verlag.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24,* 631-645.

Hegarty, M. (1992a). The mechanics of comprehension and comprehension of mechanics. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 428-443). New York: Springer-Verlag.

Hegarty, M. (1992b). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 1084-1102.

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

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

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92,* 126-136.

Lankford, C. (2000a). Effective eye-gaze input into Windows™. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 23-27). New York: ACM Press.

Lankford, C. (2000b). GazeTracker™: Software designed to facilitate eye movement analysis. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 57-63). New York: ACM Press.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93,* 390-397.

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,* 312-320.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

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

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84,* 429-434.

Paas, F., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79,* 419-430.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7,* 219-226.

Salvucci, D. D., & Anderson, J. R. (2001). Automated eye-movement protocol analysis. *Human-Computer Interaction, 16,* 39-86.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71-78). New York: ACM Press.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-296.

Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2001). The modality effect in multimedia instructions. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 1024-1029). Mahwah, NJ: Erlbaum.

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

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training.* Englewood Cliffs, NJ: Educational Technology Publications.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 43,* 111-121.

CHAPTER 6 - General discussion

The main aim of this thesis was to take a closer look at the modality effect in multimedia learning and to provide a set of refined guidelines for the use of spoken text in multimedia instructions. Based on the existing evidence for the modality effect, the questions were raised whether the effect can be generalised to longer instructions from a domain other than the exact sciences, whether the effect can be replicated in a regular classroom setting, and whether the effect still occurs if the time-on-task is varied by introducing learner-pacing in the instructions. In the previous chapters, five studies were presented in which the generalisability of the modality effect was tested and the interaction of modality and pacing was investigated. This chapter briefly reviews the results of these studies, and discusses the implications of the findings for theories of multimedia learning. Furthermore, a set of refined design guidelines is presented that can be derived from the results, and suggestions are made for further research. The chapter concludes with some final remarks.

Review of the results

The studies in this thesis were set up to test the generalisability of the modality effect in multimedia learning. Thus, the instructional materials used in these studies differed in several ways from the materials used in previous research on the modality effect (Jeung, Chandler & Sweller, 1997; Kalyuga, Chandler & Sweller, 1999, 2000; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi, Low & Sweller, 1995; Tindall-Ford, Chandler & Sweller, 1997). First of all, the subject matter of the instructions did not originate in the exact sciences but concerned the teaching of an instructional design strategy. Moreover, the presentation of the instructions lasted at least twenty minutes in even the fastest versions of the task, and not just a few minutes as in previous studies. Finally, the pacing of the instructions was varied, whereas earlier research used either system-pacing based on the pace of the narration or, with paper-based instructions, restricted time-on-task based on the total time of the narration. So the only similarity with the instructions used in earlier research was the presentation of pictorial and verbal information that had to be integrated mentally.

The results of the studies in this thesis are partly a replication of the results of previous research on the modality effect. First, in almost all studies, replacing visual text with spoken text yielded lower mental effort scores reported by the learners during the instructions (Chapter 2, Chapter 3: first study, and Chapter 4). Only two studies (Chapter 3: second study, and Chapter 5) failed to find any differences on the subjective mental effort measures. Such lower mental effort scores with spoken text have also been reported by Tindall-Ford et al. (1997). Second, with system-paced instructions based on the pace of the narration, replacing visual text with spoken text resulted in higher retention scores (Chapter 3: second study) and higher transfer scores (Chapter 3: second study, and Chapter 4). One study did not find superior learning results for spoken text with system-paced instructions (Chapter 3: first study), but the test scores in this specific study were quite low, so that the absence of an effect can possibly be attributed to a bottom-effect. Overall, the learning gains obtained with spoken text are fairly in line with the results of earlier research on the modality effect (Kalyuga et al., 1999, 2000; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Tindall-Ford et al., 1997).

Changing the pace of the instructions, however, resulted in a totally different pattern with regard to the learning results. When the pacing was slowed down, the superiority of spoken text over visual text on both retention and transfer test disappeared (Chapter 4). Moreover, allowing the learners in the visual-text conditions to determine the pace of the instructions resulted in retention scores that were just as good as the scores in the spoken-text conditions (Chapter 4) or even better (Chapter 2 and Chapter 3: second study), as well as transfer scores that were also equally good (Chapter 3: second study) or better (Chapter 2 and Chapter 4). So by slowing down the pace of the instructions, the modality effect in multimedia instructions could not be obtained anymore, and with learner-pacing the effect even turned around, with superior learning results for the visual-text conditions.

