Dynamic visualizations in multimedia learning:

The influence of verbal explanations on visual attention, cognitive load and learning outcome.

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Philosophie des Fachbereiches Psychologie

der Justus-Liebig-Universität Gießen

vorgelegt von

Florian Schmidt-Weigand

aus Marburg

2005

Abstract

The main aim of this thesis was to take a closer look at visual attention allocation, cognitive load and learning outcome in learning from dynamic visualizations with accompanying verbal explanations. Instructional design guidelines derived from cognitive theories of learning with multimedia commonly recommend presenting spoken rather than written text in order to promote learning. Based on the existing evidences for the superiority of spoken over written text presentation five empirical studies were conducted to challenge the practical scope as well as the theoretical substantiation of this modality effect. In general, the studies raised two questions: (1) How do learners distribute their visual attention during learning from multimedia instruction? And (2) which design attributes *moderate* the effects of text modality on perception and comprehension?

The studies examined several design attributes that affect perceptual and cognitive processes in multimedia learning. In order to gain direct and objective measures of perceptual and cognitive *processes* during acquisition, learning outcome measures and indices of cognitive load were complemented by the previously unexploited method of eye tracking. The material applied in the studies was a multimedia explanation on the formation of lightning. Besides the modality of text presentation (Chapters 2, 3, and 4) the studies varied the spatial distance between written text and visualizations (Chapter 2, Experiment 1), the visualizations being animated or static (Chapter 2, Experiment 2; Chapter 3, Experiment 1), and the pacing of instruction (Chapter 3, Experiment 2) and its control by the learner (Chapter 4).

The results deliver converging evidence for an effect of text modality on cognitive load and learning outcomes under serious time constraints. However, under less attentional competition, less time constraints, and learner control of pace, these effects changed, decreased, or even disappeared. Once learners were relieved from following apparent motion or from time constrained presentation, the need to split visual attention lost much of its impact on learning. These “cognitive” effects were associated to particularities of the viewing behavior. Eye tracking measures revealed that visual attention allocation in learning from visualizations with accompanying verbal explanations follows a fairly stable pattern that was moderated by design attributes of the instruction. In general, written text dragged visual attention away from inspecting illustrations. Learners adapted to surface characteristics of the visual material (e.g. apparent motion in the visual field) and the presence and degree of time constraints by distributing their visual attention between written text and visualizations differently. Furthermore, they were able to adjust the pace of presentation to a regular reading strategy that only varies in the time taken to read text. Thus, the need to read written text may or may not interfere with extracting information from visualizations depending on how seriously reading and viewing visualizations are disturbed by the design of a multimedia instruction.

As a practical consequence, the question for an instructional designer is not that much if or if not text should be presented aurally instead of visually but if the displayed information can be sufficiently extracted by an individual learner. Understanding the demands of a learning material on the learner’s perception and accounting for individual differences by implementing user interaction appears promising to advance the design of multimedia instructions in a learner-supporting fashion.

Table of contents

Chapter 1

Theoretical introduction and general research questions

> “Before information can be stored (…), it must be extracted and manipulated in working memory.” (Paas, Tuovinen, Tabbers, & Van Gerven, 2003, p.64).

Successful learning requires extracting, manipulating and storing relevant information. From the very beginning teachers were concerned with how to supply relevant information to their students. And from the very beginning media have played a prominent role in supplying this information. Information has been presented as pictures and texts, stored on stone tablets, vellum, and paper, distributed as books, displayed on blackboards or with overhead projection, realized in television programs, and, most recently, digitally transformed to be applicable to computer technology. Each progress in the application of media has yielded the hope to facilitate learning (cf. Hegarty, 2004; Kozma, 1991). But the potential learning benefits of media employed to deliver instruction have equally often been called into question (e.g. Clark, 1983).

Without a doubt, however, with each technological progress the degrees of freedom for instructional design have grown – and so do the demands on the teachers. Especially the most recent progresses in computer technology changed the role of the teacher to one of an instructional designer: from somebody who *selects* appropriate media to supply information to somebody who can *create and combine* media for instructional purposes on a single device, the personal computer. Current computer technology allows more easily creating and combining different media using different codes and addressing different modalities. Consequently, the focus of instructional research and design has changed from *learning with media* to so-called *multimedia learning*.

Within the research on multimedia learning much attention is currently paid to the integration of concurrently presented information sources, namely verbal and pictorial information. A shift has taken place from earlier studies on those media-combinations to current research in two ways. First, older media research asked, if, how, under what conditions, and to what degree illustration can facilitate text understanding (for reviews see Levie & Lentz, 1982; Levin, Anglin, & Carney, 1987). Pictures were considered as an enhancement in learning from (mainly expository) texts. In multimedia learning illustrations gain more instructional potential. They are easier to build, more complex, and – most of all! –

can be *dynamic*. Second, multimedia research has taken advances in building theoretical frameworks that integrate different findings. These theories allow formulating comprehensive design guidelines and yield directions for further research.

In this chapter I will review these theoretical advances, introduce some prominent design guidelines, and formulate the main research questions of the thesis. These issues will be discussed in some detail here. To motivate the studies they are reconsidered in more detail in the following paper-style chapters. The main purpose of the studies conducted in this thesis is to consider the role of perceptual and cognitive demands in the concurrent presentation of expository text and visualizations. The studies complement research in multimedia learning with measures of attentional processes, namely the method of eye tracking to object visual attention allocation during learning. Before introducing this method and its use to further advance instructional design guidelines and its theoretical explanations I will outline the potential benefits and problems especially with dynamic visualizations in multimedia learning.

Comprehending dynamic visualizations

Dynamic visualizations are one of the most appealing applications in computer-based instruction. Most obviously, they help visualizing processes that are dynamic by nature. For example, animation has been used in instructing Newtonian mechanics (e.g. Kaiser, Proffitt, & Anderson, 1985; McCloskey & Kohl, 1983; Rieber, 1989; Rieber, 1990a; Rieber & Hannafin, 1988) and the functioning of mechanical devices like a car braking system (Mayer & Moreno, 1998) or a flushing cistern (Narayanan & Hegarty, 2002). Dynamic visual displays can also be applied to convey more abstract information, such as statistical concepts (Bodemer, Ploetzner, Feuerlein, & Spada, 2004), changes in population over time (Ainsworth & VanLabeke, 2004), or computer algorithms (Narayanan & Hegarty, 2002). In a review, Park and Hopkins (1993) specified six instructional conditions for using dynamic visualizations: (1) demonstrating sequential actions in a procedural task (e.g. procedures for operating or repairing equipment), (2) visually manifesting invisible system functions and behaviors (e.g. visualizations of the human cardiovascular system), (3) illustrating a task difficult to describe verbally (e.g. relational reactions occurring simultaneously among many different components in a complex system), (4) simulating causal models of complex system behaviors (e.g. a computer-simulation for piloting an airplane), (5) providing a visually motional cue, analogy, or guidance (e.g. displaying the trajectory of a thrown ball), and (6) obtaining attention focused on specific tasks or presentation displays (e.g. animating the most relevant features of a visual display). These conditions can be grouped into three broad classes of instructional functions of dynamic visualization: *presentation* (1, 2, 3, 4, 5), *attentional guidance* (5, 6) and *interaction* (4) (Rieber, 1990b).

The presentation function of dynamic visualizations: Does congruency help?

The instructional function of *presentation* is assumed to count for all kinds of graphical information. It rests on an implicit convention across cultures. From the very beginning space in graphical presentations was used to represent real space and to represent abstract concepts that suggest cognitive correspondence between mental spaces and real ones. These natural cognitive correspondences can be described in terms of a *Congruence Principle*: graphics are effective if the structure and content of the external representation correspond to the desired structure and content of the internal representation (cf. Tversky, Morrison, & Betrancourt, 2002, p.249). Applying this Congruence Principle, dynamic visualizations appear to be “a natural for conveying concepts of change” (Tversky et al., 2002, p. 250). Like space in graphics conveys spatial properties of the instructional content, changes in the visual display indicate changes in the illustrated fact. Surely, representing spatial properties is independent from the illustrations being static or animated. But dynamic visualizations are richer than static ones in that they also facilitate the visualization of changes over time.

Due to the opportunity to convey concepts of space, size, distance, change, motion, acceleration, etc. all in one display, one might be seduced to expect dynamic visualizations having an enormous impact on learning. The potential use for instructional purposes by a more accurate presentation of facts, however, lacks clear empirical support. Recent reviews (e.g. Park & Hopkins, 1993; Rieber, 1990b; Tversky et al., 2002) report at best inconsistent results. Among the numerous studies on the effectiveness of dynamic visualisations in conceptual learning very few have revealed an advantage over static visualizations (cf. Hegarty, 2004). Within this weak empirical support in favour of dynamic over static visualizations, many studies do not allow to infer a facilitatory effect on learning from a dynamic visual display *per se* because static and dynamic visualizations in these studies are often not informationally equivalent. This informational equivalence, however, is necessary to attribute facilitation to the way information is displayed (Larkin & Simon, 1987). Positive learning outcomes in these studies are attributable to more or different information visualized in the dynamic than the static case (e.g. Large, Beheshti, Breleux, & Renaud, 1996; Rieber, 1990a), or of superior study procedures such as feedback (e.g. Reed & Saavedra, 1986), or prediction (e.g. Hegarty, Quilici, Narayanan, Holmquist, & Moreno, 1999). A general advantage of dynamic over static visual displays due to a more natural presentation cannot be deduced from these studies. For the present, positive effects of dynamic visualization due to the presentation function are restricted to cases in which information cannot be presented otherwise (e.g. in an animation that shows a complex manner of motion where both spatial position and timing are essential). However, the widespread use of dynamic visualizations in current multimedia instructions imposes a question: Why do they fail?

Perceptibility of dynamic visualizations: Guiding visual attention

As Rieber (1990b) points out, “animation is often used with the intent to impress rather than to teach” (p. 77). And even if used for the best, concentrating on the exciting possibilities of current technology we are in danger of loosing sight of problems connected with an improper use of dynamic visualizations. While delivering congruence with concepts of change the visual information becomes more *transient*, thus generating demands on human perception and cognition that are not present with static displays. When viewing a static display, viewers can re-inspect parts of the display as frequently as they wish, using the external display as an external memory aid. In contrast, once a dynamic visual display has advanced beyond a given frame, it is no longer available to the viewer. This places heavy demands on working memory if information presented earlier in the visualization must be integrated with information that is presented later (Hegarty, 2004). Hence, dynamic visualizations may be difficult to perceive and understand due to perceptual and cognitive limitations in processing a changing visual situation. As a consequence, dynamic visualizations may be distracting, or even harmful, to conveying important ideas. In order to be comprehensible, dynamic visualizations have to be designed with caution. Congruent representation is not sufficient for an illustration to be effective. The structure and content of the representation must also be readily and accurately perceived and comprehended. Tversky et al. (2002) refer to this notion as the *Apprehension Principle.* Dynamic visualizations of events may be ineffective because they violate this principle. The dynamic visual display must be slow and clear enough for observers to perceive movements, changes, and their timing, and to understand the changes in relations between the parts and the sequence of events.

In order to ensure that the more transient information in dynamic (compared to static) visualizations is not missed or inaccurately apprehended it is necessary to properly *guide visual attention.* In fact, the potential to attract visual attention is probably the most recognized characteristic of dynamic visualization. Dynamic changes in the visual field are well known to capture visual attention, especially when they indicate a perceptual object (Hillstrom & Yantis, 1994; Yantis, 1998). Perceptual objects can be part of the visual representation itself (e.g. a cloud in an animated instruction on the formation of lightning storms) or a visual cue to some discrete part of the visualization (e.g. a moving arrow which directs attention to keywords or graphics). The animation of an object in a static background facilitates figure-ground perception, making the animated object more salient for visual perception (Blake, 1977). The attentional capture of motion and other dynamics can help or hinder instructional purposes of a visualization. In the depiction of a complex system, dynamic visual cues can be used to highlight critical features and their relations to other components, thus giving some “reading instructions” for the visual display by attentional guidance (e.g. Kalyuga, Chandler, & Sweller, 1999; Reitmayr, 2003; Tabbers, Martens, & Van Merrienboer, 2004). However, an improper use of dynamic visualization may even undermine the instructional goal. Perceptually salient aspects of a visual display that are not necessarily thematically relevant can misguide visual attention. And perceptually salient aspects in different spatial locations that compete for visual attention at the same time can further distract attention.

Another way to provide perceptibility of dynamic visualizations is navigational *interaction*. Stopping and replaying or control of speed by sequencing can allow (re)inspecting and focusing specific parts and actions (e.g. Schwan & Riempp, 2004). Actually, interacting with dynamic visualizations is more than navigating but also includes procedures like simulation and feedback (e.g. Reed & Saavedra, 1986; Rieber, Tzeng, & Tribble, 2004). Furthermore, facilitative effects of interactivity on learning are not restricted to dynamic visualizations. Thus, allowing manipulation of the visualizations itself, interactivity is even more likely to facilitate perception and comprehension of dynamic visualizations. Thus, simple navigational devices may already advance the use of dynamic visualizations for instructional purposes. In order to avoid problems confounded with navigation, the proper selection and design of such devices has yet to be investigated (Tversky, et al., 2002).

Attention vs. comprehension: The role of accompanying text

Even if dynamic visualizations are properly designed with careful attentional guidance and/or interactivity they are seldom displayed in isolation. In fact, most visualisations are accompanied by expository text. The role of text may change with the kind of visualization, but verbal explanations commonly provide an indication of how a visualization is to be understood. In most cases, expository text is even necessary in order to recognize the purpose and the (instructional) message of an illustration. Thus, although a picture may sometimes be worth a thousand words it may sometimes also be worth nothing without being explained by a thousand words.

Accompanying text is necessary because visualizations are usually not self-explaining. Compared to human language pictorial information is only weakly formalized. Language has a finite set of basic characters (phonemes/letters) from which the symbols (words) are constituted. The physical properties of the symbols are arbitrary, i.e. the structure of the symbol and the concept expressed by it are semantically connected by convention. For example, the word “dog” has neither in its written nor its spoken form any “dog-likeness”. This becomes especially evident if we compare the words “dog”, “chien”, “inu” and “Hund”. These words have physically not much in common and one needs to be familiar with the conventions of the English, French, Japanese, and German language in order to know that they all refer to the same concept: a domesticated carnivorous mammal, sometimes called “the man’s best friend”. Furthermore, language has explicit relational symbols (e.g. prepositions) and a finite set of production rules (syntax) to combine single words to sentences. These formal properties of human language allow to unequivocally describe general concepts of any degree of abstraction.

In contrast, visual depictions are essentially concrete. The very heart of the presentation function outlined above is the congruence between physical properties of the visualization and properties of the depicted concept. Space conveys concepts of distance, motion conveys concepts of change, etc. Thus, the symbols used in visualizations are semantically connected to the depicted concepts by structural similarities. The “meaning” of a symbol is visually emergent and thus more “natural” compared to the arbitrary connection between a word and its meaning. This is why visualizations are often assumed to be easier to understand. On the other hand, the syntactical relations between the constituents of a visual depiction lack explicit relational symbols. Even for more formalized visualizations like charts and graphs, verbal labels are almost necessary to express the relations of specific visual entities (Kosslyn, 1989).

As long as a certain type of visualization is not (explicitly or at least implicitly) formalized by some language-like conventions (e.g. statistical graphs) the general concepts depicted by it remain rather implicit. And so do the learning outcomes if visualizations are presented without any form of verbal explication. For example, Rieber et al. (2004) found that realistic simulations of Newtonian mechanics promote implicit learning, which enables students to learn to play a video game encompassing Newton’s laws. Conceptual understanding, however, was only promoted if the implicit experience of the simulation was accompanied by verbal explanations of the underlying physical principles. Thus, accompanying text is useful or even necessary to support the presentation function of visualizations and to ensure a proper understanding of the depicted concepts.

Besides helping to understand what a visualization actually represents, accompanying text can serve as a guide for visual attention. Usually, text added to a visualization is *descriptive*, i.e. the text explains the most important of the depicted concepts or may even be informationally equivalent. But captions can also be *instructive* in that they give explicit directions how to “read” the visualization. Bernard (1990) found that learners benefit from both descriptive and instructive captions compared to visualizations without any captions, confirming the positive effect of accompanying text in understanding a visualization. Even without explicitly explaining the depicted concepts, a “reading instruction” increases the value of a visualization. Somewhat surprisingly, however, the effects were not additive, i.e. learners receiving descriptive and instructive captions together did not benefit more than learners receiving either one of the captions alone.

Clearly, if descriptive captions affect the understanding of a visualization they must be assumed to affect the way the visualization is attended. In fact, Hegarty and Just (1993) found that also an informationally equivalent verbal description of a diagram can serve as a guide for visual attention. In their study the authors exposed participants to depictions of pulley systems, informationally equivalent verbal descriptions or both. First of all, they found that learners benefit from the information in both the text and the diagram during learning. On subsequent tests of comprehension participants receiving a combined text-and-diagram description outperformed participants receiving either one of the information sources alone. This result is another confirmation of the positive effect of accompanying text in understanding a visualization (or vice versa). Furthermore, in order to investigate how learners integrate the verbal and pictorial information, the authors tracked participants’ eye movements in the combined text-and-diagram descriptions. The fixation patterns revealed that participants attended to text and visualization in a highly systematic manner. Most obviously, participants started the inspection with reading text. This reading was interrupted several times to inspect the diagram. The diagram was primarily inspected at the ends of clauses and sentences, checking or elaborating the representation of this clause by attending to the referential part in the diagram. Most of the clauses preceding a shift towards the diagram typically stated a configural or kinematic relation between two components. Thus, participants inspected the diagram to encode relations between components rather than characteristics of individual components. Since the diagram inspection typically focused on the referents of the preceding reading episode, the authors conclude that diagram inspection is largely text directed. Other eye tracking research confirmed that for example labels and captions in a multimedia presentation (Faraday & Sutcliffe, 1996) and verification questions about a picture or diagram (e.g. Hegarty, 1992a; Underwood, Jebbett, & Roberts, 2004) affect the way a visualization is attended.

Taken together, accompanying text can serve as a device to overcome difficulties in the perception and comprehension of visualizations. However, one can easily imagine that accompanying text especially in the case of dynamic visualizations also causes further problems. The referenced eye tracking studies indicated that written text is a highly salient stimulus for visual attention allocation. Written text and dynamics in the visual display (e.g. visual motion) may compete for visual attention allocation. Furthermore, while reading text some of the transient visual information in a dynamic display may be missed. Thus, in order to promote learning, text presentation must be treated with caution. The next section will provide some guidelines for text presentation in multimedia learning in order to prevent problems associated with accompanying text.

Guidelines for the combination of text and visualization

One goal of instructional research in multimedia learning is to figure out how the combined presentation of text and visualization must be designed in order to promote learning. Currently, there are two prominent recommendations how to combine (expository) text with (dynamic) visualizations: The *modality principle* and the *spatial contiguity principle* (Mayer, 2001). The modality principle states that it is more beneficial for learning if text in simultaneous presentation with illustrations is presented aurally rather than visually. The spatial contiguity principle states that learning is promoted if written text is presented physically close to an illustration. Note, that these recommendations are not restricted to dynamic visual displays but claim to be effective for all kinds of instructional visualizations.

These guidelines can be seen as applications of the Apprehension Principle. They are thought to avoid a split of visual attention between textual and pictorial information (Sweller, Van Merrienboer, & Paas, 1998). According to the Apprehension Principle, text can effectively help understanding a visualization only if the connection between verbal and pictorial information is readily and accurately perceivable. For the case of spatial contiguity, an integrated presentation of written text lowers the need for visual search and shortens the time to keep information elements actively represented. If text is presented aurally rather than visually, as requested by the modality principle, there is no need to split visual attention at all. The learner can inspect a visualization without ruffle while listening to accompanying verbal explanations. However, to lower the need of visual search for appropriate referents, the referential connections between a visualization and its verbal explanation must be emergent in the learning material.

Both the modality and the spatial contiguity principle are empirically well supported. A number of studies have found superior learning results when text in a multimedia instruction was presented in spoken rather than written form (e.g. Brünken & Leutner, 2001; Kalyuga et al., 1999; Kalyuga, Chandler, & Sweller, 2000; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi, Low, & Sweller, 1995; Tindall-Ford, Chandler, & Sweller, 1997). However, the modality effect cannot be achieved when the referential connections of spoken text to a visualization are not readily perceivable. In one study with pictures of high complexity spoken text only supported the understanding of a visualization when visual cues were added (Jeung, Chandler, & Sweller, 1997).

In support of the spatial contiguity principle, several studies have shown that learning is improved if split attention is prevented by placing written text elements next to the corresponding parts of a visualization (Chandler & Sweller, 1991; Chandler & Sweller, 1992; Mayer, 1989; Mayer, Steinhoff, Bower, & Mars, 1995; Moreno et al., 1999; Tindall-Ford et al., 1997). As for the modality principle, other visual cues in the written text and/or the visualization that explicate the correspondences between verbal and pictorial information have proven to be effective in order to (further) promote learning (Beck, 1984; Kalyuga et al., 1999; Reitmayr, 2003; Tabbers et al., 2004).

Certainly, perceptibility is a first necessary condition in the proper design of multimedia learning material. However, it is not clear if facilitated perception is sufficient to explain the referenced effects of modality and spatial contiguity of text presentation on learning. The next section will consider some broader theoretical approaches that have been pursued in order to provide a coherent framework for design guidelines for multimedia learning that also integrate the proposed principles.

Cognitive theories of learning in instructional areas

As implied in the previous sections, for a long time the design of instructional learning material was driven by an ever-new excitement about the potential of technological advances. As a consequence, also the search for effective guidelines in the field of multimedia learning was pushed by technical developments rather than theoretical considerations. More recent theoretical advances and, thus, the development of guidelines are based on what is known about human cognitive architecture. Currently, research on multimedia learning and instructional design rests on two theoretical frameworks, cognitive load theory (Sweller, 1988; Sweller, 1999; Sweller et al., 1998) and Mayer's cognitive theory of multimedia learning (Mayer, 1997, 2001). Both theories offer similar explanations for the above mentioned instructional design principles. The theories will be described in some detail here and reconsidered in the following chapters to motivate the particular research questions.

The role of working memory

The most central concept of human cognitive architecture in both, the cognitive load theory and the cognitive theory of multimedia learning, is *working memory*. The central role of working memory for the matter of understanding and learning stems from the assumption that, simply stated, working memory is the gateway between the external world and the existing internal cognitive entities. Meaningful learning requires the learner to select relevant information, to organize that information in a coherent structure, and to integrate this structure into existing knowledge. Working memory plays an essential role since it is here, where the selection, organization, and integration processes are assumed to take place.

Among the various models and theories of working memory (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, apart from a central regulation system, two or more separate modality-specific subsystems. Concerning the limitation of working memory, the derivation of meaningful information from learning material can be described by the following prominent metaphor: “Understanding *is* the management of [limited] working memory [resources]” (Graesser & Britton, 1996, p. 348). For example, within the issue of text comprehension this metaphor has long been recognized and some of the main predictions derived from the notion that working memory has capacity limitations have been confirmed in empirical studies (e.g. Just & Carpenter, 1992; but see Waters & Caplan, 2004 for a different view).

The notion of separate modality-specific subsystems comes into play in the explanation of effects of text modality in multimedia learning. Both, the cognitive load theory and the cognitive theory of multimedia learning rest on the crucial assumption that the presentation format affects the efficiency of the integration processes in working memory. Thus, in order to derive instructional design principles the theories need to specify how much of the limited capacity of working memory is taken up by a particular presentation format. The considered theories slightly differ in the way they conceptualize the limitations of working memory and will, thus, be discussed successively.

Cognitive load theory

Cognitive load theory (Sweller, 1988; Sweller, 1999; Sweller et al., 1998) provides a framework to integrate several findings in the research on instructional design. It has been designed to provide guidelines for the presentation of information to optimize intellectual performance. The theory rests on two assumptions: an effectively unlimited long-term memory, holding schemas of varying degrees of automation, and, as outlined above, a working memory of limited capacity with sensory-specific subsystems for visual and auditory information. The limitations of working memory are described in terms of a cognitive workload that depends on several learner and task characteristics.

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 differentiates between *intrinsic* and *extraneous* cognitive load. Intrinsic cognitive load refers to the load caused by the content of a learning material. It is determined by an interaction between the nature of the material and the expertise, prior knowledge, and cognitive abilities of the learner. In this respect, intrinsic cognitive load is the basic amount of processing required to understand an instruction. Extraneous cognitive load refers to the presentation format of the material. Extraneous load is what can be affected by manipulating instructional design. In terms of cognitive load theory instructional design is concerned to keep the overall cognitive load within working memory limits. Thus, one aim is to reduce extraneous cognitive load, i.e. to minimize the capacities required to successfully encode all relevant information. If the limits of working memory are not exhausted in a particular context, one might also encourage learners to invest extra effort in processes that are directly relevant to learning, such as schema construction. This process also increases cognitive load. To contrast this kind of cognitive load from the undesirable extraneous cognitive load, Sweller and his colleagues refer to this load as *germane* cognitive load that contributes to, rather than interferes with, learning.

Cognitive load theory offers an explanation for the modality principle introduced in the previous section. The theoretical rationale rests on the assumed subdivision of working memory. According to Baddeley's working memory model (Baddeley, 1986) visual information is processed in a "visuo-spatial sketchpad", auditory information is processed in an "auditory loop". Both systems have partly independent limited processing capacities. Effective working memory capacity can be *increased* by using both visual and auditory working memory rather than either memory stream alone. Although less than purely additive, there seems to be an appreciable increase in capacity available by the use of both, rather than a single, processor. As indicated by the vast experimental psychological literature on this topic (for a review, see Penney, 1989), many effects of text modality seem to rest on this fundamental characteristic of working memory. Thus, it can be assumed to also come into play in more complex instructional learning material. In terms of cognitive load theory, spoken and written text presentations cause different levels of cognitive load. If expository text is added to a visualization in written form, both materials have to be processed by the visual processing system. Under these conditions, an overload of the visual system is more likely to occur compared to spoken text presentation. If text is spoken rather than written, less information needs to be processed in the visual system while the processing of verbal information only requires capacity of the auditory system.

The risk of cognitive overload when text is presented in written form can be described by what Sweller and his colleagues call split-attention. It occurs whenever two or more sources of information must be processed simultaneously in order to derive meaning. If different sources of *visual* information are physically separated, one source must be held active in the visual system until the corresponding source is found and integrated. The more information must held active or the more capacity is needed for the search task the more likely it is that a cognitive overload occurs. Due to the spatial contiguity principle discussed earlier, this problem can also be reduced by physically integrating the disparate sources of information. Within cognitive load theory, this procedure may be considered to be just as effective in facilitating learning as presenting verbal material in auditory and pictorial material in visual form. In this view, effects of text modality derive from split-attention (cf. Sweller et al., 1998, p. 282).

Cognitive theory of multimedia learning

Mayer's cognitive theory of multimedia learning (Mayer, 1997, 2001) is similar to the cognitive load theory in its basic assumptions and in the resulting instructional design guidelines. In fact, both theories support the modality as well as the spatial contiguity principle. And both, Mayer and Sweller and his colleagues, refer to the working memory model of Baddeley. However, the theories slightly differ in how they conceptualize the *entities* processed in working memory. In contrast to the cognitive load theory, where the visual and auditory subsystems of working memory are closely related to what is actually presented to the sensory modalities, the subsystems in Mayer's theory are only in a first step associated with the modality of presentation in a so-called sensory memory. For the actual working memory the cognitive theory of multimedia learning postulates different internal *information codes*. With reference to dual-coding theory (Paivio, 1986) Mayer assumes that information can be stored verbally or pictorially. What is processed in working memory is not auditory or visual information but *verbal* or *pictorial representations* of information. That is, visualizations are transformed into a pictorial representation of the learning material in a subsystem that is responsible for building up a pictorial mental model. And text is transformed into a verbal representation in a subsystem for building up a verbal mental model of the content. That implies that written and spoken text is initially processed in different channels, but is subsequently represented in the same verbal system. The verbal and pictorial processing channels are, in accordance with cognitive load theory again, severely limited in their capacities.

In terms of the cognitive theory of multimedia learning, learners receiving a visualization with accompanying text construct a verbal and a pictorial mental model and build connections between these models. By referring to dual-coding theory, Mayer provides a general account for the utility of visualizations for instruction. According to dual-coding theory, information that is stored verbally and pictorially is easier to recall than information that is stored in either one of the codes alone. Thus, multimedia-based presentation of information is supposed to promote learning because it allows to more easily construct a verbal *and* a pictorial model than if words (or pictures) are presented in isolation. Mayer refers to this account as the multimedia principle: “Students learn better from words and pictures than from words alone” (Mayer, 2001, p. 63).

However, the crucial aspect of the theory for the purpose of this thesis is that meaningful learning from visualizations with accompanying text can only occur, if both, verbal and pictorial representations are present in working memory at the same time. The design guidelines for the concurrent presentation of text and visualizations that can be derived from this notion are the same as for the cognitive load theory. In fact, the terms “modality principle” and “spatial contiguity principle” used in the previous sections were originally introduced by Mayer (2001). The theoretical accounts for these recommendations offered by Mayer are similar to the explanations given by the cognitive load theory and do not refer to verbal and pictorial information codes. The rationale for the spatial contiguity principle, as formulated by Mayer, is that physical proximity of corresponding words and pictures lowers the need for visual search: “When corresponding words and pictures are far from each other on the page or screen learners have to use cognitive resources to visually search the page or screen for corresponding words and pictures. Thus, learners are less likely to be able to hold them both in working memory at the same time.” (Mayer, 2001, p. 81). Also the rationale for the modality principle does not explicitly need different internal codes: “When pictures and words are both presented visually, the visual/pictorial channel can become overloaded but the auditory/verbal channel is unused. When words are presented auditorily, they can be processed in the auditory/verbal channel, thereby leaving the visual/pictorial channel to process only the pictures.” (Mayer, 2001, p. 134).

In the formulation of the theory Mayer distinguishes between presentation modalities (visual, auditory) and presentation codes (pictorial, verbal), which are sometimes confused (Weidenmann, 2002). Thus, one might feel somewhat uncomfortable to equate a verbal channel with an auditory channel and to equate a visual channel with a pictorial channel as happened in the rationale for the modality principle. In fact, research on discourse comprehension proves the equation of verbal and auditory channels to be inappropriate (e.g. Graesser, Millis, & Zwaan, 1997). Admittedly, modalities and codes are not completely separated since only verbal but not pictorial codes can be presented to the visual and auditory modalities. But if the “visual/pictorial” channel is overloaded by written text, is it appropriate to call it a “pictorial” channel at all? Although one can conceptually distinguish modalities from codes it may be difficult to find a distinct boundary between perceptual processing and processing of an item in a code-specific short-term memory (cf. Penney, 1989, p. 399). However, the distinction between a sensory memory and a code-specific working memory points to a critical aspect of both theories. Although claiming to give theoretical accounts for instructional design based on the cognitive architecture, the recommendations for text presentation refer to limitations for processing information presented to the visual *modality*. Thus, the core of text presentation effects in multimedia learning may not be a limited cognitive process but a matter of perceptibility as expressed by the Apprehension Principle. Matter-of-factly, the observations taken to investigate effects of instructional design mainly concentrate on cognitive and not perceptual processes. The next section will provide some considerations, how the set of observations may be expanded to get a more detailed impression of the cognitive and perceptual processes a learner engages in while learning with multimedia instructions.

Measuring cognitive and perceptual processes

The cognitive frameworks described in the previous section provide theoretical accounts of instructional design principles for concurrent presentation of visualizations and verbal explanations. In order to test for the validity of these accounts we need to inspect the cognitive and perceptual processes claimed to emerge during learning with such material. The most common observations taken in the research on instructional design are rating scales (e.g. self-reported mental effort or subjective ratings of difficulty of materials) and task performance (e.g. learning outcome measures). The measures derived from these observations differ in their causal relation to the cognitive processes under inspection and with respect to their objectivity. Clearly, rating scales are essentially subjective while task performances usually suffice the requirements for objectivity. The causal relation between a measure and the assumed underlying cognitive processes is less obvious and depends on the process model. Independent from the theoretical model, however, any *subsequent* task performance is obviously only an indirect indication of the cognitive *processes* involved during acquisition. Thus, although learning outcomes are surely the most important measures of the actual effects of instructional design on learning success, they are connected to cognitive processes only by the predictions derived from a cognitive model. More direct observations of cognitive processes can be accomplished by introspection. For example, self-ratings of cognitive load have proven to be a reliable measure, i.e. people are able to introspect on their cognitive processes and have no difficulty giving a numerical indication of their perceived cognitive load (Gopher & Braune, 1984). It appears as if the research on cognitive processes underlying multimedia learning is trapped. The applied observations are either direct but inherently subjective or they are objective but give only an indirect access to cognitive processes.

Direct measures of cognitive load

In order to advance theoretical approaches current research in instructional learning is concerned to complement traditional observations with direct *and* objective measures of cognitive processes. Especially the measurement of cognitive load has achieved reasonable progress (Brünken, Plass, & Leutner, 2003; Paas et al., 2003). Brünken and his colleagues classified the currently available methods for assessing cognitive load along the two dimensions of causal relation and objectivity. Besides rating scales and subsequent task performance the authors discuss dual-task methodology, physiological measures (e.g. heart activity and eye activities), and neuroimaging techniques (e.g. fMRI).

Clearly, observing which parts of the brain are active during executing cognitive tasks (e.g. word memorization, sentence comprehension, visual rotation) delivers direct and objective measures of the amount and neural localization of cognitive processes. However, for the study of complex learning processes “the connection between memory load and prefrontal cortex activity is not yet fully understood” (Brünken et al., 2003, p. 56). Furthermore, practical considerations of neuroimaging techniques in multimedia learning call the ecological validity of the learning situation into question. The measurement apparatus is technically complex and makes its use difficult in authentic learning situations.

Another direct and objective observation that is more closely related to cognitive load and already well settled in working memory research is offered by the dual-task-paradigm. A basic assumption in the working memory model of Baddeley (1986) is that the limited resources of working memory can be distributed between simultaneous tasks. If a learner has to perform two tasks that require the same working memory resources, then the cognitive load caused by one of the tasks will directly affect the performance of the other task. Dual task methodology is known to deliver highly sensible and reliable measures of cognitive load. But although the cognitive load theory relies on Baddeley’s working memory model, dual tasks have been applied in only few studies on complex learning (e.g. Brünken, Plass, & Leutner, 2004; Brünken, Steinbacher, Plass, & Leutner, 2002; for a recent review see Paas et al., 2003). The rare application of secondary tasks in complex learning scenarios may be explained by its measurement logic. Dual tasks are intended to introduce a competition for resources. This competition undermines the ecological validity of the primary learning task. In an instructional setting one would not intentionally accompany a learning task by tasks irrelevant for the learning issue. Furthermore, since the cognitive load is measured by the interference between primary and secondary task dual-task-methodology does not allow estimating the actual cognitive load evoked by the learning task alone.

Less interfering with the learning situation than brain imaging techniques and dual tasks are physiological measures. The theoretical rationale for these techniques is that changes in physiological variables reflect changes in the cognitive functioning (Paas et al., 2003). Recent research applying measures of eye activity identified pupillary dilation and blink rate to correlate with fluctuating levels of cognitive load (Van Gerven, Paas, Van Merrienboer, & Schmidt, 2004; Van Orden, Limbert, Makeig, & Jung, 2001). Applying those objective measures in multimedia learning, however, we are once more trapped. In the case of cognitive load, eye activity measures are only an indirect indicator of cognitive processes in working memory. They may as well be a function of attentional and motivational factors associated with the learning process (Brünken et al., 2003).

Eye tracking in multimedia learning

One prominent measure of eye activity that has not been considered yet is *eye tracking*. Concerning the concept of cognitive load, eye tracking is considered to be only an *indirect* measure (Brünken et al., 2003). But although eye tracking does not provide a single numerical indication of the cognitive load a learner experiences, observing a learner's viewing behavior can nevertheless help fulfilling the claim for more direct *and* objective measures of cognitive and perceptual processes during multimedia learning. According to the so-called eye-mind assumption (Just & Carpenter, 1980) fixation or gaze durations map onto the amount of cognitive activity associated with the fixated area of a stimulus. Even in a less restricted version of this assumption it is very likely that visual information is not perceived and, thus, processed until it is fixated. The visual area perceived within a single fixation covers 5° of visual angle. With an approximate distance of 50 cm from the visual information (e.g. on a computer screen) this angle corresponds to the size of a 2 Euro coin. That is, the amount of visual information that can be perceived within a single fixation is severely limited. Due to this limitation of the eye people retrieve visual information by quickly moving the point of regard (i.e. the fixation) over the visual material. As a consequence, besides the visual attention and/or the amount of cognitive resources devoted to discrete areas of visual information, fixation patterns can reveal the time course of attending, perceiving and processing visual information.

A large body of experimental research in cognitive psychology has applied measures of eye movement. In particular, eye movement studies in the areas of reading and picture perception have generated a good understanding of the processes involved (for reviews, see Rayner, 1998; Underwood, 1998). Actually, a few eye movement studies have already addressed the extraction of information from *combinations* of words and pictures. In a previous section I outlined the work of Hegarty on the comprehension of mechanical diagrams (Hegarty, 1992a, 1992b; Hegarty & Just, 1993). Other studies investigated the visual analysis of cartoons (Carroll, Young, & Guertin, 1992), visual attention allocation in subtitled television (for an overview see (d'Ydewalle & Gielen, 1992)), and the integration of text and pictorial information in print advertisements (Rayner, Rotello, Stewart, Keir, & Duffy, 2001). Some of these studies will be reconsidered later in context to the conducted experiments.

Given the widespread use of eye tracking in the study of (visual) cognitive processes it appears quite logical to apply this method also in learning from multimedia instructions. Up to now, however, viewing behavior has rarely been considered in multimedia learning (for exceptions, see Faraday & Sutcliffe, 1996; Tabbers, Paas, Lankford, Martens, & Van Merrienboer, 2002;). This is astonishing since both cognitive theories outlined in the previous section incorporate visual processes and stress on the importance of these processes in multimedia learning. In fact, the theoretical explanations for effects of text presentation format in multimedia instructions refer to limitations of the visual system. Eye tracking offers an attempt to directly explore these limitations. According to the cognitive load theory, concurrent presentation of written text and illustration causes a split of visual attention between both information sources. Eye tracking can reveal, how visual attention is split between written text and illustrations and how much attention and/or processing resources are devoted to each of the information sources. Furthermore, both, the cognitive load theory and the cognitive theory of multimedia learning, claim visual search to be a source of cognitive load. Also the amount of visual search may become emergent in particular fixation patterns.

With reference to the Apprehension Principle outlined above, problems with dynamic visualizations and text presentation formats in multimedia learning may arise from constraints on the material's perceptibility rather than limitations of cognitive resources. Recalling the introductory quotation, information must be extracted and manipulated in working memory before it can be stored (Paas et al., 2003, p.64), but note that it also must be extracted *before* it can be manipulated. Matter-of-factly, the observations taken to investigate effects of instructional design mainly concentrate on cognitive and not on its presumably preceding attentional and perceptual processes. Eye tracking complements the set of observations, allowing to investigate these attentional and perceptual processes. In the remainder of this introductory chapter I will consider how observing these processes may help gaining a better understanding of learning from dynamic visualizations and the format of verbal explanations in multimedia learning.

Aim of the thesis: A closer look at effects of text modality

The aim of my thesis is to examine which characteristics of the learning material affect visual attention allocation, cognitive load and learning outcome. The basis for the research questions is provided by the design guidelines for concurrent presentation of text and visualizations and their theoretical explanations. Although the considered cognitive theories offer explanations for effects of text presentation on the *comprehensibility* of multimedia learning material they do not explicitly specify the influence of its *perceptibility*. However, both, the cognitive load theory and the cognitive theory of multimedia learning, refer to characteristics of the visual sensory system that are not necessarily cognitive. Thus, the studies conducted in this thesis will take a closer look not only on cognitive but also on perceptual effects of multimedia presentation formats.

Especially in the cognitive theory of multimedia learning the sensory modalities are described as a gateway between the learning material and further cognitive processing in the (code-specific) subsystems. This gateway may serve as a “bottleneck” for retrieving subsequently processed information. According to the referenced cognitive theories different units of information must be held active in working memory at the same time in order to become integrated. The smaller the bottleneck the longer some information units must be held active in working memory. However, the capacity limitations of working memory may not be exhausted and still problems in instructional design occur due to limitations of the visual system. In a dynamic multimedia presentation some information may not pass the bottleneck to enter working memory, i.e. they are simply missed. According to the Apprehension Principle, instructional design can only be effective if relevant information is readily and accurately perceivable. In order to explain effects of text modality the *perceptual* split of visual attention between written text an illustration may be sufficient to cause learning problems without referring to further cognitive processes. Thus, the core of text presentation effects in multimedia learning may not be a limitation in cognitive processing but a matter of perceptibility.

Note, that I do not ask for the validity of the modality effect itself but for the appropriateness of its explanations offered by cognitive theories. If the modality effect is a matter of perceptibility, there may be other design options to overcome difficulties with written text presentation than presenting text in spoken form instead. In order to specify how different attributes of multimedia instructions interfere with the presentation format of verbal explanations five empirical studies on the modality effect in multimedia learning were conducted that are presented in the following chapters. The studies are designed to investigate which attributes of a multimedia instruction *moderate* the modality effect. Besides the modality of text presentation the studies vary spatial properties of written text presentation, the design of illustrations, and the pacing of instruction and its control (by the learner). Common measures of learning outcome and cognitive load are complemented by the method of eye tracking. Objecting the learners' viewing behavior during acquisition provides access on how visual attention allocation is managed during multimedia learning especially for cases when visual attention has to be split between visualizations and accompanying written text.

Chapter 2 introduces eye tracking as a method previously unexploited in the context of multimedia learning. In order to check the general applicability of measures of viewing behavior, the fist study replicates a study on effects of modality and spatial contiguity of text presentation conducted by Moreno and Mayer (1999, experiment 1). The material applied in this as well as the following studies is a redesign of a multimedia explanation on the formation of lightning, originally used by Mayer and Moreno (1998; Moreno & Mayer, 1999). The main question of the first study is: How do design attributes of text presentation (written versus spoken text, integrated versus separated text) affect viewing behavior? In general, written text is attended first and gains relatively more visual attention than illustrations no matter if text is integrated to or separated from visualizations. Consequently, illustrations are inspected much longer whenever text is spoken rather than written. The second study of this chapter asks if this viewing behavior and the subsequent learning success are moderated by the salience of illustrated information. Orthogonal to text presentation format the study varies elements of the visualization as being animated or static. Performance measures of both studies deliver converging evidence for a modality effect that is restricted to animated illustration in the applied learning material but fail to replicate a spatial contiguity effect. Learning success is discussed in terms of visual attention allocation. It is suggested that the amount of time that can be spent reading and inspecting illustrations is a major source of differences in subsequent learning outcomes. In sum, learning success can be causally related to managing the attentional split between written text and animated illustration.

Chapter 3 presents two studies investigating how measures of viewing behavior relate to the actual cognitive load during multimedia learning. The first study applies several dependent measures of cognitive load and task load while using the same experimental variation as in the second study of the previous chapter. The main question is how viewing behavior relates to the learners' perceptions of the instructional material. The high salience of written text, confirmed in this study, indeed appears to drag visual attention away from illustration. Furthermore, the (in-)appropriateness of pacing turns out to be a major source of cognitive load in written text presentation. Thus, the second study of this chapter asks how the modality effect and the learners' viewing behavior vary with pacing of instruction. This study reveals that the modality effect can be described as a distracting effect of written text under serious time constraints.

In Chapter 4 a closer look is taken at the role of time-on-task for the modality effect. The study in this chapter introduces minimal learner control to the multimedia instruction applied in the previous studies. The main question is if the modality effect is a mere matter of time-on-task or if a qualitative change occurs from system- to learner-controlled instruction. Having control over the pace of instruction, learners are able to adjust the presentation in order to avoid cognitive overload and to gain a comparable learning success no matter if text is spoken or written. Time-on-task does not reveal an effect of text modality. To understand the lack of any modality effect in this study, the learners' viewing behavior is compared with observations taken in the second study of Chapter 3. Indeed, a qualitative change takes place from system- to learner-controlled instruction, expressed in the different relations of time on text and illustrations to total self- or system-controlled presentation time.

The final Chapter 5 recapitulates all five studies. The results are reviewed with respect to their theoretical and practical implications. On the basis of these considerations design recommendations for multimedia instructions will need to be specified more precisely. Furthermore, the general discussions will summarize suggestions for further research.

Chapter 2

Integrating different sources of information in multimedia learning: correspondence between viewing behavior and comprehension

In two experiments students' eye movements were recorded during presentation of a multimedia instruction on the formation of lightning and learning outcomes were measured in subsequent performance tasks. In Experiment 1 text was presented either spoken, written near or far from animated illustration. Participants receiving spoken text outperformed those receiving written text in retention and transfer. Superiority of near over far presentation of written text failed statistical significance. Participants spent less time inspecting illustrations if text was written and alternated between text and illustrations. Text was read first and gained more visual attention than illustrations. Experiment 2 varied text presentation (spoken, written) and illustration format (animated, static). Participants showed a better visual memory when text was spoken. For retention poorer performance was restricted to animated illustration. Viewing behavior replicated the results of Experiment 1. Learning outcomes are discussed in terms of visual attention allocation.

Introduction

In multimedia learning environments we are constantly forced to extract and integrate information from different information sources like words and pictures. Research in instructional design examines two major questions: How does the presentation format of information affect knowledge acquisition? And how should the combination of different sources be designed in order to promote learning? For example, a number of empirical studies have investigated whether and to what degree the modality (Brünken & Leutner, 2001; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi, et al., 1995; Tindall-Ford, et al., 1997) and spatial properties (Chandler & Sweller, 1991; Mayer, 1989; Mayer, et al., 1995; Moreno & Mayer, 1999; Sweller & Chandler, 1994; Sweller, Chandler, Tierney, & Cooper, 1990; Tindall-Ford, et al., 1997) of text presentation can foster multimedia learning, and under which conditions animation is a helpful characteristic of illustration (Narayanan & Hegarty, 2002; Zuberbuehler, 1999; for reviews, see Park & Hopkins, 1993; Rieber, 1990b; Tversky, et al., 2002). Cognitive theories of multimedia learning (e.g. Mayer, 2001; Sweller, 1999) do offer explanations on the level of real time information processing. However, attentional, perceptual, and cognitive demands of the instructional material are mostly *inferred* from learners' performance on subsequent tasks or self-reported difficulties with the materials at hand. In order to advance theoretical approaches and to refine instructional design principles it is necessary to complement these subjective or indirect measures with more direct process measures (Brünken, et al., 2003; Paas, et al., 2003). An often suggested, well suited albeit seldom used measure in multimedia learning is the observation of viewing behavior. Applying the method of eye tracking the studies presented in this chapter address two issues: (1) How do design attributes in multimedia learning environments (i.e. written vs. spoken text, integrated vs. separated text, and animated vs. static illustration) influence viewing behavior? (2) And how does viewing behavior correspond to learning outcomes?

Eye movements and other process measures in settings with multiple information sources

Eye movement studies have generated a good understanding of the processes involved in reading and picture perception (for reviews, see Rayner, 1998; Underwood, 1998). Surprisingly, only few eye movement studies have addressed the extraction of information from *combinations* of words and pictures. Notable exceptions are studies from Hegarty on the comprehension of mechanical diagrams (Hegarty, 1992a, 1992b; Hegarty & Just, 1993), Carroll, et al. (1992) on the visual analysis of cartoons, d'Ydewalle and colleagues on attention allocation in subtitled television (for an overview see d'Ydewalle & Gielen, 1992), and more recently Rayner, et al. (2001) on the integration of text and pictorial information in print advertisements. The studies varied tasks (e.g. sentence verification of static or dynamic aspects of a mechanical diagram; Hegarty, 1992a), individual factors (e.g. high vs. low spatial ability (Hegarty & Just, 1993) or familiarity with subtitles (d'Ydewalle, Praet, Verfaillie, & Van Rensbergen, 1991)), content of the material (d'Ydewalle & Gielen, 1992; Rayner et al., 2001), and participants' goals (Rayner et al., 2001).

The only study I found using eye movement measures to evaluate *instructional learning material* was conducted by Faraday & Sutcliffe (1996). They tracked eye movements while participants watched a 27 sec. animation sequence taken from a multimedia presentation for medical education (on the 'Etiology of Cancer'). Viewing behavior was qualitatively described in fixation sequences aggregated over participants. The question was in which order participants attend visual information. The authors differentiated between text captions, labels, and still or moving objects. Exploring the fixation paths they found that most of the visual attention was directed to moving objects and written text. The onset of an animation produced an attentional shift towards the object in motion. Sometimes, however, visual attention was "locked" by text elements like labels or captions. After scene changes it took some fixations to re-orientate attention.

The findings are in accordance with the eye movement literature referenced above. Studies on combinations of text and pictures consistently found a high attentional salience for written text. Text is commonly read before accompanying pictures are inspected (Carroll et al., 1992; Hegarty, 1992a; Rayner et al., 2001). And even when written text is accompanied by a dynamic visual display as in subtitled television a considerable amount of time is spent on reading (d'Ydewalle et al., 1991).

Faraday and Sutcliffe conclude that captions, labels, animations, and scene changes can easily overload attention, resulting in presentation elements being missed. They recommend to design instructional material in a way that allows shifts of attention according to the content being shown, i.e. providing time to read captions and labels, using animation to guide attention to important elements of the instruction, and pause for re-orientation when scenes are changed. Although some characteristics of the viewers' scan paths support these recommendations the observations and conclusions remain qualitative. We still do not know, how much time it takes to comprehend captions or labels, how much of a text is possibly missed by paying attention to animations and how long it takes to re-orientate after scene changes. Moreover, eye movement studies have not systematically varied presentation format so far in order to investigate the allocation of attention on illustration as a matter of presenting written or spoken text and to explore the influence of static vs. animated illustration on reading.

An approach to investigate the influence of instructional design attributes on visual attention was taken by Brünken, et al. (2002). In a pilot study, Brünken and his colleagues used dual-task methodology to assess the amount of cognitive resources occupied by different text presentation formats. They presented two multimedia computer-based training (CBT) programs (about the human cardiovascular system and the historic city of Florence) that contained expository text and closely related illustrated information. Presentation of text alternated within-subjects between written and spoken format. In addition to learning from the CBTs participants were asked to perform a secondary visual observation task. Response time on the secondary task varied as a function of text presentation format. The visual secondary task was performed slower if text was written rather than spoken. The results show that the cognitive resources of the visual system are directly influenced by presentation format. More visual attention is captured by learning material if both, verbal and pictorial information is visual compared to spoken verbal information. Thus, the observation of visual secondary task performance reveals an effect of modality of text presentation on visual attention. However, with respect to a closer linkage of design attributes with knowledge acquisition processes the study has some shortcomings. The differences in performance on the secondary task were not compared to learning outcomes for each presentation condition separately. Thus, we do not know how different attentional demands correspond to learning performance. Furthermore, a secondary task is very likely to interfere with the learning task since it puts an additional load on the learner. In contrast, measuring eye movements is not expected to interfere with the primary learning task. Moreover, eye movements reveal by which part of a presentation visual attention is captured. The amount of time spent inspecting a discrete area (words, sentences, pictures) is commonly taken as a correlate of the amount of cognitive resources allocated to the processing of the inspected area (e.g. Just & Carpenter, 1980). In order to compare viewing behavior with actual learning outcomes we need to consider cognitive processes in more detail.