Two main conclusions can be drawn from these results. On the one hand, with system-paced instructions based on the pace of the narration, the modality effect can be generalised to multimedia materials with a longer duration on a subject outside the exact sciences. The modality effect as demonstrated in previous research has been replicated, with a consistent superiority of spoken text over visual text. The mere fact that the effect was also obtained outside the laboratory walls, in an ecologically more valid classroom setting (Chapter 3: second study, and Chapter 4), emphasises the practical relevance of the modality effect for multimedia design.

On the other hand, as soon as the pacing of the instructions is changed either by slowing down the pace or by introducing learner-pacing, the modality effect vanishes into thin air, and the use of spoken text is not superior to the use of visual text anymore. With learner-paced instructions the effect even reverses, with superior learning results for visual text instead of spoken text. This interaction between modality and pacing in multimedia learning seems to be partly the result of an increase in time-on-task, but also partly of some other aspect of learner-pacing that makes it more suitable for reading text than for listening to a narration.

Theoretical implications

The findings of this thesis have some implications for the theoretical account of the modality effect in multimedia learning as given by the theories of Sweller (1999) and Mayer (2001).

First, Sweller and Mayer explained the modality effect in terms of spoken text addressing the phonological loop and preventing overload in the visual channel. However, as has been argued in Chapter 4, this explanation is not in line with Baddeley's model of working memory (Baddeley, 1992), in which visual text is converted into a phonological code and processed in the same slave system as spoken text. Moreover, the finding that spoken text leads to less mental effort and better learning results can also be explained by the fact that replacing visual text with spoken text prevents split attention. So apart from the fact that an explanation of the modality effect in terms of special mechanisms in working memory is not very convincing, it is also not needed. Consequently, the modality effect in multimedia learning can best be accounted for as the result of preventing split attention between verbal and pictorial information.

Second, both cognitive load theory and Mayer's theory of multimedia learning do not take the pacing of instructions into account. Their guidelines for multimedia design are derived from the results of experiments in which instructions were used with a pacing based on the pace and length of the narration. The efficiency of a presentation mode in terms of time-on-task was not considered. However, the results of this thesis show that pacing is a highly relevant factor for the modality effect in multimedia learning. Because spoken text and picture can be perceived simultaneously, the processing of instructions in this presentation mode is very efficient. If learners have to split their attention between visual text and picture, it will not only cost more mental effort but also more time to integrate the information sources. So with a fixed time-on-task based on the narration, split visual-text formats will be less effective than spoken-text formats. But if time-on-task is extended and no overload occurs, the difference between both presentation modes can simply be compensated by spending more time and more effort. Especially if visual search is kept to a minimum, visual text can be just as effective or even more effective than spoken text.

This explanation might also account for the somewhat paradoxical finding that presenting text and picture sequentially rather than simultaneously did not lead to the expected temporal split-attention effect, as long as the text was presented in short fragments (Mayer, Moreno, Boire & Vagge, 1999; Moreno & Mayer, 1999; Mousavi et al., 1995). In one experiment, Moreno and Mayer even found that sequential presentation of visual text and picture led to better learning results than simultaneous presentation. This can only be explained by taking the time-on-task in account. With a sequential presentation format, learners have more time to integrate text and picture, so the disadvantage of the temporal split-format can be overcome.

A final point is that both Mayer's theory of multimedia learning and cognitive load theory do not stress the qualitative differences between listening and reading. Although these differences do not surface with system-paced instructions, with learner-paced instructions they seem to become relevant as they lead to a superiority of visual text over spoken text. It might be argued that integrating spoken text and pictures is a somewhat passive process, with little opportunities for strategic learning behaviour, whereas integrating visual text and pictures is a more active process in which strategic aspects play a more important role. However, more research is needed, for example by looking at eye-movement patterns in multimedia learning (as in Chapter 5), to clarify this issue.

Refined guidelines for the design of multimedia instructions

In Chapter 1, the following three guidelines for the design of multimedia instructions were discussed that resulted from cognitive load theory (Sweller, 1999) and Mayer's generative theory of multimedia learning (Mayer, 2001).

1. Get rid of redundant information
2. Prevent split attention
3. Use spoken text, not visual text

The first guideline precedes the other two, because the prevention of split-attention and the use of spoken text are only advisable if both text and picture are necessary for complete understanding of the instructional materials. In all other cases, one of the two is redundant and should be removed. But if both text and picture are necessary in the instructions, the guidelines do not prescribe when text and picture should be physically integrated to prevent split attention, and when visual text should be replaced with spoken text. Based on the results of the studies reported in this thesis, a set of refined guidelines is proposed for the design of multimedia instructions that resolves this problem.