Instructional design principles and theoretical explanations

Current research on multimedia learning and instructional design is dominated by two theoretical frameworks, cognitive load theory (Sweller, 1999) and cognitive theory of multimedia learning (Mayer, 2001). The key issue of these theoretical approaches is to base instructional design on "how the human mind works" (Mayer, 2001, *p*. 41). Active learning requires coordinating a set of cognitive processes mainly organizing different sources of information into a coherent structure or schema. The coordination and organization processes are assumed to take place in a limited-capacity working memory. In accordance with Baddeley's working memory model (Baddeley, 1986) both theories assume sensory-specific subsystems or processing channels for visual and auditory information. Visual information is processed in a visuo-spatial sketchpad, auditory information is processed in an auditory loop. Each subsystem has limited processing capacities. In order to promote learning, the crucial task for an instructional designer is, thus, to make an optimal use of these limited capacities.

Derived from empirical evidence the best-supported instructional design principles are the modality principle and the spatial contiguity principle. The modality principle states that "students learn better when words in a multimedia message are presented as spoken text rather than printed text" (Mayer, 2001, p. 134). The spatial contiguity principle states that "students learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen" (Mayer, 2001, p. 81). Using the cognitive framework, effects of text modality and spatial contiguity can be attributed to different *cognitive loads* caused by presentation format (Mayer & Moreno, 2003; Sweller, et al., 1998). If expository text is added to an illustration in written form, both materials have to be processed by the visual processing system. Under these conditions, an overload of the visual system is more likely to occur compared to spoken text presentation. The spatial contiguity principle rests on a potential overload caused by split visual attention (Sweller et al., 1998). If different sources of visual information are physically separated (e.g. illustrations and captions), learners are forced to split visual attention between these sources. As a consequence, before integrating both sources, the source that was attended first must be held active in visual working memory until the corresponding information in the second source is found and processed. The more information is held active or the more capacity is needed to search for corresponding information the more likely a cognitive overload occurs.

Are there alternative explanations for modality and spatial contiguity effects? A necessary prerequisite for cognitive load to occur is the intake of information. As Moreno and Mayer (1999) point out "the superiority of concurrent animation and narration over concurrent animation with on-screen text might be caused by students missing part of the visual information while they are reading the on-screen text (or vice versa)" (Moreno & Mayer, 1999, p. 359). Thus, if we take properties of the sensory modalities into account, the modality effect might be explained by a load on visual perception rather than working memory. Since we cannot view two spatially separated areas at the same time presenting spoken rather than written text increases the probability of both, text and concurrent visualization, to enter working memory. The more information can be retrieved by the sensory modalities the more is possibly learned. In a similar fashion a lack of spatial contiguity can put a load on visual perception. Since written text forces learners to split their visual attention between text and visualization it initiates a visual search for corresponding information. If the physical distance between information sources complicates this visual search, less time might be available for retrieving relevant information. In both cases, the perceptual demands of a particular presentation format can have an impact on the amount of information that is actually processed.

In the study by Faraday and Sutcliffe (1996) animation produced a shift of visual attention towards objects in motion. Motion is known to capture visual attention (Hillstrom & Yantis, 1994). In this case, however, visual attention was sometimes "locked" by written text. If we assume appropriate selection of areas of information being a crucial prerequisite for successful learning to occur, animation might help or hinder gaining a maximum of information from combinations of visualization and written text. From this perspective, animated compared to static visualization is expected to (a) shift the split of visual attention between written text and visualization towards visualization, (b) facilitate visual search, or (c) increase the competition between both information sources. If attentional demands of multimedia presentation play a role in modality and spatial contiguity effects, the effects should be moderated by attentional characteristics of the visualization.

Taken together, the following experiments are designed to examine the effects of modality and spatial contiguity of text presentation on visual attention allocation and its correspondence to learning outcomes (Experiment 1) and whether visual attention allocation and learning outcomes are influenced by characteristics of the visualization (Experiment 2). Experiment 1 examines how learners allocate or split their visual attention if text accompanied by animated visualizations is written rather than spoken. How much of the total learning time is spent reading written text (if present) and inspecting the visualization? Does physical distance influence the split of visual attention between written text and visualization? And, can this viewing behavior account for modality and spatial-contiguity effects?

Experiment 1: Correspondence between viewing behavior and comprehension in modality and spatial contiguity

The purpose of this Experiment is to examine the allocation of visual attention in a common setting of multimedia learning. The material was similar to one used by Moreno and Mayer (1999) in order to compare viewing behavior with learning outcomes in a well-established learning scenario. Moreno and Mayer (1999) used a multimedia explanation about the process of lightning formation. In the first experiment of their study, animated illustrations were presented concurrently with expository text that was either spoken or written. Physical distance of written text was further varied as being integrated, i.e. physically close to the illustrations, or separated, i.e. physically far from the illustrations. They found that participants in the spoken text group performed better on retention, transfer and matching tests than both written text groups, confirming the well-established modality effect. A spatial-contiguity effect was found in two of three measures for written text presentation. The integrated text group outperformed the separated text group in the retention and transfer but not in the matching test.

Using the same experimental manipulation as Moreno and Mayer (1999), Experiment 1 tests the following hypotheses: First, the results of Moreno and Mayer (1999) will be replicated, i.e. participants in the spoken text group are expected to outperform both written text groups in verbal recall and transfer, and participants in the integrated text group are expected to outperform the separated group in verbal recall and transfer. Second, we expect that participants will spend more time viewing the visualizations in the spoken text group than in both written text groups. Participants in the written text groups need to split their visual attention between text and visualizations and, thus, will spend a reasonable amount of time reading. And third, participants in integrated and separated text groups are expected to differ in their viewing behavior. Due to the higher spatial distance between text and visualizations and the higher demands on visual search the separated text group might spend less time reading and/or viewing visualizations than the integrated text group.

Method

Participants and Design

40 students of the Justus-Liebig University Giessen participated in the experiment in partial fulfillment of a course requirement. One additional participant had to be excluded due to experimenter error. All participants were native German speakers with normal vision. People with deficient vision were excluded beforehand in order to minimize possible problems with the recordings of eye movements. Participants were randomly assigned to one of three experimental groups. The experimental design was identical to Moreno and Mayer (1999, Experiment 1). 14 participants served in the spoken text group, 14 participants in the integrated written text group, and 13 participants in the separated written text group. The groups did not differ in prior knowledge. Mean values vary around 2.2 for self-estimated prior knowledge (on a 5-point scale from very little to very much) and between 2.5 and 3 checked items on a checklist consisting of 8 domain-related items.

Materials and apparatus

The learning material consisted of a 16-step multimedia instruction on the formation of lightning programmed in Flash 4.0 (Macromedia, 1999). The multimedia instruction used by Moreno and Mayer (1999) was redesigned in order to synchronize the presentation of instruction with the eye tracking equipment and to gain full access to the material for further experimental variations. Expository text was translated into German and visualizations were designed to be equivalent in content to the instruction by Moreno and Mayer (1999). The instruction showed a sequence of 16 animated illustrations depicting the motion of cool air that becomes heated; heated air rising up and forming a cloud; the rising of the cloud beyond the freezing level; drops of water and ice crystals moving up and down within the cloud, colliding, and causing electrical charges to arise; heavy drops and crystals falling down and producing downdrafts; a stepped leader of negative charges moving down to high objects on the ground; and positive charges moving up to the cloud producing a flash light. Illustrations were accompanied by an expository text describing each of the major events. Text was spoken, written inside the illustration frame or written below the illustration frame (Figure 1). The whole text had a length of 281 words, varying between 9 and 26 words per scene. Scene length was matched to the number of words per scene with a rate of 82 words per minute. For the spoken text condition, text was spoken in a male voice at a rate appropriate for this timing. Overall duration of the instruction was 206 seconds[1].

Figure 1. A selected frame and corresponding integrated written text for a multimedia explanation on lightning formation. This scene explains the effect of downdrafts: "When downdrafts strike the ground, they spread out in all directions, producing the gusts of cool wind people feel just before the start of the rain."

The instruction was presented by a PC on a 21'' color monitor, situated approximately 80 cm from the participant. Spoken text was presented by an audio system. Fixations were monitored by an Applied Science Laboratories' corneal-reflectance and pupil-center eye tracker (ASL 504). Fixation position on the screen was measured with a sampling rate of 50 Hz and output to a PC, which controlled the recording, the camera, and the calibration (ASL Eyepos, E5000). Two additional monitors displayed the participants' tracked eye and its current fixation position on the stimulus screen to the experimenter.

Prior knowledge and performance measures were assessed by paper-and-pencil tests. The material consisted of a participant questionnaire, a retention test, and a four-item transfer test. The participant questionnaire asked for the participant's gender, age, profession and experience with meteorology. The retention test asked the participants to write down an explanation of how lightning works until they were

[1] Note, that in the original presentation Moreno and Mayer (1999) adjusted the speed of presentation to the rate of the speaker. This resulted in a 140 s presentation with approximately 123 words per minute. This pace appeared rather fast to me. Thus, I decided to slow down the whole presentation in order to gain an instructionally reasonable density of information for both presentation formats.

told to stop. The transfer test contained four questions: (1) “Which physical conditions must be given in order to decrease the intensity of a lightning storm?”, (2) “Why do you often see clouds in the sky but no lightning?”, (3) “What does air temperature have to do with lightning?”, and (4) “What are the physical causes of lightning?”. Answers were asked to be given in a 5-alternative forced choice format. Alternatives were selected from a pool of open answers objected in a prior experiment.

Procedure

Participants were tested in single sessions and were randomly assigned to one of the three treatment groups. They were given general instructions explaining the procedure and introducing the topic. Participants were instructed to acquire as much information as possible about the formation of lightning from the following multimedia presentation in order to perform subsequent tasks. Next, participants were asked to fill out the questionnaire. Then, participants were seated in front of the stimulus PC and the eye-tracking system was calibrated. After that, the experimenter started the multimedia presentation. After participants had seen the presentation, they were given instructions to work on the retention test. Participants had 8 min to process the test. The retention test was followed by the transfer test. Instructions for the transfer test were handed out together with the first of four questions and the questions were handed out successively. Participants were given 4 min to answer each of the questions. After completion, participants were debriefed and thanked for their participation. The experimental session took about 40 min.

Results

Viewing behavior

For the n=40 participants calibration failed in 11 cases. The remaining 29 cases were processed in the following manner. Viewing positions were transformed into single fixations and saccades by using ASL-Eynal software. Areas of interest (AOI) were defined to cumulate single fixations and saccades into viewing time on text and illustration respectively. An AOI in the presentation was a part in which either a portion of text or an illustration was displayed. Figure 2 shows the area of written text and the area of illustration for an exemplary scene of the presentation. In order to detect inaccurate calibration the resulting viewing times were further inspected in the following manner. Data sets in which viewing time on AOIs summed up to less than 75% of the total presentation time were taken as possibly invalid. Applying this criterion, three further participants had to be excluded. Thus, the following analyses were calculated with a set of 26 data cases.

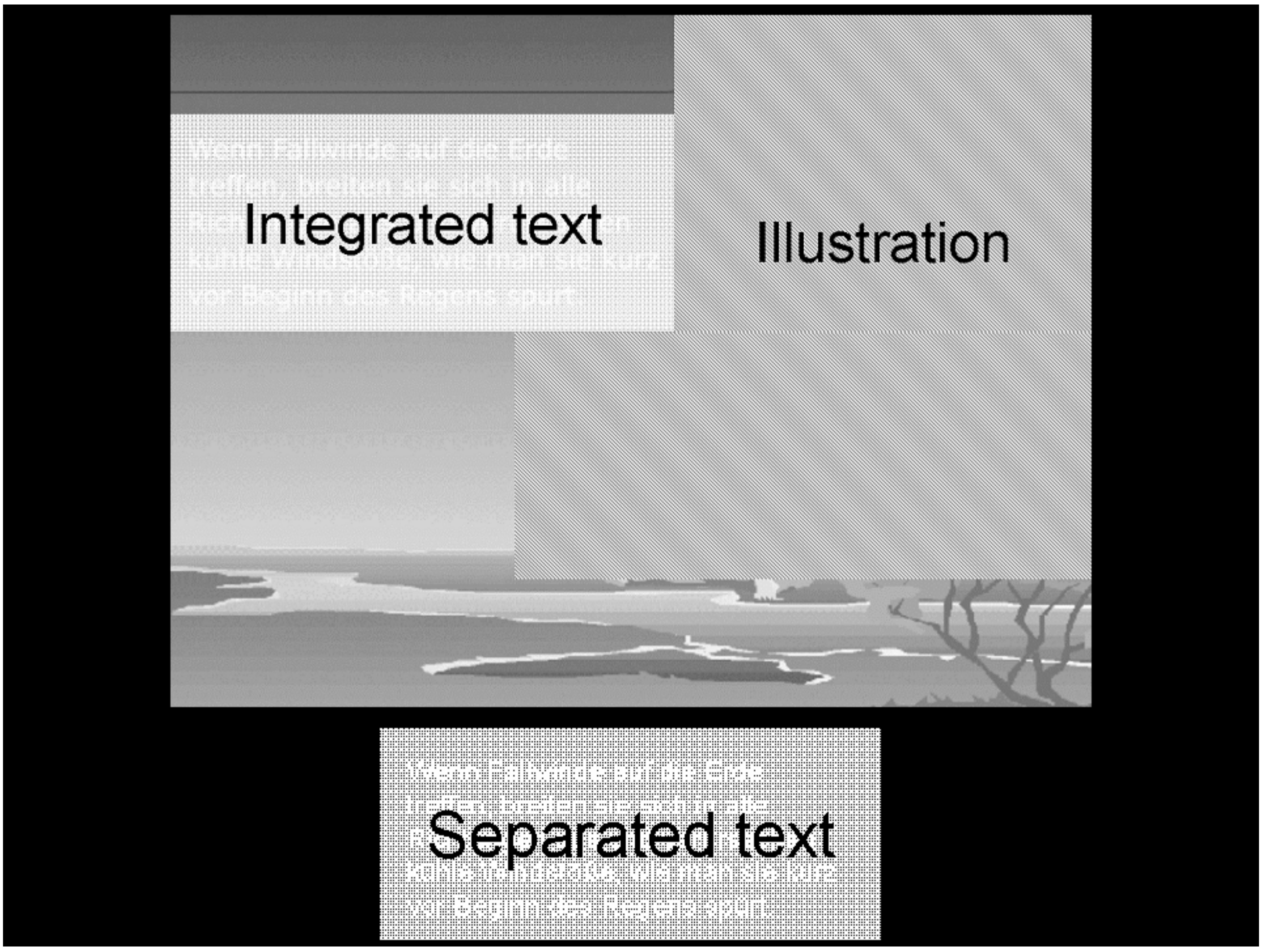

Figure 2. An example of areas of interest (AOI) for illustration and integrated and separated written text. Note that the areas vary in size and location from scene to scene depending on the text length and the location of the illustrations.

Overall, participants spent 190 s (SD=6.83) or 92% of their viewing time on AOIs. Means and standard deviations of viewing time on illustration and text as well as summed viewing time on AOIs for each group are shown in Table 1.

Table 1
Mean viewing time on areas of interest (AOI) for illustration, written text (if present) and sum of text and illustration areas (sum) for each of the three experimental groups.

		Groups					
		Spoken text		Integrated text		Separated text	
		M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
AOI	Illustration	187.4	4.7	92.1	17.5	74.8	13.8
	Text	-	-	101.9	21.5	115.1	16.8
	Sum	187.4	4.7	194.0	6.3	189.9	8.1

An analysis of variance (ANOVA) conducted on viewing time on illustration showed that the groups differed significantly, $F(2,23)=195.31$, $MSE=32{,}738$, $p<.001$, $\eta^2=.94$. Tukey tests (based on an alpha of .05) revealed that all three groups differed from each other in the amount of time spent inspecting illustrations. The spoken text group spent significantly more time inspecting illustrations than the written text groups. This result confirms that modality of text presentation affected viewing behavior on illustration. Effect size indicates this as a massive effect. Within written text groups the integrated text group spent more time inspecting illustrations than the separated text group. Thus, the spatial contiguity of written text presentation affected the time spent inspecting illustrations. Viewing times on AOIs are shown in Figure 3.

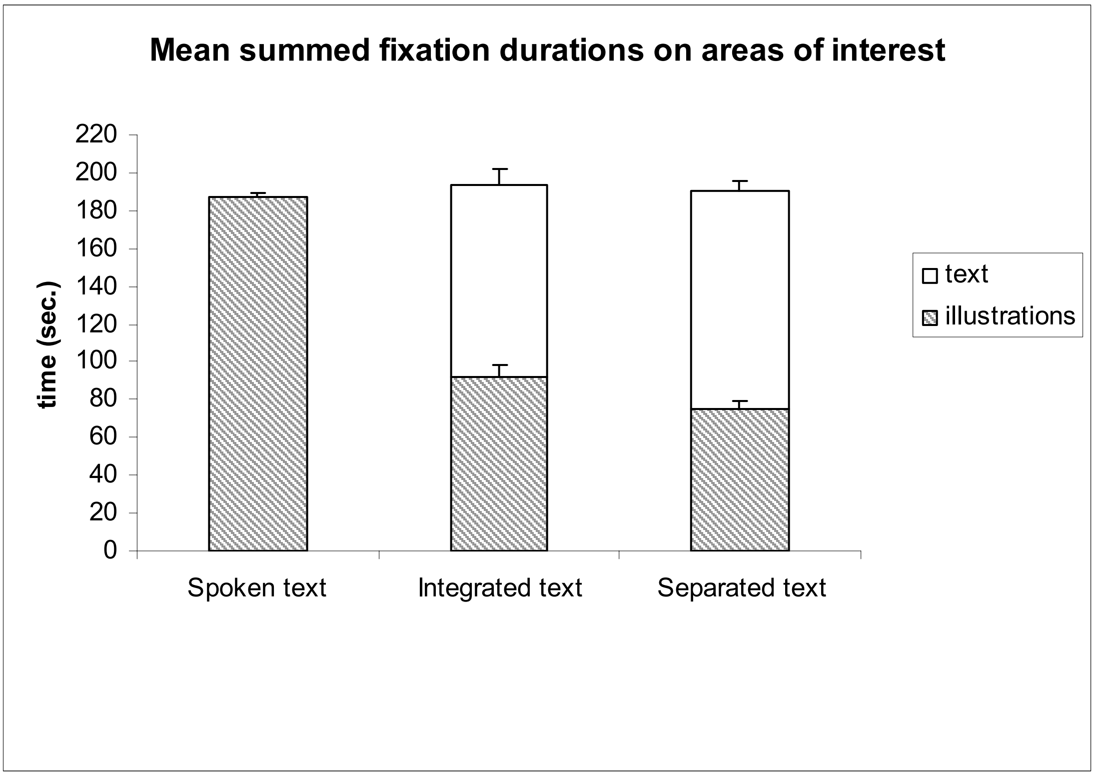

Figure 3. Mean viewing time on areas of interest (AOI) separated into viewing time on illustration and viewing time on written text (if present) for each of the three experimental groups.

Within written text conditions participants split their visual attention in the following manner. Overall, participants alternated between text and illustration 3.3 times per scene. There was no significant difference in this behavior between integrated and separated text groups ($t(15)=0.68$, $p>.10$). Exploring the scan paths revealed that after a scene change 93% of the first five fixations were on written text. Integrated and separated text groups did not differ ($t(15)=0.95$, $p>.10$). Participants apparently objected text as soon as new text occurs and initiated a reading sequence. Illustrations were mainly ignored at the beginning of a new scene. Overall, the mean ratio of the time spent reading to the time spent inspecting illustrations was 1.4. Participants spent about 40% more time reading than inspecting illustrations. This ratio did not significantly change with illustration format ($t(15)=1.71$, $p>.10$). Although participants differed in their total time spent inspecting illustrations (see above) the time spent reading text was not significantly influenced ($t(15)=1.42$, $p>.10$). Taken together, participants in written text conditions inspected the multimedia explanation in a comparable fashion. However, if text was integrated rather than separated, illustrations gained some more visual attention.

Performance Measures

Each participant's performance on the retention test was scored with two scorers being unaware of the participant's identity. Participants were given 1 point for each of nineteen main ideas of the casual chain of lightning formation. The inter-rater reliability was $r=.96$. Scores for the problem-solving transfer were obtained by counting the number of correct marks in the forced-choice items, i.e. a maximum of 4 points could be obtained in the transfer test. No participants had to be excluded from further analyses of the performance measures, thus the following calculations are based on n=40 data sets. Mean scores and standard deviations for both measures are shown in Table 2.

Table 2
Mean scores and standard deviations by the three experimental groups on performance measures.

	Groups					
	Spoken text		Integrated text		Separated text	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Retention	15.6	1.8	11.9	3.4	10.2	3.3
Transfer	3.1	0.9	2.0	1.4	1.2	0.9

An ANOVA conducted on retention scores with groups as between-subjects factor indicated a significant difference, $F(2,37)=12.06$, $MSE=101.22$, $p<.001$, $\eta^2=.40$. Tukey tests (based on an alpha of .05) revealed that the spoken text group recalled more idea units than did both written text groups. An ANOVA conducted on transfer scores with groups as between-subjects factor indicated a significant difference, $F(2,37)=9.46$, $MSE=11.57$, $p<.001$, $\eta^2=.34$. Tukey tests (based on an alpha of .05) revealed that the spoken text group selected more correct alternatives than did both written text groups. The differences between integrated and separated written text in both scores shown in Figure 4 failed to reach statistical significance. Thus, both performance measures replicated the modality but not the spatial contiguity effect.

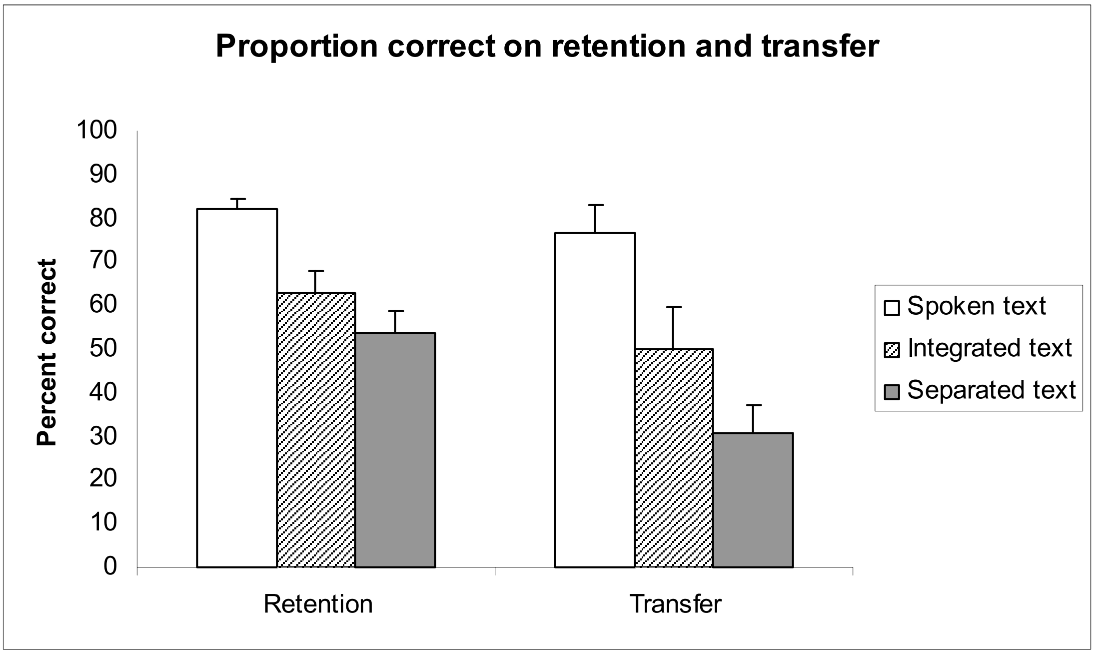

Figure 4. Proportion correct on retention and transfer in Experiment 1 (with standard error bars).

Discussion

Applying a learning material similar to one used by Moreno and Mayer (1999) the experiment delivered converging evidence for the modality effect. The spoken text group performed better on retention and transfer tests than both written text groups. This superiority of spoken text can be explained in terms of visual attention allocation. Participants in the written text conditions spent much less time exploring the illustrations than the spoken text group and alternated between reading text and inspecting illustrations several times per scene. The observed viewing behavior indicated that presenting written text distracted participants' visual attention from illustrations. Written text was read first before illustrations were inspected and participants spent more than 50% of their time reading. These results replicate earlier findings on split visual attention between textual and pictorial information (Faraday & Sutcliffe, 1996; Hegarty & Just, 1993; Rayner et al., 2001). Written text proved to be a highly salient stimulus for visual attention allocation. In the context of the instructional material used in the experiment this viewing behavior appears reasonable. The textual information helps interpreting the illustrations. However, due to the amount of visual attention on written text and the alternation between text and illustration some of the visual information was possibly missed or less thoroughly processed compared to spoken text groups.

Integrating written text into the illustrations was expected to lower the burden of attentional split and thus lead to better learning performance compared to separated text presentation. Spatial distance between written text and illustrations had no significant influence on fixation paths and time spent reading text. But the illustrations apparently benefited from an integrated text format. More time was spent inspecting illustrations if text was integrated rather than separated from illustrations. This effect of spatial contiguity failed to significantly affect subsequent learning performance. Descriptively, however, the differences in retention and transfer tests pointed in the direction one would expect if the time spent inspecting illustrations is a predictor for subsequent learning performance, as suggested by the attentional interpretation of the modality effect. However, the variation of physical distance was probably not strong enough to provoke clearer differences in viewing behavior and learning performance.

Taken together, Experiment 1 demonstrated that effects of instructional design on learning outcomes correspond to attentional demands of the learning material indicated by a particular viewing behavior. Presenting written rather than spoken text caused a split of visual attention between text and illustration, and decreased learning performance. No clear effect could be confirmed for spatial contiguity. However, the split of visual attention was affected by the salience and spatial accessibility of both information sources. Thus, varying the relative salience of competing visual information sources should change viewing behavior. If managing the attentional split between written text and illustration is causally related to subsequent learning outcomes, the modality effect might be moderated by the relative salience and accessibility of pictorial information. In order to test this hypothesis Experiment 2 was conducted.

Experiment 2: The influence of animated and static illustration on viewing behavior and the modality effect

Experiment 1 revealed that the modality effect can – in terms of visual attention allocation – be interpreted by the processing of illustrations. Especially illustrations might suffer from less attention given to them whenever visual attention has to be split between written text and illustrations. Alternatively, it is also possible that the processing of both information sources is affected: (1) the mere presence of animated illustrations might impede reading comprehension, (2) the time needed to alternate between written text and illustrations and to visually search for corresponding parts of information might cause information loss in *both* sources, or (3) the demand to manage attention allocation and visual search might put an additional load on the learner's cognitive system. (Note that these interactions between written text and illustrations are not mutually exclusive.)

The purpose of Experiment 2 is, thus, to examine whether characteristics of illustration moderate the effects of text modality on viewing behavior and learning. By comparing animated with static illustrations Experiment 2 asks whether and to what degree attentional salience of illustration influences visual attention allocation and subsequent learning performance. How much attention is devoted to animated compared to static illustrations? How is visual attention on written text affected by the presence or absence of visual motion in an illustration? And, does the presence or absence of visual motion moderate the modality effect in learning outcomes? To answer these questions, eye movements and learning performance of students were observed who received one of the following four presentation formats: a multimedia instruction presenting (1) a sequence of animated illustrations together with written text, (2) a sequence of animated illustrations with spoken text, (3) a sequence of static illustrations with written text, and (4) a sequence of static illustrations with spoken text.

Concerning the modality of text presentation Experiment 2 is expected to replicate the effects found in Experiment 1. More time will be spent viewing illustrations when text is spoken rather than written. This modality effect should also appear in learning performance.

The manipulation of attentional salience of illustration should lead to more visual attention on animated compared to static illustrations. This effect is mainly expected within written text conditions where illustrations compete with text. Visual attention has to be split between both information sources. The attentional split can be expected to change viewing behavior in favor of animated over static illustration. More time will be spent inspecting animated than static illustration. Consequently, less time can be spent reading if illustrations are animated rather than static.

What are possible consequences for learning outcomes? If the comparably higher salience of animation drags visual attention away from written text, it might disturb reading comprehension. In this case, the modality effect should be increased by animated compared to static illustrations. If, however, animation facilitates visual search for illustrations that correspond to some portion of written text, animated illustration should decrease the modality effect.

To measure the selective influence of attentional salience on text comprehension and processing of illustrated information I introduce a visual memory test. I expect participants to perform better on a visual memory test the more time they spend inspecting illustrations, and to perform better on a verbal retention test the more time they spend reading.

Method

Participants and Design

50 students of the Justus-Liebig University Giessen voluntarily participated in the experiment. All participants were native German speakers and had normal or corrected to normal vision. Participants

were randomly assigned to one of four experimental groups. 13 participants served in the groups receiving animated illustrations with spoken text and static illustrations with spoken text respectively. 12 participants served in each of the other two experimental groups. The groups did not differ in prior knowledge. Mean values vary around 2.5 for self-estimated prior knowledge (on a 5-point scale from very little to very much) and between 2.5 and 3.5 checked items on a checklist consisting of 8 domain-related items.

Materials and apparatus

The learning material was designed on the basis of Experiment 1. Material for animated illustration with spoken text was identical to the spoken text condition of Experiment 1. Material for animated illustration with written text was identical to the integrated text condition of Experiment 1. The integrated text format was chosen in order to avoid confounding effects of text modality and spatial contiguity. Static illustrations were prepared to be informationally equivalent to animated illustrations. Simple movements like “cool air moving over a warmer surface” were indicated by arrows. In the case that the final state of an animated illustration did not match the initial state (e.g. positive charges in the cloud moving to the top and negative charges in the cloud moving to the bottom of the cloud), static illustrations visualized the process leading to the final state (e.g. arrows indicating the direction of movement). Scene length was matched to the number of words per scene with a rate of 66 words per minute. For spoken text conditions, text was spoken in a male voice at a rate appropriate for this timing. Overall duration of the instruction was 256 seconds[2].

Stimulus presentation and eye tracking equipment as well as prior knowledge and performance measures were the same as in Experiment 1. In order to investigate the influence of presentation format on the processing of illustrated information, a visual memory test was applied. The test contained instructions to sketch (1) a cloud with a sufficient condition for electrical charges to arise, (2) how electric charges arise in a thundercloud, (3) the distribution of charges within a thundercloud before the stepped leader builds up, and (4) a stepped leader as it arises before a lightning. Answers were supposed to be given on four sheets containing a simplified background scene of the presentation.

[2] Note, that the timing of the presentation was even slower than in Experiment 1. Participants in Experiment 1 still reported difficulties in keeping up with the speed of the presentation.

Procedure

The procedure was identical to Experiment 1 except that after the transfer test the visual memory test was given. Participants were given 5 min to work on the sketches. Tasks were handed out successively. After completing the visual memory test, participants were given additional 3 min to write comments on their sketches in a different color. This was done in order to facilitate scoring of ambiguous sketches. The experimental session lasted about 45 min.

Results

Viewing behavior

For the n=50 participants calibration failed in 4 cases. The remaining 46 cases were processed according to Experiment 1. 16 participants whose viewing times on AOIs summed up to less than 75% of total presentation time were excluded from further analyses. Thus, the following analyses were conducted with a set of 30 data cases. Means and standard deviations are shown in Table 3.

Table 3
Mean viewing time on areas of interest (AOI) for illustration, written text (if present) and sum of text and illustration areas (Sum) for each of the four experimental groups.

			Text presentation format			
			Spoken text		Written text	
		AOI	*M*	*(SD)*	*M*	*(SD)*
Illustration format	Animated	Illustration	218.8	(13.6)	85.3	(15.3)
		Text	-	-	137.8	(22.0)
		Sum	218.8	(13.6)	223.1	(10.5)
	Static	Illustration	215.1	(17.3)	70.1	(23.9)
		Text	-	-	152.2	(25.4)
		Sum	215.1	(17.3)	222.3	(13.2)

ANOVA with text presentation (spoken vs. written) and illustration (animated vs. static) as between-subjects factors and with viewing time on illustration as dependent measure revealed a main effect for text presentation format, $F(1,26)=416.94$, $MSE=139{,}556$, $p<.001$, $\eta^2=.94$. As shown in Figure 5, participants in the spoken text groups spent more time viewing illustrations than participants in the written text groups. This result replicates the effect of text modality on viewing time allocated to illustration that was found in Experiment 1. No main effect for illustration ($F(1,26)=1.91$, $p>.10$) and no interaction ($F<1$) occurred.

Within written text groups, participants split their visual attention between text and illustration in a comparable fashion. Overall, participants in written text conditions alternated between reading text and inspecting illustrations 5.1 times per scene. There was no significant difference in this behavior between animated and static illustration groups ($t(13)=0.62$, $p>.10$). Within the first five fixations after a scene change 90% of fixations were on text. Animated and static illustration groups did not differ ($t(13)=1.42$, $p>.10$). Participants apparently started reading text as soon as new text occurred after a scene change. The mean ratio of the time spent reading to the time spent inspecting illustrations was 2.3. Participants spent twice as much time reading than inspecting illustrations. This ratio did not significantly change with illustration format ($t(13)=1.20$, $p>.10$). Participants did neither differ in their total time spent reading ($t(13)=1.23$, $p>.10$) nor in time spent inspecting illustrations ($t(13)=1.37$, $p>.10$). The hypothesis that animated illustrations drag more visual attention away from written text than static illustrations was not statistically confirmed. However, the 15 sec shift of viewing time in favor of animated over static illustration points in the expected direction. The right panel of Figure 5 shows the viewing times on written text and illustrations.

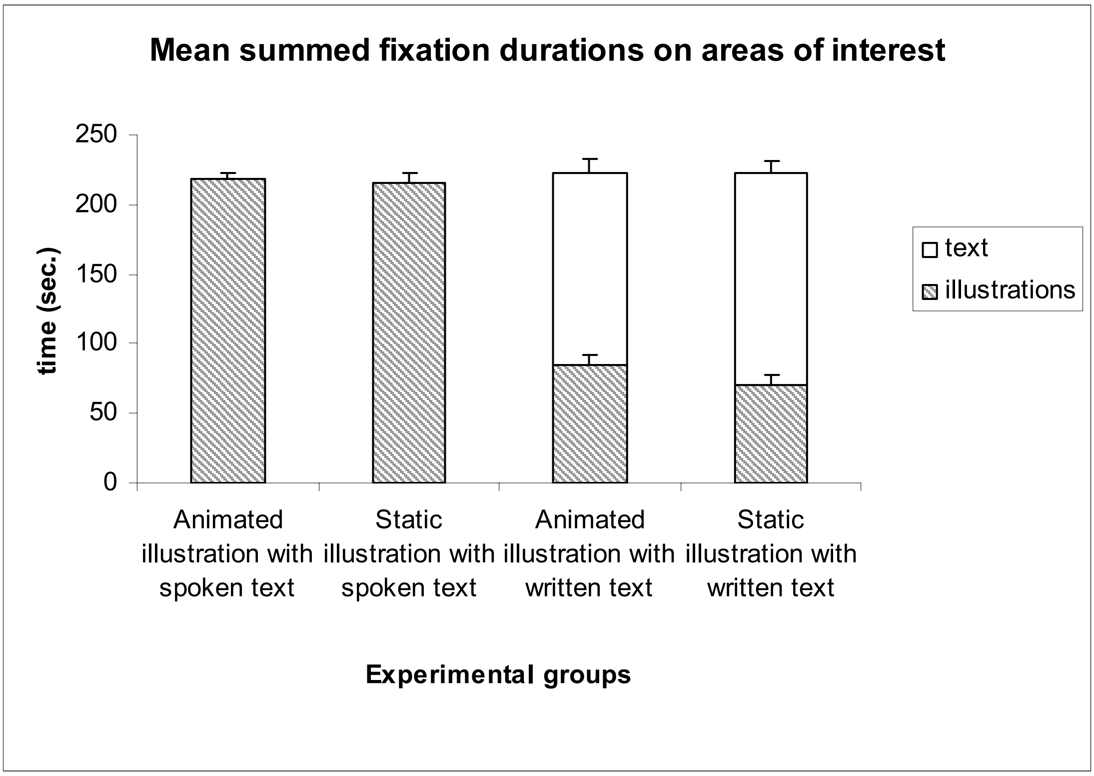

Figure 5. Mean viewing time on areas of interest (AOI) separated into viewing time on illustration and viewing time on written text (if present) for each of the four experimental groups.

Performance Measures

Retention and transfer scores were obtained in the same way as in Experiment 1. Visual memory was scored by two independent scorers being unaware of the participants' identity. Participants were given 1 point for each appropriate and identifiably sketched visual element, with a maximum of 2 points obtainable for sketches 1, 3, and 4, and a maximum of 3 points for sketch 2. Examples for acceptable answers are: a straight line with temperature symbols indicating that the cloud extends above the freezing level (sketch 1), the collision of water and ice crystals in the cloud (sketch 2), negative charges at the bottom of the cloud (sketch 3), and a stepped leader between the cloud and a higher object from the ground (sketch 4). Inter-rater-reliability was r=.77 for the retention and r=.87 for the visual memory test.

No participants had to be excluded from analyses of performance measures, thus the following analyses were conducted with n=50 data sets. Table 4 shows mean values and standard deviations of performance scores for retention, transfer and visual memory tests.

Table 4
Mean values and standard deviations of performance scores for visual memory, retention, and transfer tests.

		Text presentation format			
		Spoken text		Written text	
	Illustration format	*M*	*(SD)*	*M*	*(SD)*
Visual memory	Animated	6.8	(2.0)	6.2	(1.1)
	Static	7.0	(1.6)	5.3	(2.0)
Retention	Animated	13.4	(2.5)	11.4	(2.6)
	Static	12.7	(2.7)	13.2	(2.7)
Transfer	Animated	3.5	(0.7)	3.3	(0.7)
	Static	3.5	(0.7)	3.3	(1.2)

Effects on visual memory. ANOVA with the between-subjects factors text presentation (spoken vs. written) and illustration (animated vs. static) revealed a main effect for text presentation format (F(1,46)=5.36; MSE=15.91, p=.025, η^2=.10). As shown in Table 4, participants in both spoken text conditions performed better on visual memory than participants in the written text conditions. Illustration format (F<1) and interaction (F(1,46)=1.35, p>.10) failed to reach statistical significance. This result replicates the well-established modality effect with a visual memory task.

Effects on verbal recall. ANOVA with the between-subjects factors text presentation (spoken vs. written) and illustration (animated vs. static) revealed no significant main effects for text presentation format (F(1,46)=1.01, MSE=6.90, p>.10) and visualization format (F<1). The interaction, however, was marginally significant (F(1,46)=2.83, MSE=19.4, .10>p≥.05, η^2=.06). As shown in Figure 6, the animated illustration with written text group performed worse than the other three groups. One-tailed post-hoc t-test for both animated illustration groups showed a significant difference between these two groups (t(23)=1.965, p<.05) in the direction predicted by the modality principle. These results confirm the modality effect for animated but not for static illustration.

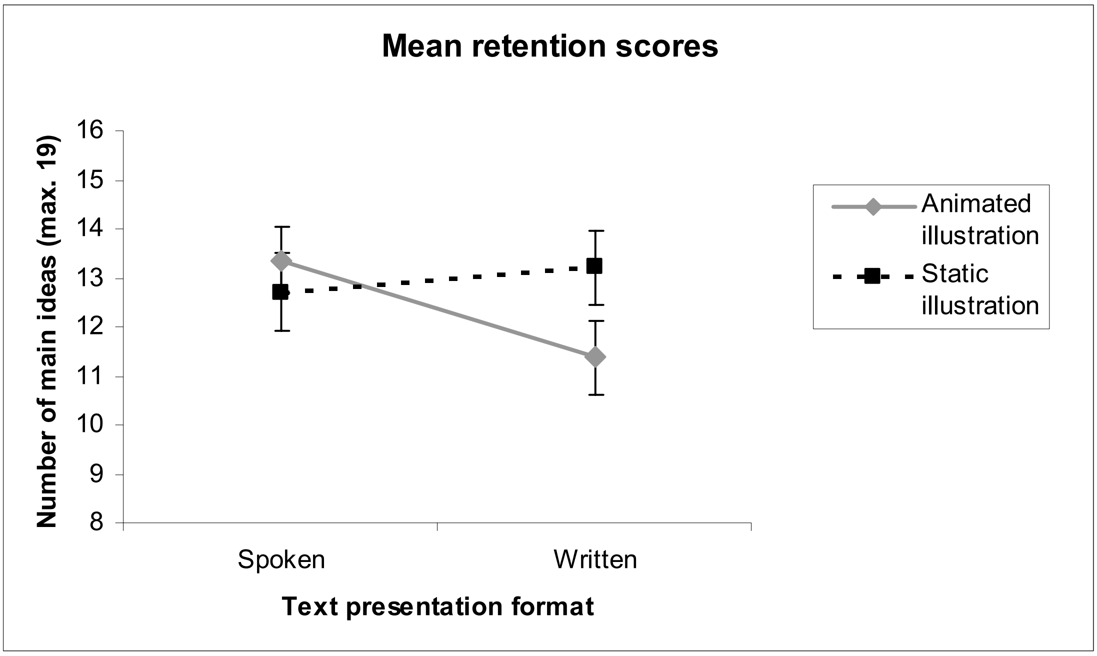

Figure 6. Mean retention scores (with standard error bars) for each of the four experimental groups.

Effects on problem-solving transfer. ANOVA with the between-subjects factors text presentation (spoken vs. written) and illustration (animated vs. static) revealed no significant main effects for text presentation format and visualization format, and no interaction (all Fs<1).

Discussion

The observed pattern of viewing behavior replicated the results of Experiment 1. Illustrations got less visual attention if text was written rather than spoken. This modality effect on viewing times was as massive as in Experiment 1. Within written text conditions, text was attended first and captured relatively more visual attention than accompanying illustration. Thus, written text proved to be a highly salient stimulus for visual attention allocation that is at least initially preferred to illustration.

The effect of text modality on visual attention allocation is mirrored by the performance in the visual memory test. Participants performed worse in this test if text was written rather than spoken. The poorer performance of written text groups in this test can be attributed to viewing time on illustration. Written text groups had less time processing illustrations since they spent much of their time reading text. Consequently, they had more difficulties remembering the illustrations and sketching main steps of the process of lightning formation.

The retention test replicated the modality effect of text presentation for animated illustration as found in Experiment 1. However, the negative effect of written text presentation disappeared in static illustration conditions. Thus, for text based recall the modality effect was moderated by the format of illustration. Apparently, animated illustrations disturbed reading comprehension. Similar to Experiment 1, however, the variation of illustration format failed to significantly change viewing behavior. Animation was expected to shift the attentional split between text and illustration. Participants in the written text groups did not differ in their general viewing behavior. Both groups attended to written text first and alternated between written text and illustration equally often. Descriptively, however, participants spent some more time on animated compared to static illustration and thus had less time reading text.

Taken together, Experiment 2 confirmed the interpretation of the modality effect given in Experiment 1. The modality effect for visual memory can be explained in terms of loss of visual information whenever visual attention has to be split between illustrations and written text. The relative salience of illustration compared to written text might have an additional influence especially on reading comprehension.

General discussion

The goal of the experimental studies presented in this paper was to examine attentional processes in multimedia learning. Applying the measure of eye movements it was found that viewing behavior is influenced by characteristics of instructional design. Most obvious, whenever illustrations were accompanied by written text learners alternated between reading some portion of text, inspecting illustrations, going back to the text and so on. In general, written text was attended first and gained relatively more attention than illustrations. Consequently, learners spent much less time inspecting illustrations if explanatory text was written rather than spoken. The relative distance between written text and illustrations only had a mild effect on this viewing behavior, i.e. integrated text facilitated the allocation of visual attention on illustrations without significantly affecting reading behavior. Varying attentional salience by presenting illustrations animated or static failed to significantly influence the described split of visual attention between text and illustration.

The observed viewing behavior complements current research in the field. Other eye tracking studies on simultaneous presentation of text and illustration consistently report that viewers tend to read several portions of text before inspecting corresponding parts of pictorial information (Carroll et al., 1992; d'Ydewalle & Gielen, 1992; Hegarty & Just, 1993; Rayner et al., 2001). This preference for textual information can be interpreted in different ways. As d'Ydewalle points out reading is possibly more efficient than watching pictures, as in subtitled television (d'Ydewalle & Gielen, 1992; d'Ydewalle et al., 1991). Considering more complex pictorial information, reading text can also guide the processing of illustration. For example, while viewing a diagram of a pulley system with an additional text describing the diagram, students successively read one or two sentences of text and then inspected the described portion of the diagram (Hegarty & Just, 1993). In accordance with these observations our results proved written text to be a highly salient stimulus for visual attention allocation in our material. Learners attended to written text first and devoted a high amount of visual attention to written text.

In the current setting, this apparently stable viewing behavior opens a different view on the modality effect. Since presentation time was limited learners were forced to weigh a trade-off between attention allocation on written text and pictorial information. The high salience of written text dragged visual attention away from pictorial information. Under these conditions it is very likely that the illustrations were not as thoroughly processed as with spoken text. Thus, the modality effect in learning outcomes found in both experiments can be explained by visual attention allocation. Learners in written text conditions did not sufficiently elaborate the illustrations to perform equally good on subsequent performance tasks. This interpretation becomes especially evident in the selective influence of text modality observed in Experiment 2. Performance measures revealed a modality effect in a task based on illustrated information while no main effect occurred when the task was based on verbal recall. The overall inspection time of the illustrations dropped considerably from spoken to written text presentation. Thus, the time spent inspecting illustrations served as a predictor for subsequent performance in a visual memory task. In order to comprehend written text, however, the time spent reading apparently sufficed to solve a verbal recall task as good as if text was spoken.

In *cognitive* theories of multimedia learning (Mayer, 2001; Sweller et al., 1998) this interpretation of the modality effect is not made explicit. These theories ascribe the modality effect to limitations of (visual) working memory. However, an obvious limitation in the material was the amount of time that could be spent viewing relevant parts of the instruction. Several theories on visual attention allocation (e.g. Allport, 1989; Van der Heijden, 1996) suggest that the eye itself is a limiting factor for information processing. In fact, the resources of working memory may or may not be sufficient to process all information taken in by the eye. But the eye itself is surely limited in the amount of information that can be fixated and retrieved in a discrete time interval. Thus, if a learner has enough time to read written text *and* inspect illustrations, the superiority of spoken over written text presentation possibly disappears. Further research is necessary to examine if the modality effect can be moderated in this manner by varying presentation time.

In terms of visual attention allocation spatial contiguity of written text and illustration is expected to have an effect on learning performance if it affects the visual access to corresponding verbal and pictorial information. Participants spent some more time inspecting illustrations if text was presented near rather than far from an illustration. The observed differences in retention and transfer tests failed to reach statistical significance. Thus, facilitated access to illustrated information did not lead to better learning performance. It cannot be ruled out that this lack of effect was a matter of effect size. However, the general pattern of viewing behavior was not influenced by the physical distance between text and illustration. No matter if text was presented near or far from the illustration, participants attended to text first, alternated between text and illustrations equally often, and did not significantly differ in the amount of time spent reading.

Why did the manipulation of spatial distance in Experiment 1 fail to provoke clearer differences in viewing behavior? The maximal distance between text and accompanying illustration for separated text in our material was 15 cm or 12° of visual angle. This distance can easily be covered by one or two saccades. The 16 scenes presented discrete steps in the formation of a lightning storm. Each scene contained an illustration that was commented by only one or two sentences. Within written text conditions, the captions always appeared at the same location. The illustration did not show more than the aspect described in the text. Furthermore, reading text allowed predicting in which part of the visual scene the next piece of information was likely to appear. Thus, the mere physical distance in our material does possibly not influence the visual search for appropriate referents.

In contrast, most studies referenced in support of a spatial contiguity principle confound the physical distance between text and illustration with a manipulation of referential cohesion (e.g. Chandler & Sweller, 1991; Mayer, 1989; Mayer et al., 1995; Sweller & Chandler, 1994; Sweller et al., 1990; Tindall-Ford et al., 1997). In these studies larger portions of text were usually integrated into illustrations by separating the text into smaller parts and placing these discrete parts of a label or caption close to spatially discrete referential parts of a picture, diagram, or table. Thus far we cannot estimate the relative contribution of physical distance and referential cohesion on the positive effects of text integration. But recent evidence shows that guiding visual attention to appropriate referents without manipulating physical distance can have the same positive effect as text integration (Kalyuga, et al., 1999; Reitmayr, 2003; Tabbers, et al., 2004). These findings support that visual search for appropriate referents is mainly a matter of attentional guidance. In this view spatial contiguity appears to have less impact than “referential contiguity”. More research is necessary to estimate the selective influence of spatial contiguity and referential cohesion between verbal and pictorial messages. Observing viewing behavior in those variations will gain further insight in the effects of spatial and referential properties on visual attention allocation and subsequent learning outcomes.

A step to vary attentional guidance was taken in Experiment 2. The manipulation of still vs. animated illustration was expected to have an impact on the split of visual attention between written text and illustration. Viewing behavior did not significantly vary with the presence or absence of motion within illustrations. Animation did not change the amount of visual attention allocated to written text and illustrations or the number of alternations between both visual information sources. As pointed out, the referential connections between text and illustrations in our learning material were rather obvious. Thus, similar to spatial contiguity of written text the potential of visual motion to guide visual attention did not influence the visual search for corresponding information. Participants' visual attention was apparently more guided by referential properties of the content of the learning material than by surface properties of the presentation format. Further research with other learning material is necessary to clear under which conditions animation influences visual attention allocation and subsequent learning in multimedia instructions.

The learning performances in Experiment 2 give a hint why there is so little empirical support for facilitative effects of animation over static illustration on learning (Chandler, 2004; Hegarty, 2004; Tversky et al., 2002). Animation is supposed to be a more natural way of conveying concepts of change such as in weather patterns, the cardiovascular system, or the mechanics of a bicycle pump. As a consequence, animation should help building up a good image-based representation and, thus, a more elaborated mental model of the process (Park & Hopkins, 1993; Rieber, 1990b). However, Experiment 2 did not reveal any positive effect of animation. The information given by animated illustration could equally effective be presented as static illustration. Furthermore, Experiment 2 revealed that animation can influence the processing of information even if the observable viewing behavior is not affected. When animation was accompanied by written text it rather hindered than helped learning. Although animated illustrations did not drag more visual attention away from written text than static illustrations, the presence of visual motion can moderate the modality effect. Suppressing the attentional capture of motion (Faraday & Sutcliffe, 1996; Hillstrom & Yantis, 1994) in the periphery of the visual field while reading expository text might have disturbed reading comprehension. Thus, the presumed facilitative effects of animation were confounded with other attentional demands caused by the presence of visual motion. As a practical consequence, instructional designers should weigh the potential advantages of animation against the cost for other concurrently presented information.