1. Get rid of redundant information
2a. With system-paced instructions, use spoken text
2b. With learner-paced instructions, use visual text
3. Prevent visual search

Guideline 1: Get rid of redundant information

This guideline is identical to the one presented in Chapter 1. The message is simple but important, to remove any information elements that are not necessary for learning.

Guideline 2a: With system-paced instructions, use spoken text

If instructions are system-paced, for example because time pressure is high, the text accompanying a picture or animation should be presented as spoken text. In this way, learners do not have to split their attention and can process text and picture simultaneously. Working memory load will be less and learning results will be better than with visual text.

Guideline 2b: With learner-paced instructions, use visual text

If instructions are learner-paced, the text accompanying a picture or animation should be presented as visual text. Although learners have to split their attention between visual text and picture, they can process the information flexibly, read at their own pace, easily skip over information they already know, and pay more attention to information that is new to them. As long as no cognitive overload occurs, it is worth the extra mental effort needed to integrate text and picture.

Guideline 3: Prevent visual search

Present information elements that refer to each other in such a way that visual search is minimised. This can be accomplished in several ways:

- use visual cues that relate the text to the appropriate part of the picture (also with spoken text);
- present not all the text accompanying a picture or animation at once, but in smaller pieces, so that learners do not have to search the text;
- physically integrate text and picture, so that each piece of text is placed next to the appropriate part of the picture.

These new guidelines are a refinement of the original guidelines in that they are more specific in determining when spoken text should be used. Only in situations in which time-on-task is a crucial variable and the instructions are system-paced based on the pace of the narration, should spoken text be first choice in multimedia instructions. Because the narration and the picture or animation can be perceived simultaneously, it is the most efficient presentation mode in terms of time-on-task. However, it also has the disadvantage of the static and linear nature of spoken text. Therefore, in all other cases and as long as no cognitive overload occurs, visual text is to be preferred over spoken text in multimedia learning. Processing the information can be much more flexible, which makes visual text more effective than spoken text in terms of learning results. From a practical viewpoint, another advantage of visual text is that it is often much cheaper to produce and easier to deliver than spoken text. Finally, to prevent too much mental effort having to be invested as a result of the split-attention format of visual text and picture, visual search should be minimised.

Directions for further research

The studies reported in this thesis convincingly show the generalisability of the modality effect and the interaction of pacing and modality in multimedia learning. However, some aspects of the results of the studies need to be corroborated through further research and other aspects can be expanded into new directions.

As discussed in some of the previous chapters, the studies reported in this thesis have some methodological drawbacks as a result of the classroom setting in which most experiments were conducted. For example, students could see each other wearing headphones or reading from the screen, and a somewhat unstable environment like the Internet was used to deliver the instructions. On the one hand, these circumstances do not seem to have played a decisive role, as consistent results were obtained that were in line with earlier research.

Moreover, the setting of the studies also reflects actual educational practice, which gives the results a great ecological validity. On the other hand, a replication of the studies in more controlled laboratory settings where learners cannot see each other would strengthen the conclusions.

Furthermore, whereas the modality effect found in previous studies has been replicated with system-paced instructions, it would be interesting to try replicate the reverse modality effect found in this thesis with a learner-paced version of the instructional materials of Mayer and Sweller. Getting the same results with their instructions would be a good indication of the generalisability of the interaction of pacing and modality in multimedia learning.

Another aspect of the results that could be supported with more substantial evidence is the mental effort measurements. Although the use of self-report scales gives a useful indication of relative differences in cognitive load, it also has the disadvantage of being an indirect measurement. Moreover, the perceived load of the instructions may be influenced by certain preconceptions of the learners on the ease of listening versus the ease of reading (Cennamo, 1993; Salomon, 1983, 1984). More objective measures such as dual-task methods are needed to give further empirical support for the assumptions on working memory load during multimedia learning (see for example Brünken, Plass & Leutner, 2002).

An interesting new direction for the research on multimedia learning is the use of eye-tracking methods. As demonstrated in Chapter 5, measuring eye movements in multimedia learning yields useful data on the way people integrate text and picture that cannot be obtained with other methods. That makes it possible to test certain assumptions of theories of multimedia learning. Moreover, it can be used to assess whether learners need both the pictures and text to understand the instructions or whether one of the two is redundant. Also, more fine-grained cognitive models of the learning process can be tested by tracing the eye-movement protocol (see Salvucci & Anderson, 2001, for an example of this technique).