Taken together, the measure of eye movements was successfully applied to investigate effects of instructional design. Viewing behavior was added to a set of common observations in order to complement subsequent product measures of instruction with a measure of attentional processes. Doing so, we gained insight into how learners attend to different sources of information. Attentional demands of a particular learning material can help explaining learning differences caused by varying presentation formats. As a practical consequence the observed viewing behavior supports the general warning not to accompany animation with written text. However, the attentional demands of concurrent presentation of written text and illustration vary with the visual presentation format of both information sources. Additional research is necessary to further examine how animation, presentation duration, spatial distance and referential cohesion affect visual attention allocation and subsequent learning performance in multimedia instructions.

Chapter 3

Visual and cognitive load in multimedia learning: Effects of text modality, split-attention and pacing of instruction

This chapter examines different sources of visual and cognitive load in multimedia learning. In two experiments students watched an instruction on the formation of lightning while their eye movements during learning were recorded. Cognitive load was measured with self-ratings and subsequent task performances. Experiment 1 varied text presentation (spoken, written) and type of illustration (animated, static). Participants reported more difficulties with written compared to spoken text and rated written text presentation as less appropriate in its pace. Participants spent less time inspecting static than animated illustrations and less time on illustrations at all if text was written. No differences occurred in subsequent performance measures. Experiment 2 varied text presentation (spoken, written) and pacing of instruction (fast, medium, slow). Written compared to spoken text increased self-ratings of cognitive load and led to poorer performance in a visual memory task. Learners in written text conditions felt significantly more distracted by textual information and perceived the pace of presentation as less appropriate. Self-ratings and learning outcomes also indicated a higher cognitive load for fast presentation pace. An interaction of text modality with pacing was shown by viewing behavior and subjective load. The slower the pace of instruction, the more attention was given to illustrations in relation to reading time, and written text caused a higher cognitive load especially for fast presentation. The results underscore the important role of written text as a highly salient and potentially disturbing source of information as long as there is not enough time to inspect other information sources as well.

Introduction

In multimedia instructions learners must pay attention to different information sources, find and select corresponding parts of information and mentally integrate these parts into a coherent structure or schema. Thus, learning with multimedia instructions is not always a “joyful and effective experience” (cf. Tabbers, 2002, p. 31) but can put a high workload on the learners’ cognitive system. In Chapter 2 I presented converging evidence for one of the most prominent guidelines to keep this workload within bounds – the so-called modality principle: “Students learn better when words in a multimedia message are presented as spoken text rather than printed text” (Mayer, 2001, p. 134). Besides measuring learning outcomes in subsequent performance tasks the method of eye tracking was applied in order to investigate effects of instructional design on visual attention allocation. The observed viewing behavior revealed which information was attended, how long it was attended and presumably processed, and which information was not attended and, thus, possibly missed. Learning performance corresponded to the observed pattern of viewing behavior. The results suggested that the amount of time that can be spent reading and inspecting illustrations is a major source of differences in subsequent learning outcomes.

Lower learning performance is usually attributed to a higher workload. However, the direct effects that attentional demands have on this workload are still not fully understood. In terms of the learners' perceptions on specific task characteristics one might ask to which parts of the learning material learners attribute attentional and cognitive demands. Does text drag visual attention away from illustrations? Or is reading disturbed by the presence of illustrations? How much time is needed to extract sufficient information from both, text and illustration? In order to gain a better understanding of viewing behavior as a measure of cognitive activities in multimedia learning the present study addresses two issues: (1) How does viewing behavior relate to the learners' perceptions of the instructional material? (2) And how does viewing behavior vary with different sources of cognitive load like modality of text presentation and pacing of instruction?

Theoretical framework: Cognitive Load Theory

A framework to describe working memory load in learning with multimedia instructions is provided by cognitive load theory (Sweller, 1999; Sweller, et al., 1998). In accordance with Baddeley's working memory model (Baddeley, 1986) the theory assumes partly independent sensory-specific subsystems for visual and auditory information. Visual information is processed in a visuo-spatial sketchpad; auditory information is processed in an auditory loop. Both subsystems have limited capacities. How many of these limited capacities are occupied (or how much the processing systems are "loaded") during learning depends on several learner and task characteristics.

The theory differentiates between *intrinsic* and *extraneous* cognitive load. Intrinsic cognitive load refers to the load caused by the content of a learning material. It is determined by an interaction between the nature of the material and the expertise, prior knowledge, and cognitive abilities of the learner. Extraneous cognitive load refers to the presentation format of the material, or the activities required of a learner. Extraneous load is what can be affected by manipulating instructional design. Thus, proper instructional design is concerned to reduce extraneous cognitive load, i.e. to minimize the capacities required to successfully encode all important parts of an information source.

In terms of cognitive load theory, presenting an expository text accompanied by illustrations in spoken rather than written form is a way to vary extraneous cognitive load. Several studies showed that presenting text in spoken rather than written form lowers the cognitive load perceived by the learner (Kalyuga, et al., 1999, 2000; Tabbers, 2002; Tabbers, Martens, & Van Merrienboer, 2001; Tindall-Ford, et al., 1997) and leads to better learning outcomes (Brünken & Leutner, 2001; Kalyuga et al., 1999; 2000; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi, et al., 1995; Tindall-Ford et al., 1997). The theoretical explanation for this well-established modality effect offered by cognitive load theory rests on the limitation of the visual processing system. If expository text is added to an illustration in written form, both materials have to be processed by the visual processing system. Under these conditions, an overload of the visual system is more likely to occur than in spoken text presentation. If text is spoken rather than written, less information needs to be processed in the visual system while the processing of verbal information only requires capacity of the auditory system.

Cognitive overload in the visual processing system can occur whenever written text and illustrations are presented concurrently. In this case learners need to split their visual attention between both information sources. To integrate information in working memory, for example some portion of text must be held active while learners search for corresponding information in the illustration. Several studies showed that lowering the need for visual search by placing portions of written text close to referential parts of accompanying illustrations (Chandler & Sweller, 1991; Sweller & Chandler, 1994; Sweller, et al., 1990; Tindall-Ford et al., 1997) or by visually cueing appropriate referents (Kalyuga et al., 1999; Reitmayr, 2003; Tabbers, et al., 2004) leads to a lower perceived cognitive load and/or better learning outcomes. Thus, visual properties of the learning material are a potential source of extraneous cognitive load. The actual (over-)load of the visual system in these studies, however, is only inferred from subjective or subsequent measures. While also varying visual properties of the learning material in Chapter 2, I put subsequent learning outcomes in relation to directly objected viewing behavior. The observed fixation patterns during learning helped describing the actual attentional and perceptual demands of the instructional material and its potential effects on subsequent learning outcomes. In order to continue this approach it is necessary to investigate how viewing behavior relates to the concept of cognitive load and how the method of eye tracking contributes to its measurement.

Measuring cognitive load in multimedia learning

As other psychological constructs, that of cognitive load is not directly observable. The most widely used indicators of cognitive load are rating scales and task performance measures (Brünken, et al., 2003; Paas, et al., 2003). Self-ratings of cognitive load have proven to be a reliable measure, i.e. people are able to introspect on their cognitive processes and have no difficulty giving a numerical indication of their perceived cognitive load (Gopher & Braune, 1984). A more objective observation is given by task performance. For example, differences in subsequent learning outcome measures are commonly attributed to different cognitive loads during acquisition. Note, however, that in cognitive load theory poorer learning performance is related to a factual *overload* of the cognitive system. As long as the capacity of working memory is not exhausted, no differences in subsequent performance are expected to occur even if the total amount of cognitive load might have differed between learners.

Within the area of instructional design researchers only recently complemented traditional measures of cognitive load with more direct observations during the learning phase, namely secondary task performance and physiological measures. Dual task methodology, although delivering highly sensible and reliable measures of cognitive load during the learning process, is applied in only few studies (e.g. (Brünken et al., 2002; Brünken, et al., 2004; for a recent review see Paas et al., 2003). One reason is that this method undermines the ecological validity of the learning task. The very logic of dual tasks is to introduce a competition for limited cognitive resources. Thus, the secondary task interferes with the primary learning task and does not allow estimating the actual cognitive load evoked by the learning material. Less interfering with the learning process are physiological measures. The theoretical rationale for these techniques is that changes in physiological variables reflect changes in the cognitive functioning (Paas et al., 2003). Recent research applying measures of eye activity identified pupillary dilation and blink rate to correlate with fluctuating levels of cognitive load (Van Gerven, et al., 2004; Van Orden, et al., 2001).

The relation of directly measured *overall* cognitive load to attributes of the learning material can still only be accomplished by varying these attributes. There is no direct or obvious connection between indices of cognitive load and the *contents* of the learning material. In the context of eye activity measures eye tracking offers a direct indication of which part of a visual stimulus is currently processed. Fixation or gaze durations are assumed to map onto the amount of cognitive activity associated with the fixated area. Thus, it is reasonable to object which area of a learning material is attended in order to estimate the amount of cognitive resources devoted to that area. In contrast to common indices this observation does not deliver a measure of overall cognitive load. But it can be used to estimate the relative load of information located in a discrete area in comparison to other information areas of the learning material at hand.

So far, only few studies investigated eye movements in multiple information sources like concurrent presentation of text and pictures (Carroll, et al., 1992; d'Ydewalle & Gielen, 1992; Hegarty, 1992a, 1992b; Hegarty & Just, 1993; Rayner, et al., 2001; Tabbers, et al., 2002; Underwood, et al., 2004). The existing studies consistently revealed that viewers read several portions of text *before* they inspect corresponding parts of the pictorial information. This general finding is in accordance with findings reported in Chapter 2. Varying the modality of text presentation in a multimedia explanation it was found that illustrations were inspected significantly shorter whenever written text was presented compared to presentation conditions in which the same text was spoken. Within written text conditions text was read first before illustrations were inspected and learners spent relatively more time reading than inspecting illustrations. Thus, written text proved to be a highly salient stimulus for visual attention allocation that is at least initially preferred to illustration.

Since learners had to split their visual attention between written text and associated illustrations they alternated between reading text and inspecting illustrations several times. Due to the amount of visual attention on written text and the alternation between text and illustration some of the visual information was possibly missed or less thoroughly processed compared to spoken text groups. In Chapter 2 I concluded that especially illustrations suffered from the split of visual attention. Most obviously, illustrations were much less attended whenever text was written rather than spoken. Furthermore participants in written text conditions performed worse than their counterparts in the spoken text conditions especially in a visual memory task. However, some evidence also suggested that the presence of illustrations – especially when they were animated – might have disturbed reading comprehension.

In order to understand better how a particular fixation pattern relates to subsequent learning outcomes we might consider cognitive load as an intermitting variable between visual attention allocation and learning. Hence, the present experiments ask learners to introspect their cognitive processes while learning with the material. Besides common rating-scales of global, intrinsic, and extraneous cognitive load (e.g. Kalyuga et al., 1999; Paas & Van Merrienboer, 1993, 1994; Swaak & de Jong, 2001) subjective time estimation will be applied as an alternative index of cognitive load (Fink & Neubauer, 2001) by asking learners to rate the appropriateness of pacing of instruction. Furthermore I will introduce specific questions on several design attributes. Asking learners directly if they missed parts of text or parts of an illustration or how difficult it was to connect textual and pictorial information will help identifying the critical attentional and cognitive demands of the instructional material. Furthermore it can help understanding the time course of fixations on illustrations and written text.

Considering subjective time estimations or the time learners spent inspecting a discrete part of information leads us to a critical aspect of cognitive load theory and measurement: the time on task. In most studies on the modality effect time on task – or better: presentation duration – is recognized as a possible source for cognitive load (e.g. Mousavi et al., 1995). In order to control time on task presentation duration is often determined by the pace of spoken text (e.g. Mayer & Moreno, 1998; Moreno & Mayer, 1999). The pace of an instruction can be seen as an independent source of cognitive load. Just like visual cueing (Kalyuga et al., 1999), the pacing of instruction possibly interacts with the modality effect. If appropriately attending to important information is crucial for successful learning, cognitive load should be lower for longer presentation durations. Eventually, the modality effect might disappear as soon as the timing is appropriate to fully attend both information sources, i.e. to read written text and inspect illustrations. This should be observable in viewing behavior. If the modality effect is due to the fact that learners miss parts of important information when they split their visual attention between written text and illustrations under time constraints, fixation patterns are expected to vary with presentation duration. For longer presentation durations relatively more viewing time will be devoted to the formerly “missed” information. Before testing these hypotheses in Experiment 2, rating scales on cognitive load and on more specific design attributes of a particular learning material will be compared with fixation patterns in Experiment 1.

Experiment 1: The influence of animated and static illustration on viewing behavior and the modality effect

The purpose of Experiment 1 is to introduce self-ratings of cognitive load in a common setting of multimedia learning and to compare these ratings with learners’ viewing behavior. Material and experimental variation are adopted from Experiment 2 of Chapter 2. In that experiment we used a multimedia explanation about the process of lightning formation. Illustrations were presented concurrently with expository text that was either spoken or written. Illustrations were further varied as being animated or static. Thus, the material contains two variations of attentional salience for visual attention allocation: the presence or absence of written text and the presence or absence of apparent motion in the illustrations.

Viewing behavior is expected to replicate the findings of Experiment 2 in Chapter 2. Illustrations will get less visual attention if text is written rather than spoken. Within written text conditions, text is expected to be attended first and to capture relatively more visual attention than accompanying illustration. Descriptively, participants in Experiment 2 of Chapter 2 spent some more time on animated than on static illustration. The current Experiment might reveal if the assumed higher salience of animated compared to static illustration becomes statistically evident.

Cognitive load is expected to be lower in spoken than in written text presentation. The need to split visual attention between written text and illustration should be perceived as more demanding than if text is spoken. Participants are further expected to attribute their split attention and perceived cognitive load to distinct aspects of the presentation format. If the higher portion of time spent reading indicates a loss of pictorial information, participants might report that they missed part of the illustrated information or felt distracted from inspecting illustrations by the textual presentation format. Furthermore a modality effect should be observable in rating the pace of presentation as less appropriate whenever text is written rather than spoken.

Learning outcomes in the prior study were quite complex. A modality effect occurred in a test of visual memory. For verbal memory the modality effect was restricted to animated illustration indicating that visual cues might rather hinder than facilitate learning if text is written. It will be exciting to see if this pattern can be found again. Furthermore, self-ratings might reveal if animated compared to static illustration is perceived as helpful or bothersome for learning.

Method

Participants and Design

60 students of the Justus-Liebig-University Giessen participated in the experiment in partial fulfillment of a course requirement. All participants were native German speakers and had normal or corrected to normal vision. Participants were randomly assigned to one of four experimental groups. The experimental design was identical to Experiment 2 in Chapter 2. 15 participants served in each of four experimental groups (derived from a 2x2 experimental design) viewing either animated illustrations with spoken text, animated illustrations with written text, static illustrations with spoken text, or static illustrations with written text. The groups did not differ in prior knowledge. Mean values varied around 2 for self-estimated prior knowledge (on a 5-point scale from very little to very much) and between 4 and 5 checked items on a checklist consisting of 11 domain-related items.

Materials and apparatus

The learning material consisted of a 16-step multimedia instruction on the formation of lightning. The instruction showed a sequence of 16 illustrations depicting the motion of cool air that becomes heated; heated air rising up and forming a cloud; the rising of the cloud beyond the freezing level; drops of water and ice crystals moving up and down within the cloud, colliding, and causing electrical charges to arise; heavy drops and crystals falling down and producing downdrafts; a stepped leader of negative charges moving down to high objects on the ground; and positive charges moving up to the cloud producing a flash light.

The presentation was programmed in Flash 4.0 (Macromedia, 1999). Static illustrations were prepared to be informationally equivalent to animated illustrations. Simple movements like “cool air moving over a warmer surface” were indicated by arrows. In the case that the final state of an animated illustration did not match the initial state (e.g. positive charges in the cloud moving to the top and negative charges in the cloud moving to the bottom of the cloud), static illustrations visualized the process leading to the final state. The illustrations were accompanied by expository text that was either spoken or written inside the animation frame (*Figure 7*). The whole text had a length of 281 words, varying between 9 and 26 words per scene. Scene length was matched to the number of words per scene with a rate of 82 words per minute. For spoken text conditions, text was spoken in a male voice at a rate appropriate for this timing. The overall duration of the instruction was 206 seconds.

Figure 7. A selected frame and corresponding on-screen text for multimedia explanation on lightning formation.

The instruction was presented by a PC on a 21’’ color monitor, situated approximately 80 cm from the participant. Spoken text was presented by an audio system. Fixations were monitored by an Applied Science Laboratories’ corneal-reflectance and pupil-center eye tracker (ASL 504). Fixation position on the screen was measured with a sampling rate of 50 Hz and output to a PC, which controlled the recording, the camera, and the calibration (ASL Eyepos, E5000). Two additional monitors displayed the participants' tracked eye and its current fixation position on the stimulus screen to the experimenter.

Prior knowledge, performance, and cognitive load measures were assessed by paper-and-pencil tests. The material consisted of a participant questionnaire, a retention test, a four-item transfer test, a visual memory test, and two cognitive load rating sheets. The participant questionnaire asked for the participant's gender, age, profession and experience with meteorology. The retention test asked the participants to write down an explanation of how lightning works until they were told to stop. The transfer test contained four questions. Answers were asked to be given in a 5-alternative forced choice format. The visual memory test contained four sketch tasks. Answers were asked to be given on four sheets containing a simplified background scene of the original learning material. (More detailed descriptions of the tests can be found in chapter 2.)

The first of two rating sheets on cognitive load, given directly after presenting the multimedia instruction, contained the following three questions: (1) "How easy or difficult was it for you to learn something about lightning from the presentation you just saw?", (2) "How easy or difficult would you consider the *content*?", (3) "How pleasant or bothersome would you consider the *presentation format*?". Participants were instructed to place a check mark for each question on a 7-point rating scale from very easy (pleasant), easy (pleasant), rather easy (pleasant), medium, rather difficult (bothersome), difficult (bothersome), to very difficult (bothersome). Question one is a standard item for subjective ratings of cognitive load (e.g. Kalyuga et al., 1999; Paas & van Merrienboer, 1993, 1994). Questions two and three are introduced to differentiate between intrinsic cognitive load, i.e. due to an interaction between learner and content, and extraneous cognitive load, i.e. due to the presentation format (e.g. Swaak & de Jong, 2001).

The second rating sheet, given after completion of the performance tests, contained 9 detailed statements on the presentation: (1) "I would have preferred to stop the presentation myself at certain points", (2) "I would have preferred to look at some illustrations again", (3) "I would have preferred to rewind and repeat parts of the text", (4) "I missed parts of the textual information", (5) "I missed parts of the illustrations", (6) "It was difficult for me to relate textual and pictorial information to each other", (7) "The illustration distracted me from textual information", (8) "The textual information distracted me from the illustration", and (9) "How did you perceive the presentation pace? The pace was …". Statements 1 to 8 had to be rated on a 6-point scale from completely false, false, rather false, rather true, true to completely true. Statement 9, concerning the pace of presentation had to be answered on a 7-point scale from very slow, slow, rather slow, optimal, rather fast, fast, to very fast.

Procedure

Participants were tested in single sessions and were randomly assigned to one of the four treatment groups. They were given general instructions explaining the procedure and introducing the topic. Participants were instructed to acquire as much information as possible about the formation of lightning from the multimedia presentation in order to perform subsequent tasks. Next, participants were asked to fill out the questionnaire. Then, participants were seated in front of the stimulus PC and the eye-tracking system was calibrated. After that, the experimenter started the multimedia presentation. After participants had seen the presentation they rated their perceived cognitive load on the first (of two) rating sheets. Then they were given instructions to work on the retention test. Participants had 8 min to process the test. The retention test was followed by the transfer test. Instructions for the transfer test were handed out together with the first of four questions and the remaining questions were handed out successively. Participants were given 5 min to answer all questions. After the transfer test the visual memory test was given. Participants had 5 min to work on the sketches. Tasks were handed out successively. After completing the visual memory test, participants were given three additional minutes to write comments on their sketches in a different color. This was done in order to facilitate scoring of ambiguous sketches. Finally, the second rating sheet was handed out. After completion, participants were debriefed and thanked for their participation. The experimental session took about 50 min.

Results

Subjective ratings

No participants were excluded from further analyses of subjective ratings since all participants filled in the rating sheets appropriately. Thus, the following calculations were based on n=60 data sets.

In addition to rating cognitive load in general, two more detailed estimations were requested. Besides estimating the overall load while learning with the instruction, participants were asked to distinguish between load caused by content (i.e. intrinsic cognitive load) and load caused by presentation format (i.e. extraneous cognitive load). Although one can argue that this differentiation is quite difficult for participants or that learners are not sensitive to this differentiation at all, correlations between the three items varied between $r=.29$ ($p<.05$) and $r=.69$ ($p<.01$) indicating that participants answered the questions differently. Thus, separate analyses were conducted for each of the three items. Analyses of variance (ANOVA) with the between-subjects factors text presentation (spoken vs. written) and illustration (animated vs. static) provided the following results. For overall cognitive load ANOVA revealed no main effect for text presentation format ($F(1,56)=1.35$, $MSE=1.67$, $p>.10$), no main effect for illustration format ($F(1,56)=1.35$, $MSE=1.67$, $p>.10$), and no interaction ($F(1,56)=1.35$, $MSE=1.67$, $p>.10$). Concerning difficulties with the content of the presentation, ANOVA revealed a marginally significant main effect for text presentation format ($F(1,56)=3.40$, $MSE=3.75$, $.10<p<.05$, $\eta^2=.06$), but no main effect for illustration format ($F(1,56)=1.23$, $MSE=1.35$, $p>.10$), and no interaction ($F(1,56)=2.56$, $MSE=2.82$, $p>.10$). Participants tended to report more difficulties with the content (!) if textual information was written. However, no significant effects were obtained concerning the load caused by the presentation format (all Fs<1).

After completion of the performance tasks, participants were asked to give more detailed descriptions of their cognitive load by judging statements about several aspects of the presentation. Before analyzing each of the items separately one can – in order to control for alpha-inflation – consider the nine items as a multidimensional scale of cognitive load. Thus, a multivariate analysis of variance (MANOVA) with the between-subjects factors text presentation (spoken vs. written) and illustration (animated vs. static) and with the nine judgments as dependent measures was conducted. The MANOVA revealed a significant main effect of text presentation format ($F(9,48)=4.52$, *Wilks-Lambda*=0.54, $p<.001$, $\eta^2=.46$), no main effect for illustration format ($F<1$) and no interaction ($F(9, 48)=1.24$, *Wilks-Lambda*=0.81, $p>.10$). Post-hoc ANOVAs revealed that the main effect is explained by different judgments between spoken and written text presentation groups in two of the statements. Participants in the written text groups scored higher when asked if they were distracted from illustrations by the textual information ($F(1,56)=33.02$, $MSE=29.4$, $p<.001$, $\eta^2=.37$). Furthermore they estimated the pace as less appropriate than participants in the spoken text conditions ($F(1,56)=5.66$, $MSE=7.35$, $p<.05$, $\eta^2=.09$). Mean scores for all judgments are shown in *Table 5*.

Table 5
Mean values and standard deviations of rating scores for cognitive load items. Higher scores indicate a higher cognitive load or a higher agreement with the statement.

		Text presentation format			
		Spoken text		Written text	
Item description	Illustration format	*M*	*(SD)*	*M*	*(SD)*
Overall load (0-6)	Animation	1.9	(1.1)	1.9	(1.1)
	Static illustrations	1.9	(0.8)	2.5	(1.4)
Content (0-6)	Animation	1.5	(1.0)	1.5	(0.7)
	Static illustrations	1.3	(0.9)	2.3	(1.4)
Presentation format (0-6)	Animation	1.5	(0.9)	1.9	(1.1)
	Static illustrations	1.8	(1.0)	2.0	(1.3)
Stop presentation (0-5)	Animation	2.3	(1.6)	2.8	(1.6)
	Static illustrations	2.1	(1.8)	3.1	(1.6)
Review illustrations (0-5)	Animation	2.2	(1.8)	3.1	(1.2)
	Static illustrations	2.6	(1.9)	2.8	(1.4)
Repeat text (0-5)	Animation	2.4	(1.5)	2.5	(1.5)
	Static illustrations	2.6	(1.6)	3.1	(1.3)
Missed text (0-5)	Animation	2.5	(1.7)	1.5	(1.4)
	Static illustrations	2.3	(1.9)	2.0	(2.0)
Missed illustrations (0-5)	Animation	1.7	(1.5)	1.6	(1.5)
	Static illustrations	1.4	(1.3)	1.9	(1.0)
Problems connecting text and illustration (0-5)	Animation	1.0	(1.1)	1.1	(1.3)
	Static illustrations	1.3	(1.1)	1.8	(1.0)
Distracted by illustration (0-5)	Animation	1.1	(1.4)	1.5	(1.1)
	Static illustrations	1.0	(1.5)	1.8	(1.1)
Distracted by text (0-5)	Animation	0.9	(0.7)	1.6	(1.1)
	Static illustrations	0.5	(0.6)	2.6	(1.2)
Pacing of instruction (0-6)	Animation	3.1	(1.2)	3.6	(0.7)
	Static illustrations	2.5	(1.2)	3.5	(1.3)

Performance Measures

Scores for performance measures were obtained in the following manner. Performance on the retention test was scored with two scorers being unaware of the participant's identity. Participants were given 1 point for each of nineteen main ideas of the causal chain of lightning formation. The inter-rater reliability for the scores was r=.96. Mean values of scores obtained by the two scorers were used in the following analyses. Scores for the problem-solving transfer were obtained by counting the number of correct marks in the forced-choice items, i.e. a maximum of 4 points could be obtained in the transfer test.

Visual memory was scored by two independent scorers being unaware of the participant's identity. Participants were given 1 point for each appropriate and identifiably sketched visual element, with a maximum of 2 points obtainable for sketches 1, 3, and 4, and a maximum of 3 points for sketch 2. Inter-rater-reliability for the visual memory test was *r*=.95.

No participants had to be excluded from further analyses of the performance measures, thus the following calculations were based on n=60 data sets. Mean scores and standard deviations for all three measures are shown in Table 6.

Table 6
Mean values and standard deviations of performance scores for retention, transfer, and visual memory tests.

		Text presentation format			
		Spoken text		Written text	
Test	Illustration format	*M*	*(SD)*	*M*	*(SD)*
Retention	Animation	11.3	(3.7)	11.1	(3.5)
	Static illustrations	9.2	(3.6)	10.3	(3.2)
Transfer	Animation	3.6	(0.7)	3.3	(1.1)
	Static illustrations	3.4	(0.7)	3.4	(0.7)
Visual memory	Animation	6.7	(2.5)	7.2	(1.4)
	Static illustrations	6.3	(2.1)	6.0	(2.4)

Analyses of variance (ANOVA) with the between-subjects factors text presentation (spoken vs. written) and illustration (animated vs. static) provided the following results. For retention ANOVA revealed no significant main effect for text presentation format (F<1), no main effect for illustration format (F(1,56)=2.62, MSE=32.27, p>.10), and no interaction (F<1). ANOVA on problem solving transfer revealed no significant main effects for text presentation format and illustration format and no interaction (all Fs<1). Also, ANOVA on scores of the visual memory test revealed no significant main effect for text presentation format (F<1), no main effect for illustration format (F(1,56)=1.86, MSE=8.63, p>.10), and no interaction (F<1).

Viewing behavior

For the n=60 participants calibration failed in 16 cases. The remaining 44 cases were processed in the following manner. Viewing positions were transformed into fixations and saccades using ASL-Eyenal software. Areas of interest (AOI) were defined to cumulate single fixations and saccades into viewing times and numbers of fixations on text and illustration. An AOI in the presentation was a part in which either a portion of text or an illustration was displayed. *Figure 8* shows an area of written text and an area of illustration for one scene of the presentation. In order to detect inaccurate calibration the resulting viewing times were further inspected in the following manner. Data sets in which viewing time on AOIs summed up to less than 75% of the total presentation time were taken as possibly invalid. Applying this criterion, 13 further participants had to be excluded. Thus, the following analyses were calculated with a set of 31 data cases.

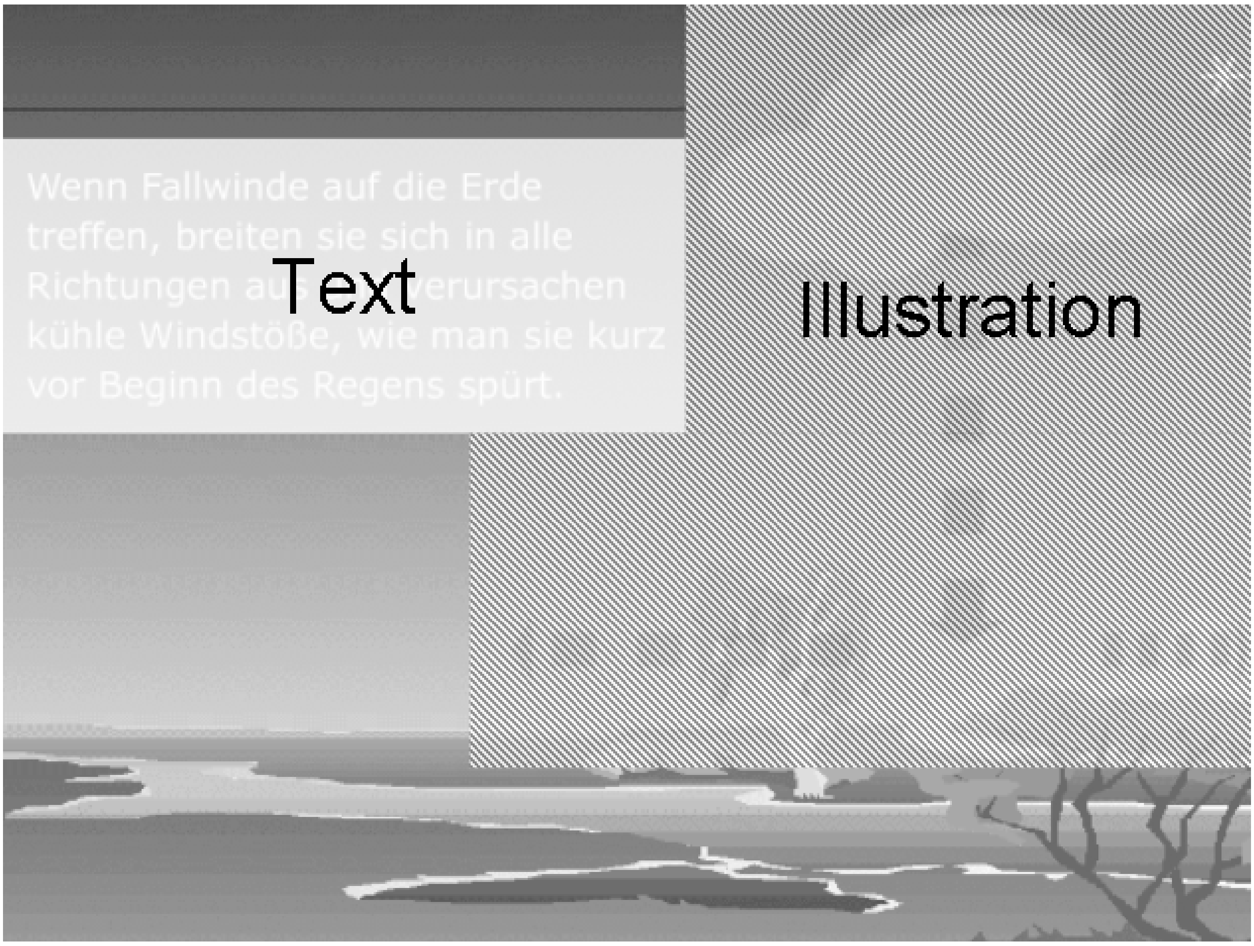

Figure 8. An example of areas of interest (AOI) for illustrations (striped) and on-screen text (white). Note that the areas vary from scene to scene depending on text length and location of the illustrations.

Overall, participants spent 182 s (SD=10.97) or 88% of their inspection time on AOIs. Means and standard deviations of viewing time on illustration and text as well as summed viewing time on AOIs for each group are shown in Table 7.

Table 7
Mean viewing durations on areas of interest (AOI) for illustrations, on-screen text (if present) and sum of text and illustration areas (Total AOI) for each of the four experimental groups.

			Text presentation format			
			Spoken text		Written text	
			M	*(SD)*	*M*	*(SD)*
Illustration format	Animation	Illustration	177.8	(12.5)	86.5	(18.8)
		Text	-	-	97.8	(16.2)
		Total AOI	177.8	(12.5)	184.3	(6.7)
	Static illustrations	Illustration	173.5	(11.8)	60.3	(12.7)
		Text	-	-	121.7	(13.0)
		Total AOI	173.5	(11.8)	182.0	(5.2)

An ANOVA with the between-subjects factors text presentation (spoken vs. written) and illustration (animated vs. static) and with summed fixation times on illustration as dependent measure revealed a main effect for text presentation format, $F(1,27)=321.88$, $MSE=70{,}756$, $p<.001$, $\eta^2=.92$. Participants in the spoken text groups spent more time inspecting illustrations than participants in the written text groups. There was also a main effect for illustration format, $F(1,27)=7.15$, $MSE=1{,}571$, $p<.05$, $\eta^2=.21$. Participants spent more time inspecting animated than static illustrations. The interaction between text presentation and illustration format was marginally significant, $F(1,27)=3.67$, $MSE=807.45$, $.10>p>.05$, $\eta^2=.12$. The marginal interaction indicates that the main effect of illustration format was mainly caused by written text presentation conditions. Within written text groups more visual attention was given to animated than to static illustration. This result was mirrored by the ratio of viewing times on written text and illustration. The mean ratio of the time spent reading to the time spent inspecting illustrations was 1.5. Participants spent 50% more time reading than inspecting illustrations. Comparing these ratios between animated and static illustration groups revealed a significant difference ($t(13)=2.68$, $p<.05$). As shown in Table 7 participants spent 26 sec. more time inspecting animated rather than static illustrations. Consequently they also spent on average 24 sec. less time reading.

Exploring the fixation paths revealed that participants in written text groups split their visual attention between text and illustration differently. Overall, participants in written text conditions alternated between reading text and inspecting illustrations 3.7 times per scene. The static illustration group alternated significantly more often than the animated illustration group ($t(13)=2.42$, $p<.05$). Within the first five fixations after a scene change 87% of the fixations were on text. Animated and static illustration groups did not differ in their initial viewing behavior ($t(13)=1.61$, $p>.10$). Participants in both groups apparently started reading as soon as new text occurred after a scene change.

Discussion

The observed pattern of viewing behavior replicated the results of Experiment 2 in Chapter 2. Participants in written text conditions split their visual attention between text and illustration, attending to written text first and spending an equal or even larger amount of time reading text than inspecting illustrations. Consequently they spent much less time inspecting illustrations than spoken text groups.

The potential of animated illustration to shift visual attention towards illustration, descriptively observable in Experiment 2 in Chapter 2, became statistically evident now. Relatively more time was spent inspecting illustrations if they were animated rather than static. Furthermore, animation “locked” visual attention. Participants receiving animated illustrations with written text alternated less often between text and illustrations than if illustrations were static.

Self-ratings of cognitive load allow a more detailed view on these attentional aspects of presentation format. First of all, participants reported a marginally higher difficulty with the content of the learning material for written compared to spoken text presentation. Although this outcome delivers converging evidence for a modality effect it is somewhat surprising that participants attribute their different cognitive loads to the content and not to the format of presentation. Obviously, participants were not aware of the experimental variation. Participants in written text conditions might not have considered the particular presentation format as unusual and/or spoken text presentation as helpful to reduce their cognitive load.

Asked in more detail, participants reported to have been distracted from inspecting illustrations when text was written rather than spoken. This outcome supports the interpretation of results in the studies of Chapter 2. The high salience of written text, indeed, appears to drag visual attention away from illustration. Furthermore, participants rated presentation time as less appropriate in written text conditions. Thus, the modality effect can be described as a distracting effect of written text in a time limited presentation condition.

Comparing viewing behavior with subjective load, the split of visual attention was subjectively time consuming. Participants felt they needed more time to sufficiently attend to all offered information sources. As a consequence, one should expect that with longer presentation duration learners devote relatively more time to illustrations than to written text. Once enough time is given to attend and integrate all information sources the modality effect should disappear.

Although participants perceived written text as comparably uncomfortable, performance measures did not reveal any significant differences. Apparently, there was still enough time to compensate for attentional and cognitive demands caused by presentation format. In fact, the pacing of instruction was lower in our material than in comparable studies by Mayer and Moreno (Mayer & Moreno, 1998; Moreno & Mayer, 1999). Faster pacing might increase the modality effect in the same way as slower pacing might decrease it. To test these hypotheses Experiment 2 was conducted.

Experiment 2: The influence of pacing on the modality effect

The purpose of Experiment 2 is to examine the influence of pacing on the modality effect. Varying the pacing of instruction independently from text modality the following hypotheses will be tested. Since more time is spent on both information sources, longer presentation duration should lead to less cognitive load and better learning performance. Within written text conditions viewing behavior is expected to change with pacing. If illustrations are missed or processed superficially in short presentation durations I expect that relatively more time is devoted to them than to written text the longer the presentation lasts. As a consequence the modality effect should be stronger for shorter presentation durations. Longer presentation durations may help compensating the cognitive load caused by split visual attention.

Note that the interaction of pacing and text modality might differ between self-ratings and learning outcomes. Poorer learning theoretically only occurs if the learning material causes a cognitive overload. Learners might be able to compensate higher attentional and/or cognitive demands of the learning material and reach a comparable level of learning performance, thus not suffering from a cognitive overload. In Experiment 1 self-ratings revealed modality effects in different aspects of the learning material while no differences occurred in outcome measures. Participants might still be sensitive for differing cognitive demands while not suffering from these demands in performing subsequent tasks.

Method

Participants and Design

90 students of the Justus-Liebig University Giessen participated in the experiment in partial fulfillment of a course requirement. All were native German speakers and had normal or corrected to normal vision. Participants were randomly assigned to one of six experimental groups, with 15 participants in each group. The experimental groups were derived following a 2 (spoken vs. written text) x 3 (fast, medium, and slow pace) experimental design. The groups did not differ in prior knowledge. Mean values varied around 2.5 for self-estimated prior knowledge (on a 5-point scale from very little to very much) and between 4.5 and 5 checked items on a checklist consisting of 11 domain-related items.

Materials and apparatus

The learning material consisted of the same multimedia instruction as in Experiment 1. Animated illustrations on the formation of lightning were accompanied by expository text. The text was either presented in spoken or written format. The variation of pacing was derived in the following manner. In the fast condition, timing was set on a ratio of 120 words per minute resulting in a presentation duration of 140s. This pace approximates a timing originally applied in Mayer and Moreno (1998) by simply adjusting the pace of presentation to a normal speaker's rate. Medium and slow paces were obtained by reducing the ratio successively with a factor of 0.75. Thus, the ratio was 90 words/min for medium pace and 67.5 words/min for slow pace resulting in durations of 187s and 249s respectively. Stimulus presentation and eye tracking equipment as well as prior knowledge, subjective ratings and performance measures were the same as in Experiment 1.

Procedure

The procedure was identical to Experiment 1.

Results

Subjective ratings

Subjective ratings were computed in the same way as in Experiment 1. No participants had to be excluded from further analyses, since all participants filled in the rating sheets appropriately. Thus, the following calculations are based on n=90 data sets.

Answers to items on overall cognitive load, difficulty of the content and appropriateness of the presentation format correlated between r=.28 and r=.59 (all ps<.01). Separate analyses were conducted for each of the three items. Mean scores for each experimental group are shown in Table 8.

Table 8
Mean values and standard deviations of rating scores for cognitive load items. Higher scores indicate a higher cognitive load or a higher agreement with the statement.

		Pacing					
		fast (140 s)		medium (187 s)		slow (249 s)	
Item description		*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Overall load (0-6)	Spoken text	2.0	(1.1)	1.9	(1.2)	1.7	(1.3)
	Written text	2.1	(1.1)	1.9	(0.9)	1.3	(0.8)
Content (0-6)	Spoken text	1.9	(1.2)	1.7	(1.0)	1.7	(1.1)
	Written text	2.2	(1.0)	1.5	(0.6)	1.9	(0.9)
Presentation format (0-6)	Spoken text	1.5	(1.2)	1.2	(0.9)	1.3	(0.8)
	Written text	2.9	(1.6)	1.3	(1.2)	1.1	(1.0)
Stop presentation (0-5)	Spoken text	3.5	(1.6)	3.0	(1.4)	2.1	(1.4)
	Written text	3.9	(1.2)	3.5	(1.7)	2.9	(1.5)
Review illustrations (0-5)	Spoken text	3.3	(1.8)	2.9	(1.4)	2.6	(1.6)
	Written text	3.6	(1.3)	2.9	(1.2)	2.9	(1.7)
Repeat text (0-5)	Spoken text	3.3	(1.5)	3.4	(1.4)	2.9	(1.3)
	Written text	3.9	(0.7)	3.1	(1.3)	2.9	(1.8)
Missed text (0-5)	Spoken text	3.5	(1.9)	2.3	(1.4)	2.3	(1.2)
	Written text	1.3	(1.2)	1.3	(0.7)	0.9	(0.9)
Missed illustrations (0-5)	Spoken text	1.5	(1.2)	1.5	(0.8)	1.9	(1.7)
	Written text	2.4	(1.7)	1.5	(1.4)	1.5	(1.0)
Problems connecting text and illustration (0-5)	Spoken text	1.4	(1.0)	1.3	(0.9)	1.1	(1.0)
	Written text	1.7	(1.6)	1.2	(1.0)	0.9	(1.1)
Distracted by illustration (0-5)	Spoken text	1.4	(1.0)	1.7	(1.3)	1.4	(0.9)
	Written text	1.5	(1.1)	1.7	(1.2)	1.5	(1.2)
Distracted by text (0-5)	Spoken text	1.3	(0.9)	0.9	(0.6)	1.0	(0.8)
	Written text	2.9	(1.3)	1.7	(1.2)	1.9	(1.4)
Pacing of instruction (0-6)	Spoken text	3.7	(0.6)	3.4	(0.7)	2.8	(0.9)
	Written text	4.5	(0.8)	4.0	(0.8)	2.8	(1.0)

Analyses of variance (ANOVA) with the between-subjects factors text presentation (spoken vs. written) and pacing (fast, medium, slow) provided the following results. For overall cognitive load ANOVA revealed no main effect for text presentation format ($F<1$), no main effect for pacing ($F(2,84)=2.18$, $MSE=2.54$, $p>.10$), and no interaction ($F<1$). Concerning difficulties with the content of the presentation, ANOVA revealed no main effect for text presentation format ($F<1$), no main effect for pacing ($F(2,84)=1.43$, $MSE=1.41$, $p>.10$), and no interaction ($F<1$). However, asking for the load caused by presentation format, an ANOVA revealed a marginally significant main effect for text presentation (F(1,84)=3.88, MSE=4.90, .10>p>.05, η^2=.04), a main effect for pacing (F(2,84)=7.20, MSE=9.10, p<.01, η^2=.15), and a significant interaction (F(2,84)=3.51, MSE=4.43, p<.05, η^2=.08). As shown in *Figure 9*, participants in written text conditions tended to perceive the presentation format as more bothersome compared to participants in spoken text conditions. More obviously, post-hoc Scheffé tests on pacing revealed that participants perceived the presentation format as more bothersome at fast pace than at medium and slow paces. The interaction revealed that especially in written text presentation the fast pace was rated as being more bothersome than in the other presentation conditions.

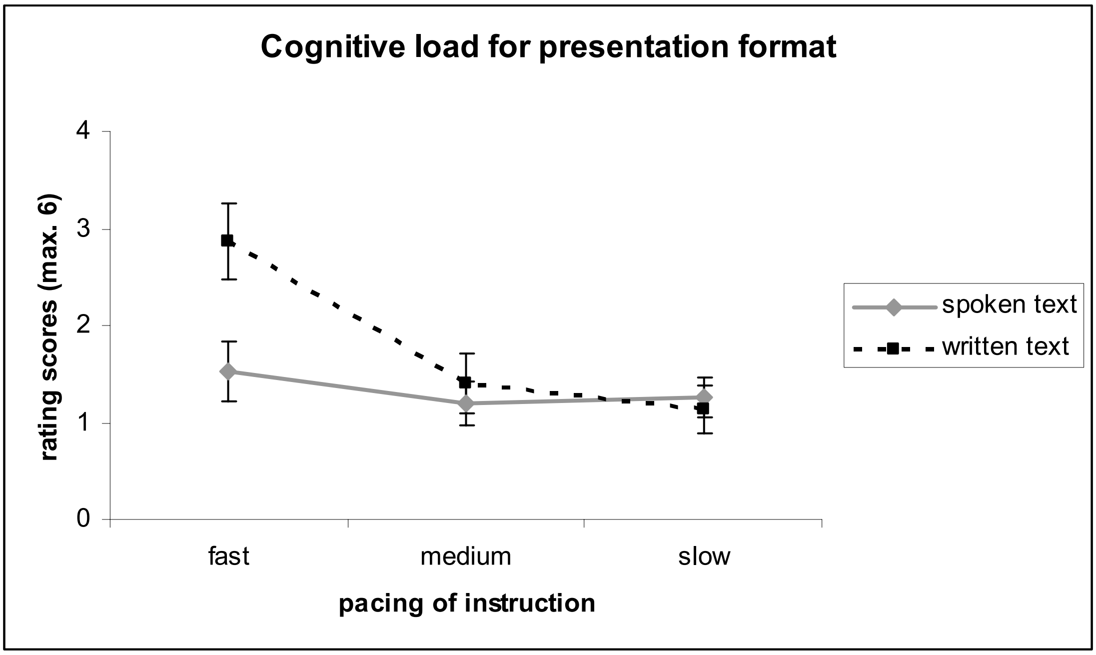

Figure 9. Mean ratings of cognitive load caused by the format of presentation.

After completion of the performance tasks, participants were asked to give more detailed descriptions of their cognitive load by judging statements about several aspects of the presentation. Mean scores for each experimental group are shown in Table 8. A multivariate analysis of variance (MANOVA) with the between-subjects factors text presentation (spoken vs. written) and pacing (fast, medium, slow) and with the nine judgments as dependent measures revealed a main effect of text presentation format (F(9,76)=9.358, *Wilks-Lambda*=0.47, p<.001, η^2=.53), a main effect for pacing (F(18,152)=2.84, *Wilks-Lambda*=0.56, p<.001, η^2=.25), but no interaction (F(18,152)=1.37, *Wilks-Lambda*=0.74, p>.10). Participants, although unaware of the experimental manipulation, significantly differed in their perceptions of the presentation depending on both, text presentation format and presentation duration. Post-hoc

ANOVAs revealed that the main effects were caused by different judgments in the following statements. Participants in the written text conditions were less likely to agree that they missed part of the textual information ($F(1,84)=30.63$, $MSE=49.88$, $p<.001$, $\eta^2=.27$) but showed higher agreement when asked if they felt distracted from illustrations by textual information ($F(1,84)=21.03$, $MSE=25.6$, $p<.001$, $\eta^2=.20$). Furthermore, they estimated the pacing to be faster than participants in the spoken text conditions ($F(1,84)=7.21$, $MSE=4.9$, $p<.01$, $\eta^2=.08$). The main effect of pacing in the MANOVA was caused by the same three items but in a different order of effect sizes plus one additional item. As expected, participants estimated the presentation duration in accordance with the actual experimental variation, i.e. fast pace groups perceived the pacing as “rather fast”, medium pace groups between “rather fast” and “medium”, and slow pace groups between “medium” and “rather slow” ($F(2,84)=20.43$, $MSE=13.88$, $p<.001$, $\eta^2=.33$). Participants felt more distracted by textual information in the fast pace conditions ($F(2,84)=4.94$, $MSE=6.01$, $p<.01$, $\eta^2=.11$). Furthermore, the faster the actual pace the more participants would have liked to stop the presentation at certain points ($F(2.84)=4.892$, $MSE=10.544$, $p<.05$, $\eta^2=.10$). Finally, participants were more likely to agree that they missed part of the textual information when pacing was fast ($F(2,84)=3.28$, $MSE=5.34$, $p<.05$, $\eta^2=.07$).

Performance Measures

Scores for performance measures were obtained in the same way as in Experiment 1. No participants had to be excluded from analyses of performance measures. For retention and visual memory tests a second rater scored a subset of 20 participants’ data independently. Inter-rater-reliability for these subsets were $r=.93$ for retention and $r=.89$ for visual memory. Thus, the following analyses were conducted with the scores obtained by the first rater for n=90 data sets. Table 9 shows mean values and standard deviations of performance scores for retention, transfer and visual memory tests.

Table 9
Mean values and standard deviations of performance scores for retention, transfer, and visual memory tests.

		Pacing					
		fast (140 s)		medium (187 s)		slow (249 s)	
		M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Retention	Spoken text	10.0	(1.9)	11.7	(3.2)	11.3	(3.3)
	Written text	9.1	(5.5)	10.9	(3.6)	11.9	(3.5)
Transfer	Spoken text	3.2	(1.2)	3.7	(0.6)	3.7	(0.5)
	Written text	3.3	(0.8)	3.3	(0.8)	3.3	(0.8)
Visual Memory	Spoken text	5.9	(1.5)	6.8	(1.7)	6.6	(1.8)
	Written text	4.8	(2.6)	5.5	(2.2)	6.5	(2.0)

Analyses of variance (ANOVA) with the between-subjects factors text presentation (spoken vs. written) and pacing (fast, medium, slow) provide the following results. For retention ANOVA revealed a marginally significant main effect for pacing ($F(2,84)=2.78$, $MSE=36.78$, $.10>p>.05$, $\eta^2=.06$). Participants in the medium and slow pace conditions tended to remember more main ideas than in the fast presentation condition. No significant main effect for text presentation format and no interaction were obtained (Fs<1). ANOVA on problem solving transfer revealed no main effect for text presentation format ($F(1,84)=1.60$, $MSE=1.11$, $p>.10$), no main effect for pacing ($F<1$), and no interaction ($F(2,84)=1.07$, $MSE=0.74$, $p>.10$). ANOVA on scores of the visual memory test revealed a significant main effect for text presentation format ($F(1,84)=4.22$, $MSE=16.9$, $p<.05$, $\eta^2=.05$), a marginally significant main effect for pacing ($F(2,84)=2.58$, $MSE=10.35$, $.10>p>.05$, $\eta^2=.06$), but no interaction ($F<1$). Participants in spoken text conditions performed better on visual memory than participants in written text conditions. Furthermore participants tended to perform better the more time they had for inspecting the instruction.

Viewing behavior

Calibration failed in 5 cases. The remaining 85 cases were further processed in the same way as in Experiment 1. Viewing positions were transformed into single fixations and saccades. Fixations and saccades were further cumulated into viewing durations and numbers of fixations on areas of interest (AOI, see *Figure 8*). Four participants whose viewing times on AOIs summed up to less than 75% of total presentation time were excluded from further analyses. Thus, the following analyses were conducted with a set of 81 data cases.

Overall, participants spent 92% of their time viewing AOIs. Means and standard deviations of viewing time on illustration and text as well as summed viewing time on AOIs for each group are shown in Table 10.

Table 10
Mean viewing durations on areas of interest (AOI) for illustrations, on-screen text (if present) and sum of text and illustration areas (Total AOI) for each of the six experimental groups.