The results of this thesis also underline the importance of extending the research on multimedia learning to more interactive learning environments. Effects that apply under more strict system-paced conditions might not work or have different outcomes when learners interact with the program. This can be simply tested by comparing system-pacing with other kinds of pacing. An example is a study by Mayer and Chandler (2001), who found that presenting an animation in smaller parts produced better learning results than presenting it as a whole. Also interactions with other factors than control over the pacing could be investigated. For example, giving the learner the choice over the preferred presentation mode (Plass, Chun, Mayer & Leutner, 1998), or providing external aids to relieve working memory load, are interesting explorations of the relationship between user interactions and multimedia learning.

Final remarks

People learn. And multimedia instructions can be used to make this learning efficient and effective. Maybe they can even turn it into a pleasant experience. Nevertheless, this will only be achieved if instructional designers have guidelines at their disposal that tell them how to produce multimedia learning environments that work. This thesis has tried to make a contribution to the development of these guidelines, by taking a closer look at the use of spoken word in multimedia instructions, by critically testing the underlying assumptions behind the existing guidelines, by investigating the interaction of modality and pacing in multimedia learning, by applying methods that have not been used before like eye-tracking, and, last but not least, by presenting a set of refined guidelines that make multimedia learning more effective. So multimedia learning can become fun again, I hopefully proclaim.

References

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

Brünken, R., Plass, J. L., & Leutner, D. (2002). *Direct measurement of cognitive load in multimedia learning.* Manuscript submitted for publication.

Cennamo, K. S. (1993). Learning from video: Factors influencing learners' preconceptions and invested mental effort. *Educational Technology, Research & Development, 41,* 33-45.

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

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

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92,* 126-136.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93,* 390-397.

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,* 312-320.

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

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

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

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90,* 25-36.

Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist, 18,* 42-50.

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76,* 647-658.

Salvucci, D. D., & Anderson, J. R. (2001). Automated eye-movement protocol analysis. *Human-Computer Interaction, 16,* 39-86.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

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

SHORT SUMMARY

According to cognitive load theory (Sweller, 1999) and Mayer's theory of multimedia learning (Mayer, 2001), presenting information in two sensory modalities rather than one leads to a more efficient use of working memory resources so that cognitive overload is prevented and learning is improved. This is called the modality effect and has been demonstrated before in several experiments. The resulting guideline for multimedia instructions is that the text accompanying a picture or animation should be presented as spoken text, rather than visual text. However, previous research on the modality effect used only short, system-paced instructions, on subjects from the exact sciences like geometry, so it is yet unclear if the guideline to use spoken text is as generally applicable as is suggested. The main aim of this thesis is to take a closer look at the modality effect in multimedia learning and to provide a set of refined guidelines for the design of multimedia instructions. Five experimental studies are presented in which the generalisability of the modality effect to another content area is tested with a multimedia lesson on the subject of instructional design, consisting of diagrams and explanatory text. Furthermore, the relationship of the modality effect with the use of visual cues, time-on-task and the pacing of instructions is investigated, and learners' fixation patterns are studied with a tool for measuring eye movements in a multimedia environment.

Chapter 1 gives a general introduction to the thesis. First, the limited capacity of working memory and its modality-specific subsystems are discussed in relation to multimedia learning. Subsequently, cognitive load theory and Mayer's generative theory of multimedia learning are introduced, as well as the guidelines for the design of effective multimedia instructions that result from these theories. The following three guidelines are discussed together with their empirical support:

1. Get rid of redundant information
2. Prevent split attention
3. Use spoken text, not visual text

The evidence for the third guideline that is based on the modality effect in multimedia learning is critically reviewed, and the two main research questions are introduced. First, does the modality effect generalise to longer instructions on a subject other than the short multimedia messages on subjects from the exact sciences used in previous research? Second, does the modality effect also occur with instructions that are not system-paced based on the pace of the narration? Finally, an overview is given of the other chapters of the thesis.