	Pacing											
	Fast (140s)				Medium (187s)				Slow (249s)			
	Spoken text		Written text		Spoken text		Written text		Spoken text		Written text	
Viewing times (sec.)	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Illustration	124.4	(6.7)	39.3	(15.3)	171.9	(2.8)	58.8	(27.9)	228.4	(10.7)	97.0	(21.3)
Text	-	-	91.5	(16.4)	-	-	116.7	(27.6)	-	-	139.4	(21.5)
AOI total	124.4	(6.7)	130.9	(6.4)	171.9	(2.8)	175.5	(5.0)	228.4	(10.7)	236.36	(8.1)

An analysis of covariance (ANCOVA) with the between-subjects factors text presentation (spoken vs. written) and pacing (fast, medium, slow) and with presentation duration as covariate on summed fixation times on illustration as dependent measure revealed a main effect for text presentation format, $F(1,75)=926.16$, $MSE=241,236$, $p<.01$, $\eta^2=.93$. Participants in the spoken text groups spent more time inspecting illustrations than participants in the written text groups. Pacing had no main effect on fixation times irrespective of presentation duration ($F<1$). The interaction between text presentation and pacing was significant, $F(1,75)=15.09$, $MSE=3,931$, $p<.01$, $\eta^2=.29$. Relative to the pacing of instruction more time was spent inspecting illustration the longer the presentation lasted. This interaction can be explained by the viewing behavior of participants in written text conditions. An ANOVA on the ratio of time spent viewing text to time spent inspecting illustrations as dependent measure and with pacing (fast, medium, slow) as between-subject factor revealed a significant difference ($F(2,37)=3.93$; $MSE=5.98$; $p<.05$; $\eta^2=.18$). As shown in *Figure 10*, this ratio dropped from 2.8 for fast pace to 2.1 for medium and 1.5 for slow pace of presentation, respectively. Post-hoc Tukey-tests revealed a significant difference between fast and slow pace conditions. Participants spent relatively more time inspecting illustrations compared to reading text the longer the presentation lasted. The value of 1 was not included in the 95% confidence intervals for each of the ratios. Thus, in all pacing conditions still relatively more time was spent reading than inspecting illustrations.

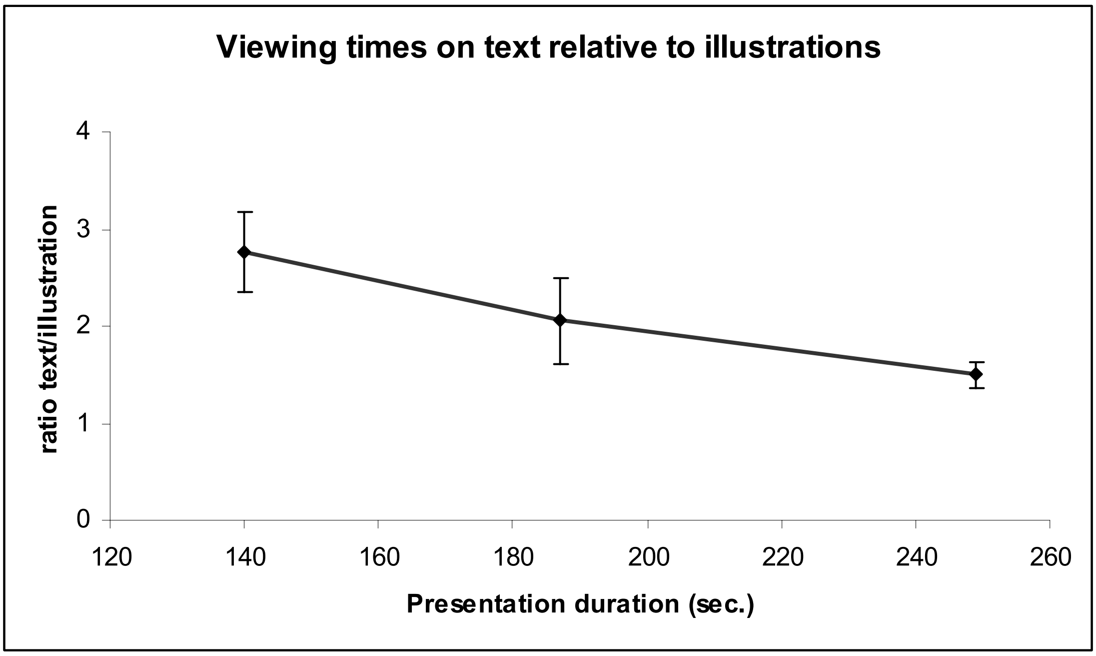

Figure 10. Ratio of viewing times on areas of written text to areas of illustration (with standard error bars).

Exploring the fixation paths revealed that participants in written text groups split their visual attention between text and illustration differently. On average, participants in written text conditions alternated between reading text and inspecting illustrations 0.27 times per second, i.e. once every 4 seconds, ranging from 0.22 for fast pace to 0.29 for medium and 0.30 for slow pace of presentation, respectively. An ANOVA with the between-subjects factor pacing (fast, medium, slow) on the number of alternations per second revealed a significant change of this aspect of viewing behavior with pacing ($F(2,37)=3.37$; $MSE=2.46*10^{-2}$; $p<.05$; $\eta^2=.15$). Post-hoc Tukey-tests revealed the difference between fast and slow pacing as significant. In slow presentation conditions learners alternated more often between text and illustrations than in fast presentation. Within the first five fixations after a scene change 91% of the fixations were on text. An ANOVA revealed that this amount did not significantly change with pacing ($F(2,37)=1.77$; $MSE=5.78*10^{-3}$; $p>.10$). In all three groups text is read almost immediately after a scene change.

Viewing behavior differs with respect to pace of presentation. However, the previous analyses did not reveal *when* these changes appear. Learners may adapt their viewing behavior during learning to the pace of presentation, i.e. time pressure in fast pacing may already have an impact on viewing behavior at the beginning of each scene. However, the apparently stable initial reading behavior suggests that changes only appear in the additional time given in slower pacing of instruction. In order to confirm this preliminary conclusion the following analyses compare mean viewing times and numbers of alternations only for a fixed time interval, i.e. each individual's viewing behavior is only regarded for the "minimal" scene lengths of the fast pacing condition. An analysis of variance (ANOVA) with the between-subjects factors text presentation (spoken vs. written) and pacing (fast, medium, slow) on summed fixation times on illustration replicated the main effect for text presentation format, $F(1,75)=907.09$, $MSE=145.502$, $p<.01$, $\eta^2=.92$. Participants in the spoken text groups spent more time inspecting illustrations than participants in the written text groups. Pacing had no main effect on fixation times in the matched time interval ($F<1$). The interaction between text presentation and pacing, significant in the ANCOVA for overall fixation times, failed statistical significance now, $F(1,75)=1.18$, $MSE=188.46$, $p>.10$. This lack of interaction becomes also evident in the ratio of the time spent reading to the time spent inspecting illustrations within written text conditions. An ANOVA with the between-subjects factor pacing (fast, medium, slow) on this ratio revealed no differences within the minimum time interval ($F<1$). Participants alternated between reading text and inspecting illustrations 0.25 times per second within the minimum scene durations. An ANOVA revealed that also this ratio did not vary with pacing ($F<1$). Taken together, participants' viewing behavior is not distinguishable when viewing behavior is only regarded in the time interval of each scene that is equal for all participants.

Discussion

A modality effect was observed in viewing behavior, subjective ratings and subsequent task performance. Measures of viewing times and visual memory replicated the findings of previous experiments. Learners spent less time inspecting illustrations when the accompanying text was written rather than spoken. As a consequence, they performed worse on a subsequent visual memory task. This result once more indicates that the modality effect might be especially caused by a loss of pictorial information when learners have to split their attention between illustrations and written text. This interpretation is supported by self-ratings. Participants in written text conditions reported a higher distraction from illustration by the textual information than the spoken text groups. Furthermore, they estimated the pacing of instruction as less appropriate. Taken together, these statements indicate that participants would have liked to spend more time especially on illustrations.

The pacing of instruction proved to be an independent source of cognitive load. Participants' ratings as well as their performance in subsequent tasks revealed a higher cognitive load the faster the presentation pace was. Faster presentation was perceived as being more bothersome, participants would have liked to stop presentation at certain points, and they felt to have missed textual information. These statements were mirrored by their task performance. Participants tended to remember more main ideas and more aspects of illustrations the longer the presentation lasted.

As hypothesized the study revealed interactions of text modality with pacing. Participants who had to split their visual attention between written text and illustrations spent relatively more time inspecting illustrations the longer the presentation lasted. This viewing behavior confirmed the distracting effect of written text in time limited presentation as a likely source of the modality effect. The additional time given was devoted to inspecting illustrations rather than further reading. Interestingly, initial viewing behavior did not vary with presentation pace. In all written text conditions participants started reading immediately after a scene change. Deviating viewing behavior like the higher number of alternations and the relatively longer time spent inspecting illustrations in longer presentation durations obviously settled in the additional time given by slower pacing.

Written text also caused a higher perceived cognitive load especially if the pace was (too) fast. This interaction of text modality with pacing on the self-rating of presentation format failed to reach statistical significance in the performance measures. Descriptively, however, performance scores showed the expected pattern. As noted above, subjective rating scales may be more sensitive to variations in cognitive load than performance measures which only reveal differences if an actual cognitive <u>over</u>load occurred.

General discussion

The goal of the studies reported in this chapter was to relate viewing behavior to cognitive load theory and measurement. I explored how viewing behavior is influenced by different possible sources of cognitive load. Concerning the modality of text presentation, the fixation patterns in both experiments replicated the findings of Chapter 2 and are in good accordance with other eye tracking studies on concurrent presentation of written text and pictorial information (Carroll & Young, 1992; d'Ydewalle & Gielen, 1992; Faraday & Sutcliffe, 1996; Hegarty, 1992a; Rayner et al., 2001; Underwood et al., 2004). Whenever pictures are combined with written text much visual attention is paid to reading text. In both experiments written text was attended first and captured relatively more visual attention than illustrations. The general pattern is that participants started reading text as soon as a new “scene” appeared and then successively alternated between text and illustrations several times. Alternations were a function of the pace of presentation, i.e. their number was positively related to presentation duration.

This observation led to an interpretation of the modality effect in terms of visual attention allocation (cf. Chapter 2). Presenting text in written rather than spoken form along with an illustration leads to poorer learning because parts of the (illustrated) information are possibly missed or only superficially processed. This interpretation is supported by the current observations. Presenting written rather than spoken text led to a higher cognitive load indicated by self-rating and performance measures. In both experiments participants rated their cognitive load as higher whenever expository text was written rather than spoken thus delivering further evidence for the validity of the modality effect. In Experiment 2 participants also showed a modality effect in a subsequent visual memory task. Asked in more detail, the need to split visual attention between written text and illustration was subjectively time consuming. Participants in both experiments rated the timing of instruction as less appropriate whenever text was written rather than spoken. Furthermore, participants in written text conditions were more distracted from inspecting illustrations by textual information than the spoken text groups.

In system-paced and, thus, time-limited presentations learners are forced to weigh a trade-off between attention allocation on written text and illustration. On grounds of the fixation data one might be seduced to conclude that people voluntarily attended to written text first. In fact, however, participants felt distracted from inspecting illustrations by the presence of written text. Thus, attending to written text first appears to be a rather unintended and automatic behavior that can, at least initially, not be suppressed. Nevertheless, this viewing behavior is reasonable. Expository text is a highly structured information source and people are used to gain much information from reading. Illustrations are usually not self-explaining and are often accompanied by written text. Reading text first might in the past have been experienced as being helpful in order to understand illustrated information. As a consequence, presenting written rather than spoken text in a multimedia instruction leads to an indeed reasonable but rather automatic initial reading behavior that is, compared to spoken text presentation, perceived as a time-consuming process.

Presentation time revealed to be a moderating variable for the modality effect in Experiment 2. In general, longer presentation durations led to lower ratings of extraneous cognitive load and marginally higher performance in visual memory, hence proving the pacing of instruction to be an independent source of cognitive load. A statistically significant effect of text modality on ratings of extraneous cognitive load only occurred for fast presentation pace. Descriptively, performance measures were in accordance with this interaction of text modality with pacing of instruction. Lowering the pace of presentation apparently lowered the burden of split attention between written text and illustration.

An interaction of text modality with pacing is also reflected in changes of the fixation patterns. Although initial reading appeared to be an automatic behavior that was not influenced by presentation pace, other aspects of the fixation patterns seemed to adapt to differing characteristics of the stimulus material. As hypothesized participants receiving written text and illustrations spent relatively more time inspecting illustrations the longer the presentation lasted. The unintended initial reading behavior made learners feel the risk of missing illustrated information. Longer presentation durations offered to compensate for this potential loss of information. This shift in viewing behavior confirms the distracting effect of written text in time-limited presentation as a likely source of the modality effect.

The observed patterns of viewing behavior and its contribution to cognitive load might also be understood in terms of particularities of reading. People are known to differ enormously in reading speed (cf. Just & Carpenter, 1987). This speed reflects individual abilities (e.g. Jackson & McClelland, 1975, 1979; Just & Carpenter, 1992) but is also adjusted to text characteristics (e.g. Graesser, Hoffman, & Clark, 1980). In terms of cognitive load theory individual reading speed can be described as a derivate of cognitive load. Like cognitive load reading speed also varies inter-individually. Poor readers (by definition) need more time reading and comprehending written text than good readers. Thus, individual reading speed becomes an intermitting variable in system paced instructions because it interacts with other sources of cognitive load. If the presentation pace does not meet individual needs, system paced presentation negatively influences learning. First, the faster the pace the more likely some – especially illustrated – information is missed due to the general tendency to attend to written text first. Second, the faster the pace the more especially poor readers will suffer from a loss of information. And, third, even if the pace might be sufficient for an individual reader to read all the text and attend to additional illustrations, the system-paced instruction might influence reading. The reader feels forced to adjust her reading speed to a not self-controlled pacing instead of adjusting her individual reading speed to the complexity of the content. Thus, the modality effect might at least partly be caused by a mismatch of system-paced instruction with self-paced reading. Further research is necessary to reveal how viewing behavior, cognitive load, and learning success may change if learners are under control of the pace of presentation.

Taken together the results deliver converging evidence that a multimedia instruction with written text presentation is perceived as more cognitively demanding than spoken text presentation. Most evidently, the cognitive demands of single mode presentation (illustration plus written text) are attributed to a distraction caused by written text and a less appropriate pacing of instruction. These results underscore that especially the processing of illustrated information suffers from the attentional split as indicated by the eye tracking data. The risk of missing (illustrated) information can easily be compensated by longer presentation duration. No modality effect occurred for longer presentation durations. As a consequence, the general recommendation to present expository text in a multimedia instruction in spoken rather than written form needs to be modified. It seems to hold only under time-limited conditions. If learners have enough time to attend all information sources for a sufficiently long period, presenting text in written form is not inferior to spoken text presentation in terms of its cognitive load. How much time is enough depends on characteristics of the learner and the material. Further research will help specifying the constraints under which written text in multimedia instructions can be as effective as spoken text.

Chapter 4

Control and cognitive load: The influence of minimal user interaction on the modality effect in multimedia learning

This chapter examines how the modality effect and visual attention allocation are affected by learner control. 31 participants watched a 16-step multimedia instruction on the formation of lightning on a computer screen. Text was presented concurrently to the illustrations either visually or aurally. Learners controlled the pace of presentation by pressing the spacebar to continue with the next step of instruction. Their eye movements were recorded during learning. Learning outcome was measured by retention, transfer, and visual memory tests. In addition, participants rated their cognitive load. Learning outcomes, self-ratings of cognitive load, and average presentation durations did not differ significantly between the text conditions. Adjusting the speed of presentation, learners were able to keep cognitive load within bounds and to gain optimal learning results independent from the format of text presentation. Eye movement patterns in written text presentation revealed that individual pacing was triggered by reading speed. The fixation durations on illustrations and number of alternations did not vary systematically with pace. In contrast, when the pace of presentation was system-controlled (as done in Experiment 2 of Chapter 3) also fixation times on illustrations and number of alternations increased for longer presentation durations. These eye movement patterns revealed that visual attention allocation was highly influenced by the matter of control (learner vs. system). The huge variance of individually chosen presentation durations suggests that individual factors like reading speed or text comprehension were much more important for an optimal pacing than the modality of text presentation. As a practical consequence, the design of multimedia instructions should allow at last minimal learner control to adjust the pace of presentation to individual needs.

Introduction

Unlike listening, reading is an inherently self-paced action. Occasionally, we may ask somebody to speak more slowly. But usually we are committed to a speakers' rate. In contrast, reading written text allows an individually chosen rate. Thus, for written text presentation in a computer-controlled multimedia instruction there is an obvious mismatch of system-paced presentation with self-paced reading. This mismatch might at least partly be responsible for the superiority of spoken over written text presentation in multimedia learning, the well-established *modality effect* (e.g. Mayer, 2001; Sweller, et al., 1998). The issue of the present study is to investigate particularities of the reading task in multimedia instructions, asking if and how the modality effect becomes manifest in a learner-controlled presentation format: (1) Are students able to adjust the pace of presentation to their individual needs, i.e. do they experience a comparable amount of cognitive load and do they reach a comparable level of learning performance when they are in control of the pace of presentation? (2) Are different cognitive demands of spoken compared to written text presentation reflected in individually chosen paces? (3) And how does the viewing behavior change from system- to learner-controlled pacing of instruction?

Individual reading speed and pacing of instruction

Experiment 2 in Chapter 3 revealed interactions of the pace of presentation with text modality. Presenting written rather than spoken text led to higher cognitive load the faster the pace of presentation was. Furthermore, participants felt distracted from inspecting illustrations by the presence of written text. Eye tracking revealed that learners started reading the text as soon as it appeared on the screen and then successively alternated between text and illustrations several times depending on the pace. Distracting effects of initial reading were compensated by spending relatively more of the additional time in longer presentation durations on the inspection of illustrations. I argued in the former chapters that faster pacing bears the risk for learners to miss important, especially illustrated, information. Furthermore, fast pacing might also have had an impact on reading comprehension. Thus, the modality effect presumably only exhibits under additional constraints. It might be bypassed when other sources of cognitive load like time pressure or pacing of instruction (Paas, et al., 2003) are kept within bounds.

In studies varying the text modality in the learning material, the presentation duration is usually fixed in order to experimentally control the time on task. As a side-effect learners may experience time pressure depending on how the pacing of instruction is gained. The upper speed limit is logically determined by the rate of the speaker. The pace of presentation is often adjusted to the spoken text conditions without explicitly specifying the rate, e.g. in terms of words per minute. Estimating the pace applied in studies supporting the modality effect reveals that pace of presentation largely varies among these studies. The average paces range from 60 words per minute (e.g. Tindall-Ford, et al., 1997, Experiment 2) to a more than doubled rate of 123 words per minute (e.g. Mayer & Moreno, 1998; Moreno & Mayer, 1999).

How can we determine critical time constraints for a modality effect to occur? Experiment 2 of Chapter 3 varied the rate in three steps from 120 to 90 to 67.5 words per minute. Only in the fast pace condition (120 words per minute) participants reported a higher cognitive load for written compared to spoken text presentation. Most obviously, if this pace was too fast to engage in a *normal* reading behavior the modality effect can be explained by a disturbed text comprehension. Reading on average takes place with a rate of about 240 words per minute (cf. Just & Carpenter, 1987). Given this average, even a rate of 120 words per minute appears slow enough to allow for reading written text and still leaving time to inspect accompanying illustrations.

However, we should not rule out reading speed as a potential source for the modality effect without further consideration. Normal readers are known to adjust their reading speed to several task demands. Texts containing low frequent words or explaining a complex matter are read more slowly than texts made up by high frequent words or containing easy-to-understand statements (cf. Just & Carpenter, 1987). Besides these text characteristics, reading speed also varies with instruction. In a study by (Hartley, Stojack, Mushaney, Annon, & Lee, 1994) text was presented in two self-paced reading conditions. In one condition participants were instructed to read an expository text and to rate it for interest and familiarity of material. In this reading task participants exhibited an average reading speed of 239 words per minute, replicating the above referenced estimate. In a second condition participants were instructed to read an expository text in order to recall it afterwards. Under these circumstances reading speed dropped to an average rate of 90 words per minute. Thus, in order to allow an average learner to read and understand an expository text in a multimedia instruction the system-paced timing should not exceed a rate of 90 words per minute.

This recommendation fits to the findings in Chapter 3. However, even with rates lower than this modality effects are still observable (e.g. Tindall-Ford et al., 1997; see also Chapter 2). One possible explanation is that reading speed varies enormously among individuals (cf. Just & Carpenter, 1987). A timing oriented on an average reading speed will only fit for average or above average readers. That is, about half the learners will suffer from a pace inappropriate for their individual reading speed.

This problem of system-paced presentation can only be overcome by individually adjusting the pace of presentation. Doing so, Hartley et al. (1994) found that there is more to reading than an optimal fit to individual reading speed. The authors estimated each participant's reading speed from the above-mentioned self-paced reading task without recall. Then they presented to-be-recalled expository texts sentence by sentence either in system-paced or self-paced conditions. System-paced conditions were set at 0.5, 1.0 and 2.0 times the individual reading speed of the no-recall task. In the self-paced condition participants pressed a key to move from sentence to sentence. Overall, recall performance was a (logarithmic) function of the absolute time spent reading (accounting for 94% of mean logarithmic recall performance). That is, longer time on task led to a higher recall performance. There was, however, a remarkable difference between individually determined and self-paced reading conditions. Individually determined reading speed did not correlate with recall performance, i.e. if system-controlled presentation was adjusted to individual reading speed, learners exhibited an equal recall performance no matter how much time they were given. In contrast, in the self-paced reading condition the time spent reading was *positively* correlated with recall (*r*=.37). Thus, individually different reading speeds in the self-paced reading task did not only compensate for different reading abilities but also reflected the contribution of deliberate, probably effortful, strategies for remembering expository text. Even an individually fitted system-controlled pace of presentation might hinder the learner to apply such further reading strategies.

In order to examine how these particularities of reading behavior contribute to multimedia learning, the present study asks if and how the modality effect is affected by self-paced instruction. Before turning to the empirical investigations I will shortly outline what effects of self-paced instruction can be expected on measures of cognitive load, learning performance and viewing behavior.

Learner control and cognitive load

Multimedia instructions can be designed to allow learners to choose the pace of progressing from one portion of information to another. One might suspect that navigation in multimedia instructions introduces an additional source of cognitive load. However, learners can also be expected to benefit from control options. In a study by Mayer and Chandler (2001) learners received an instruction composed of illustrations and spoken expository text. A "minimal" control option was realized by asking participants to progress in the material by clicking a "next"-button. Learners rated the self-paced presentations of the material as less cognitively demanding than the otherwise identical system-controlled presentations. Furthermore, in a self-paced presentation condition learners achieved a higher score in problem-solving transfer compared to their system-paced counterparts. Thus, controlling the pace of presentation led to a lower cognitive load.

Since the content of the material was identical in both conditions, the lower cognitive load must have been caused by a longer learning time in self-paced compared to system-paced presentation, by particularities of the matter of control, or both. It is reasonable to assume that lower ratings of cognitive load and the higher transfer scores in Mayer and Chandler (2001) were related to longer presentation durations. Experiment 2 in Chapter 3 revealed that students learned more with a multimedia instruction and experienced less cognitive load the slower the pace of presentation was. Furthermore, they rated faster paces as less appropriate and wished to stop the presentation at certain points. Taking into account that the system-controlled pace of approximately 120 words per minute in Mayer and Chandler (2001) equals the fastest pace in Chapter 3, learners in their study can be expected to have reduced their cognitive load by taking more time. Unfortunately, Mayer and Chandler (2001) did not report if the average pace of learner-controlled presentation deviated from the system-controlled pace.

As for the matter of reading, the learner-controlled pace of a multimedia instruction can also vary among participants. System-paced presentation even with spoken text must not be considered to fit each individual's needs. While listening to the expository text learners attend to illustrated information (see Chapters 2 and 3) and have to integrate textual and illustrated information into a coherent structure. Similar to reading speed, the speed of information processing in general can be seen as a correlate of individual cognitive load (Fink & Neubauer, 2001). There are probably individual differences in (spoken) text comprehension that might be compensated by longer pauses between sentences or paragraphs, differences in the speed of encoding pictorial information, and differences in the effort taken to integrate information and to apply recall strategies. Since Mayer and Chandler (2001) did not report individually chosen presentation durations these hypotheses are still up to be confirmed.

Although the time on task is recognized as an intermitting variable in learning, it has almost been neglected as a measure of cognitive load so far (cf. Paas et al., 2003). Introducing learner-control in a multimedia instruction with either spoken or written expository text can be expected to evoke differences in individual pacing for both text modalities. These differences in the time chosen to inspect the instruction can be interpreted as a direct measure of individual cognitive load. According to cognitive load theory written text causes a higher cognitive load than spoken text presentation due to limitations of the visual working memory (e.g. Sweller, 1999; Sweller et al., 1998). This difference in cognitive load caused by text modality should be reflected in the average pace of presentation.

Viewing behavior in system- vs. learner-controlled presentation

The studies reported in Chapter 3 established eye tracking to enhance the measurement of cognitive load in the visual processing system. Objecting a learner's viewing behavior allows to estimating the relative amount of resources devoted to different visual information sources. Thus, for written text presentation in a learner-controlled instruction, looking at the time spent reading and the time spent inspecting illustrations can reveal by which of these information sources an individually adjusted pace of presentation is triggered.