Chapter 2 reports a study in which the modality effect is tested in a classroom setting with a learner-paced version of the multimedia lesson. Also, the role of preventing visual search is examined by adding visual cues to the instructions that relate the text to the right parts of the diagrams. The hypotheses resulting from cognitive load theory and Mayer's theory of multimedia learning are that replacing visual text with spoken text (the modality effect) and adding visual cues to the diagrams (the cueing effect) will decrease the mental effort invested by the learners and give better learning results. Four versions of the instructions were compared varying in modality of the explanatory text (spoken vs. visual text) and the use of cues in the diagrams (cues vs. no cues). The participants were 111 second-year students of educational science (age between 19 and 25 years). They studied a web-based multimedia lesson on instructional design for about one hour. Afterwards they completed a retention and a transfer test. During both the instruction and the tests, self-report measures of mental effort were administered. The results show that adding visual cues to the pictures only resulted in higher retention scores, whereas replacing visual text with spoken text resulted in lower retention and transfer scores. Thus, only a weak cueing effect and even a reverse modality effect were found. The explanation given for the reverse modality effect is that the multimedia instructions in this study were learner-paced, as opposed to the system-paced instructions used in earlier research.

Chapter 3 presents two studies. The first study is a replication of the study in Chapter 2, only this time with a system-paced version of the instructions. The participants were 41 students from a Teacher Training College (age between 18 and 24) who got either an audio or a visual-text version of the multimedia lesson. They rated their mental effort and made a retention and transfer test. As expected, the audio group reported less effort than the visual text group on both the instructions and the transfer test. No differences in learning results were found, which is attributed to a bottom-effect as a result of the difficulty of the tests for the participants in this specific study. In the second study, two extra learner-paced versions of the multimedia instructions were included in the experimental design. This time, the participants were 130 second-year students of educational science and the procedure was identical to the first study. With system-paced instructions, the audio group showed higher test scores than the visual-text group, but not with learner-paced instructions. Thus, it is concluded that the modality effect in multimedia learning only applies with system-paced instructions.

In Chapter 4, the interaction of modality with pacing is further investigated by testing the effect of prolonged time-on-task separately from the effect of learning pacing. Six versions of the multimedia instructions are compared differing in modality (visual text vs. spoken text) and pacing (system-paced, system-paced with extended time-on-task, learner-paced). Lower mental effort was expected for all conditions with spoken text, and higher transfer was expected only for spoken text in the system-paced conditions. Ninety-four second-year students of educational science got the lesson, made a retention and transfer test and rated their mental effort. As hypothesised, the spoken-text groups reported less effort than the visual-text groups. A significant interaction between modality and pacing was found on transfer, indicating an advantage of spoken text over visual text in the system-paced conditions, no differences in the extended conditions, and a reverse pattern in the learner-paced conditions. It is concluded that the modality effect only occurs with system-paced instructions, and that the explanation for the effect should be reconsidered.

The study in Chapter 5 discusses GazeTracker™, a tool for studying eye movements in dynamic multimedia environments and applies the tool in an experiment that studies the effect of modality and pacing on eye movement patterns in multimedia learning. The GazeTracker™ software links eye movement data to information about the internal computer processes and automatically combines the two for further analysis. The functionality of the tool is illustrated with an experiment on modality and pacing in multimedia learning. Twelve students from a Teacher Training College participated, and three versions of the instructions were tested in a within-subjects design: a system-paced audio version, a system-paced visual-text version, and a learner-paced visual-text version. The only differences found in fixation patterns between the different conditions were longer fixation durations and a lower frequency in the audio condition. It is concluded that the spoken text leads to a pattern that looks calmer. Finally, the suitability of GazeTracker™ is for eye movement research with dynamic interfaces like web browsers is discussed.

Chapter 6 provides a general discussion of the studies in this thesis. First, a short review of the results of the studies is given, and it is concluded that the modality effect can be generalised to longer instructions on a subject outside the exact sciences, as long as the instructions are system-paced based on the narration. However, as soon as pacing is extended, the modality effect disappears, and with learner pacing the effect even reverses. This has some implications for the theories of multimedia learning that are subsequently discussed. First, it is stated that the explanation of the modality effect in terms of the modality-specific subsystems of working memory should be replaced by an explanation in terms of preventing split attention. Moreover, it is suggested that the pacing of instructions and time-on-task should be included in the theories as relevant factors, just as the qualitative differences between reading and listening to a text are. Also, the following set of refined design guidelines for multimedia instructions is presented:

1. Get rid of redundant information
2a. With system-paced instructions, use spoken text
2b. With learner-paced instructions, use visual text
3. Prevent visual search

Finally, directions for further research are suggested and the chapter concludes with some final remarks.

References

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.