Learners almost automatically engage in reading as soon as written text appears (on a computer screen). This behavior was consistently found in eye-tracking studies on combined presentation of text and pictures (Carroll, et al., 1992; d'Ydewalle & Gielen, 1992; Faraday & Sutcliffe, 1996; Hegarty, 1992; Hegarty & Just, 1993; Rayner, et al., 2001; Underwood, et al., 2004; see also Chapters 2 and 3). Thus, learners can be expected to start exploring a multimedia instruction by reading text before turning to illustrated information, no matter if the system or the learner controls the presentation. In system-controlled presentation learners are forced to adjust their reading speed to the pace of presentation. In contrast, viewing times for text and illustrations in a learner-controlled presentation are not bounded above by the system and need not be traded off. Viewing times for the text will directly reflect individual differences in reading speed and the appliance of further reading strategies (Hartley et al., 1994). Thus, the individual pace of presentation can be expected to depend on the time spent reading.

The time spent inspecting illustrations might also correlate with individual pace. This prediction, however, assumes individual differences in encoding illustrations similar to individual reading behavior. To my knowledge, this aspect of comprehension of illustrations has not yet been investigated. Longer presentation durations in system-controlled presentation offered to compensate for a potential loss of (illustrated) information (cf. Chapter 3). In fact, learners took disproportionately more time inspecting illustrations the longer the presentation lasted. In a learner-controlled presentation the risk to miss parts of the presentation is bypassed. Inspecting illustrations is neither bounded by pace of presentation nor by individual reading speed. Thus, an increase in the time spent inspecting illustrations for longer presentation durations in learner-controlled instructions can be expected to be lower than in system-paced instructions. Before comparing viewing behavior in system- vs. learner-controlled presentation, the influence of learner-control on the modality effect in multimedia learning has to be investigated.

Learner paced instruction

The purpose of this experiment is to examine how the learner-controlled pacing of instruction interacts with the modality of text presentation in multimedia learning. Learner control is introduced by allowing learners to watch each scene of a multimedia instruction as long as they want. The pace is controlled via pressing the space bar to progress to the next step of instruction. The instructional material used in this experiment is identical to the material used in the prior studies. In these studies illustrations depicting the formation of a lightning storm were presented concurrently with expository text that was either spoken or written.

For the written text presentation I expect that learners will adjust the pace of presentation to their individual reading speed and the perceived demands of the material. Similarly, learners exposed to spoken text are expected to adjust the pace of presentation to their individual needs. Note, that the lower bound for the resulting presentation duration in spoken text presentation is determined by the rate of the speaker. As a consequence of individual adjustment I expect no differences in self-ratings of cognitive load and in subsequent learning outcomes. A higher cognitive efficiency of spoken over written text presentation should be reflected in shorter presentation durations.

Viewing behavior is expected to replicate the findings of the former studies. Learners in the written text condition will spend relatively less time on inspecting illustrations than learners in the spoken text condition since they will spend a reasonable amount of time reading. The time spent reading and the time spent inspecting illustrations will reveal by which of the information sources the individually chosen presentation duration is triggered. Apart from these individual differences in the speed of information processing I expect learners to show a comparable viewing behavior. All learners in written text conditions will engage in reading with the start of a new scene and alternate between written text and illustrations equally often.

Method

Participants and Design

31 students of the Justus-Liebig University Giessen participated in the experiment in partial fulfillment of a course requirement. All participants were native German speakers with normal or corrected to normal vision. Participants were randomly assigned to one of two experimental groups. 16 participants served in the spoken text group, 15 participants in the written text group. The groups did not differ in prior knowledge. Mean values vary around 2.5 for self-estimated prior knowledge (on a 5-point scale from very little to very much) and between 4.5 and 5 checked items on a checklist consisting of 11 domain-related items.

Materials and apparatus

The learning material consisted of a 16-step multimedia instruction on the formation of lightning programmed in Flash 4.0 (Macromedia, 1999). The instruction showed a sequence of 16 animated illustrations depicting the motion of cool air that becomes heated; heated air rising up and forming a cloud; the rising of the cloud beyond the freezing level; drops of water and ice crystals moving up and down within the cloud, colliding, and causing electrical charges to arise; heavy drops and crystals falling down and producing downdrafts; a stepped leader of negative charges moving down to high objects on the ground; and positive charges moving up to the cloud producing a flash light. Illustrations were accompanied by an expository text describing each of the major events. Text was spoken, written inside the illustration frame or written below the illustration frame (*Figure 1*1). The whole text had a length of 281 words, varying between 9 and 26 words per scene. For the spoken text condition, text was spoken in a male voice at a net-rate, i.e. without pauses between paragraphs, of approximately 163 words per minute. The instruction was started by the participants pressing the space bar. Each of the 16 scenes lasted until participants hit the space bar again to start the next scene. During the learning period, the stimulus computer recorded the resulting scene lengths. The net-rate of spoken text defines the theoretical upper limit of pace of presentation.

Figure 11. A selected frame and corresponding written text for multimedia explanation on lightning formation

The instruction was presented by a PC on a 21'' color monitor, situated approximately 80 cm from the participant. Spoken text was presented by an audio system. Fixations were monitored by an Applied Science Laboratories' corneal-reflectance and pupil-center eye tracker (ASL 504). The fixation position on the screen was measured with a sampling rate of 50 Hz and output to a PC, which controlled the recording, the camera, and the calibration (ASL Eyepos, E5000). Two additional monitors displayed the participants' tracked eye and its current fixation position on the stimulus screen to the experimenter.

Prior knowledge, performance, and cognitive load measures were assessed by paper-and-pencil tests. The material consisted of a participant questionnaire, a retention test, a four-item transfer test, a visual memory test, and two rating sheets for the cognitive load. The participant questionnaire asked for the participant's gender, age, profession and experience with meteorology. The retention test asked the participants to write down an explanation of how lightning works until told to stop. The transfer test consisted of the following four questions, each typed on a separate sheet: (1) "Which physical conditions must be given in order to decrease the intensity of a lightning storm?", (2) "Why do you often see clouds in the sky but no lightning?", (3) "What does air temperature have to do with lightning?", and (4) "What are the physical causes of lightning?". The visual memory test contained instructions to sketch (1) a cloud with a sufficient condition for electrical charges to arise, (2) how electric charges arise in a thundercloud, (3) the distribution of charges within a thundercloud before the stepped leader builds up, and (4) a stepped leader as it arises before a lightning. Answers were asked to be given on four sheets containing a simplified background scene of the original learning material.

The first of two rating sheets on cognitive load, given directly after presenting the multimedia instruction, contained the following three questions: (1) "How easy or difficult was it for you to learn something about lightning from the presentation you just saw?", (2) "How easy or difficult would you consider the *content*?", (3) "How pleasant or bothersome would you consider the *presentation format*?". Participants were instructed to place a check mark for each question on a 7-point rating scale from very easy (pleasant), easy (pleasant), rather easy (pleasant), medium, rather difficult (bothersome), difficult (bothersome), to very difficult (bothersome). Question one is a standard item for subjective ratings of cognitive load (e.g. Kalyuga, et al., 1999; Paas & Van Merrienboer, 1993, 1994). Questions two and three are introduced to differentiate between intrinsic cognitive load, i.e. due to an interaction between learner and content, and extraneous cognitive load, i.e. due to the presentation format (e.g. Swaak & de Jong, 2001).

The second rating sheet, given after completion of the performance tests, contained 9 detailed statements on the presentation: (1) "I would have preferred to stop the presentation myself at certain points", (2) "I would have preferred to look at some illustrations again", (3) "I would have preferred to rewind and repeat parts of the text", (4) "I missed parts of the textual information", (5) "I missed parts of the illustrations", (6) "It was difficult for me to relate textual and pictorial information to each other", (7) "The illustration distracted me from textual information", (8) "The textual information distracted me from the illustration", and (9) "How easy or difficult was it for you to control the presentation pace?". Statements 1 to 8 had to be rated on a 6-point scale from completely true, true, rather true, rather false, false to completely false. Statement 9, concerning the pace of presentation had to be answered on a 7-point scale from very easy, easy, rather easy, medium, rather difficult, difficult, to very difficult.

Procedure

Participants were tested in single sessions and were randomly assigned to one of two treatment groups. They were given general instructions explaining the procedure and introducing the topic.

Participants were instructed to acquire as much information as possible about the formation of lightning from the multimedia presentation in order to perform subsequent tasks. Next, participants were asked to fill out the questionnaire. Then, participants were seated in front of the stimulus PC and the eye-tracking system was calibrated. After that, participants were told to press the space bar whenever they feel to have studied a discrete scene for an appropriate amount of time and received three practice trials. The practice trials were implemented as a PowerPoint presentation repeating the explanation how to use the space bar in the multimedia instruction. Then, participants started the multimedia presentation by pressing the space bar. After the participants had clicked through the presentation, they rated their perceived cognitive load on the first (of two) rating sheets. Then they were given instructions to work on the retention test. Participants had 8 min to process the test. The retention test was followed by the transfer test. Instructions for the transfer test were handed out together with the first of four questions and the remaining questions were handed out successively. Participants were given 5 min to answer all questions. After the transfer test the visual memory test was given to them. Participants had 5 min to work on the sketches. The tasks were handed out successively. After completing the visual memory test, participants were given three additional minutes to write comments on their sketches in a different color. This was done in order to facilitate scoring of ambiguous sketches. Finally, the second rating sheet was handed out. After completion, participants were debriefed and thanked for their participation. The experimental session lasted about 50 min.

Results

Before analysing the dependent measures we inspected the individually chosen presentation durations for peculiarities. Pressing the space bar to start the next scene allowed an easy navigation but bore the risk of accidentally skipping single scenes. If a participant had viewed a discrete scene for less than one second (i.e. before the written text appeared or the narration started), he or she must be considered not to have seen the “same” instruction as the other participants. Within the 31 data sets 4 participants were identified (all in the spoken text condition) who skipped at least one of the 16 scenes in this manner and were excluded from further analyses. After that, participants whose chosen presentation duration (summed over scenes) was three standard deviations above or below mean presentation duration were defined as outliers. Applying this criterion one further participant (from the written text condition) who had spent more than 357 sec. inspecting the instruction was excluded from further analyses. Thus, the following analyses (if not otherwise noted) are based on n=26 participants, 12 in the spoken and 14 in the written text condition, respectively.

Performance Measures

A scorer being unaware of the participant's identity scored each participant's performance on retention, transfer and visual memory tests. For retention participants were given 1 point for each of 19 main ideas of the casual chain of lightning formation. Scores for problem solving transfer were obtained by giving 1 point for each acceptable solution with a maximum of 3 points for each question. Acceptable answers included for example "less positive charges on the ground" (question 1), "the clouds did not reach the freezing level" (question 2), "cool air becomes heated from a warmer surface" (question 3), and "the appearance of different electrical charges within the cloud" (question 4). Visual memory scores were obtained by giving 1 point for each appropriate and identifiably sketched visual element, with a maximum of 2 points obtainable for sketches 1, 3, and 4, and a maximum of 3 points for sketch 2. Acceptable answers included for example a straight line with temperature symbols indicating that the cloud extends above the freezing level (sketch 1), the collision of water and ice crystals in the cloud (sketch 2), negative charges at the bottom of the cloud (sketch 3), and a stepped leader between the cloud and a higher object from the ground (sketch 4). A second rater scored a subset of 10 participants' data independently. Inter-rater-reliabilities for these subsets vary between r=.87 and r=.96. Analyses were conducted with the scores obtained by the first rater. Mean scores and standard deviations for all three measures are shown in Table 11.

Table 11
Mean values and standard deviations of performance scores for retention, transfer, and visual memory tests.

	Text presentation format			
	Spoken text		Written text	
Test	*M*	*(SD)*	*M*	*(SD)*
Retention	13.5	(5.9)	11.0	(5.7)
Transfer	3.3	(1.3)	3.1	(1.3)
Visual memory	7.9	(3.4)	7.0	(3.1)

T-tests of performance scores between spoken and written text groups revealed no significant differences, $t(24)$=1.05 for retention, $t(24)$=.21 for transfer, and $t(24)$=.66 for visual memory. No modality effect occurred in any of the three performance measures. The lack of a modality effect in these measures indicates that participants were able to adjust the presentation pace in a way that allowed comparable learning performance between spoken and written text presentation formats.

Overall performance did not correlate with time on task. Retention (r=.14, n.s.), transfer (r=-.06, n.s.) and visual memory (r=.19, n.s.) test scores were independent from the time learners spent with the presentation. Analyses of covariance (ANCOVA) with the between-subjects factor text presentation (spoken vs. written) and with time on task as covariate confirmed that there was no modality effect buried under the variance of performance scores explained by time on task: $F(1,23)=1.14$, $MSE=38.78$ for retention, $F(1,23)=0.04$, $MSE=0.08$ for transfer, and $F(1,23)=0.51$, $MSE=5.36$ for visual memory (all ps>.10).

Subjective ratings

In the first rating sheet participants were asked to estimate their cognitive load in general, and to further distinguish between load caused by content and load caused by presentation format. Although one can argue that this differentiation is quite difficult or that learners are not sensitive to this differentiation at all, correlations between the three items vary between $r=.00$ (n.s.) and $r=.60$ ($p<.01$) indicating that participants answered the questions differently. Thus, separate t-tests were conducted for each of the three items. None of the ratings differed between written and spoken text presentation groups, $t(24)=.43$ for overall load, $t(24)=1.48$ for content and $t(24)=.46$ for presentation format. Mean scores for both experimental groups are shown in Table 12.

Table 12
Mean values and standard deviations of rating scores for cognitive load items. Higher scores indicate a higher cognitive load or higher agreement with the statement.

	Text presentation format			
	Spoken text		Written text	
Item description	*M*	*(SD)*	*M*	*(SD)*
Overall load (0-6)	1.8	(1.0)	1.9	(1.1)
Content (0-6)	2.0	(1.3)	1.4	(0.8)
Presentation format (0-6)	1.3	(0.8)	1.4	(1.2)
Stop presentation (0-5)	1.3	(1.6)	1.5	(1.1)
Review illustrations (0-5)	3.1	(1.7)	3.1	(1.3)
Repeat text (0-5)	4.2	(1.2)	3.3	(1.1)
Missed text (0-5)	2.7	(1.4)	1.6	(1.6)
Missed illustrations (0-5)	1.1	(1.0)	1.4	(1.2)
Problems connecting text and illustration (0-5)	1.2	(1.3)	1.3	(1.1)
Distracted by illustration (0-5)	1.1	(0.9)	1.3	(1.4)
Distracted by text (0-5)	0.7	(0.8)	1.1	(1.3)
Controlling the pace of instruction (0-6)	2.6	(1.1)	3.3	(1.5)

After completion of the performance tasks, participants were asked to give nine more detailed descriptions of their cognitive load by judging statements about several aspects of the presentation. A multivariate analysis of variance (MANOVA) with the between-subjects factor text presentation (spoken vs. written) and with the nine judgments as dependent measures revealed no effect of text presentation format, $F(9,16)=2.01$, Wilks-Lambda=0.47, $p>.10$. Text presentation did not significantly influence the possible problems with discrete aspects of the presentation like “missing parts of text” or “integrating textual and pictorial information”. Overall, participants reported a medium difficulty for controlling the presentation pace. Difficulties with discrete aspects were rather denied except of two items. Participants in both conditions reported that they would have preferred to look at some illustrations again (Item 2) and to rewind and repeat parts of the text (Item 3). Controlling the pace of instruction apparently induced the wish for further navigation possibilities.

Presentation duration

No effect of text modality occurred in learning performance and subjective ratings of cognitive load. The participants obviously paced the presentation in a way that fitted their individual needs. If written text presentation causes a higher cognitive load than spoken text presentation, this load was expected to be reflected in longer individually chosen presentation durations. A t-test on presentation duration between spoken and written text groups revealed no significant difference, $t(24)=.03$. In fact, mean presentation durations as well as variance, minimal and maximal durations in spoken and written text conditions are almost equal. Participants in spoken text conditions spent 183.5 sec. (SD=37.4) on average inspecting the presentation, participants in written text conditions spent an average of 183.0 sec. (SD=47.7). In spoken text conditions the presentation durations varied between 132 sec. and 271 sec., in written text conditions between 132 sec. and 285 sec. Learner paced presentation durations did not differ from spoken to written text presentation groups but varied strongly between participants.

Viewing behavior

To analyze viewing behavior the 26 cases remaining after the first exclusion procedure were processed in the following manner. Viewing positions were transformed into fixations and saccades using ASL-Eyenal software. Areas of interest (AOI) were defined to cumulate single fixations and saccades into viewing times and numbers of fixations on text and illustration. An AOI in the presentation was a part in which either a portion of text or an illustration was displayed. *Figure 12* shows an area of written text and an area of illustration for one scene of the presentation. In order to detect inaccurate calibration the resulting viewing times were further inspected in the following manner. Data sets in which viewing time on AOIs summed up to less than 75% of the total presentation time were taken as possibly invalid. Applying this criterion, 4 further participants had to be excluded. Thus, the following analyses were calculated with a set of 22 data cases.

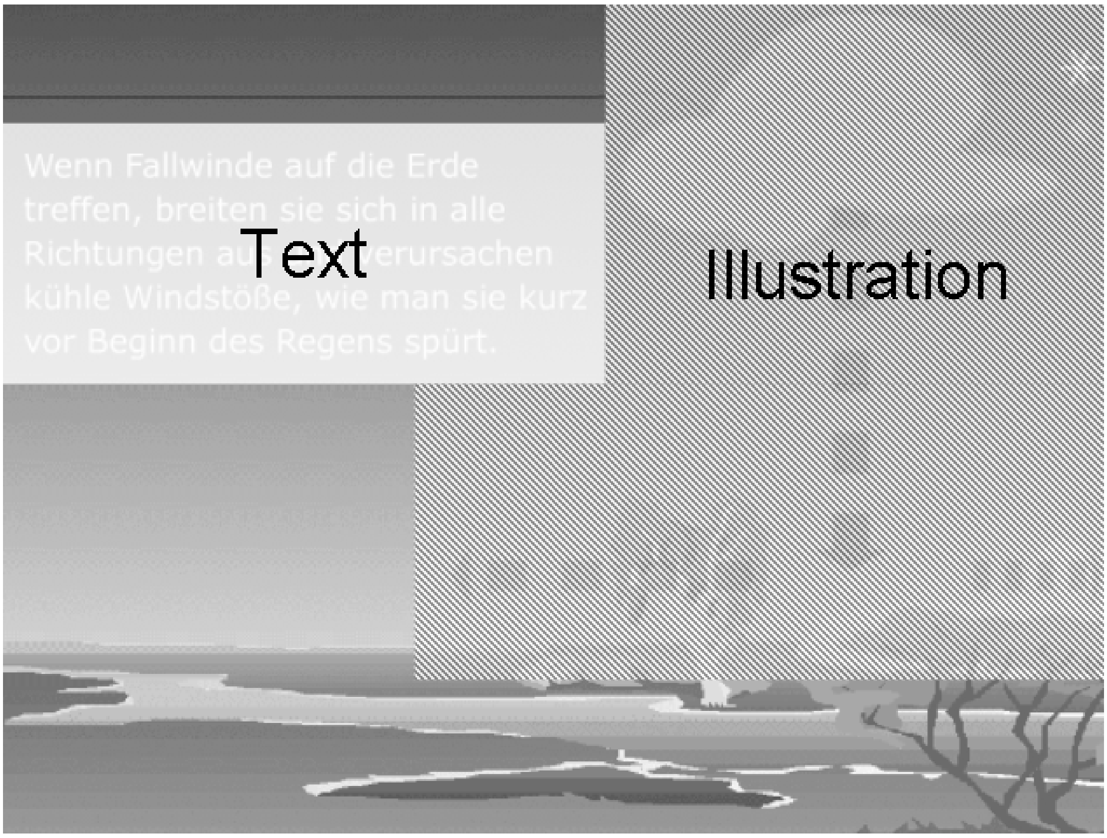

Figure 12. An example of areas of interest (AOI) for illustration (striped) and written text (white). Note that the areas vary from scene to scene depending on text length and location of the illustrations.

Overall, participants spent 92% of their fixations on AOIs. An analysis of covariance (ANCOVA) with the between-subjects factor text presentation (spoken vs. written) and with presentation duration as covariate on summed fixation times on illustration as dependent measure revealed a main effect for text presentation format, $F(1,19)=168.49$, $MSE=80{,}273$, $p<.01$, $\eta^2=.90$. Participants in the spoken text groups spent, relative to their individually chosen presentation pace, more time inspecting illustrations than participants in the written text group (see *Figure 13*). As shown in *Figure 13*, participants receiving written text split their visual attention between text and illustration. The time spent reading and the time inspecting illustrations did not significantly correlate ($r=.51$, n.s.). While the time spent reading systematically varied with the individually chosen presentation durations ($r=.97$, $p<.01$) the time spent inspecting illustrations did not significantly correlate with presentation duration ($r=.62$, n.s.). Participants in the written text condition alternated between written text and illustrations on average 3.4 times per scene. The number of alternations did not significantly correlate with the presentation duration ($r=.43$, n.s.). The right panel of *Figure 5* depicts the number of alternations for individual presentation durations. Within the first five fixations after a scene change 91% of fixations were on text. Also this viewing behavior did not vary with the presentation duration ($r=-.24$, n.s.). Assuming that participants read all the text at least once allows calculating a lower limit of the applied reading speed. This speed varied between participants from 78 to 222 words per minute with an average of 152 words per minute.

Taken together, all participants in the written text condition showed similar patterns of viewing behavior. After a scene change they started reading the text and then turned to inspect the illustrated information, reread some portion of text and then returned to the illustration again before starting the next scene. The only source of individual difference was the time spent reading which almost perfectly fitted the individually chosen pace of instruction.

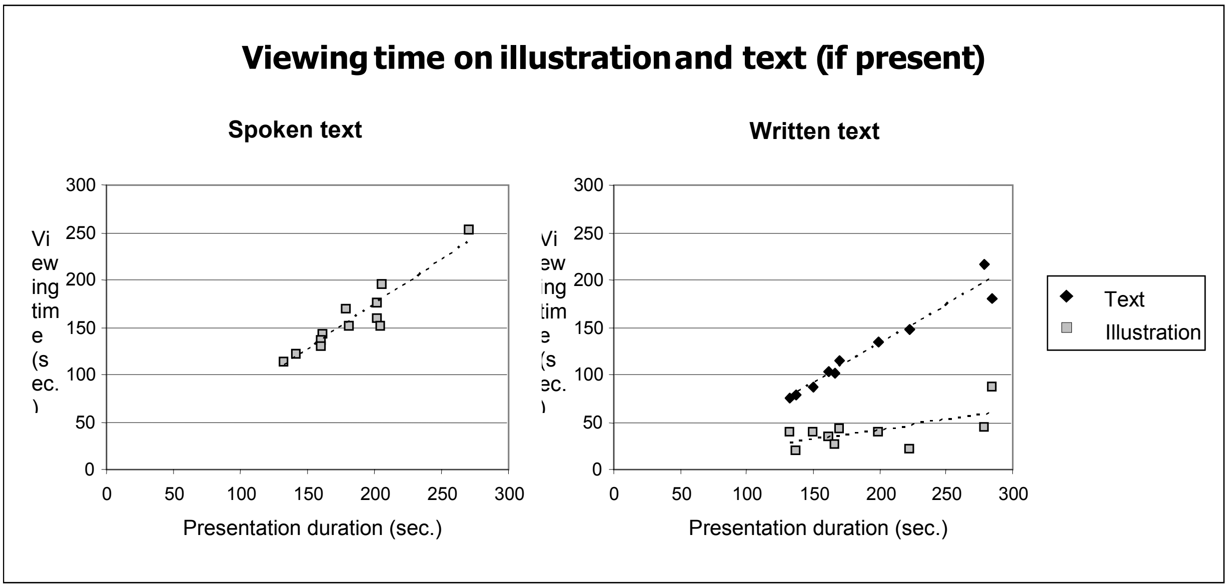

Figure 13. Individual viewing time on illustration and written text (if present) compared to individually chosen presentation durations.

Discussion

Viewing behavior replicated earlier findings showing a modality effect for the time that could be spent inspecting illustrations. Participants in written text conditions split their visual attention between text and illustrations and thus spent relatively less time inspecting illustrations than participants who received spoken text. Within the written text condition an equal amount of time was spent inspecting illustrations no matter how long the participants watched each scene.

Performance measures and subjective ratings of cognitive load did not vary with the text presentation format. As expected no modality effect occurred in these measures. Since learners controlled the pace of instruction they could adjust the speed of presentation to the assumingly different demands caused by written and spoken text presentation. Thus, a modality effect was expected to occur in the individually chosen presentation durations. Interestingly, however, the durations did *not* vary with text presentation format. But the variance between participants indicates large individual differences in the optimal pacing of multimedia instructions. One major source for these individual differences can be found in reading time. Analyzing participants' viewing behavior revealed that, as predicted, the time spent reading largely varied. Average reading speed was slower than in normal reading tasks but somewhat faster than expected. However, the time spent reading in the written text condition almost perfectly correlated with presentation duration while the amount of time inspecting illustrations did not. Given that individual reading or text comprehension abilities are pre-experimentally set, the chosen pace of instruction in the written text condition was determined by reading speed.

Comparing written and spoken text conditions revealed that not only the mean of chosen durations was equal, but also the distribution of durations in terms of standard deviation and range of duration values were (almost) equal in both text presentation conditions. Thus, generalizing the interpretation of optimal pacing of instruction for both groups it appears that learners choose a presentation pace that fits to their *individual speed of text comprehension*. The role of control and individual text comprehension factors may be understood better if we take a closer look at viewing behavior in learner-paced compared to system-paced instructions.

Comparing viewing behavior in system- and learner-paced instruction

The second experiment in Chapter 3 (henceforth Experiment 1) varied text presentation (spoken, written) and pace of presentation in three steps while using the same multimedia instruction as in the current study. The variation of system paced presentation durations (140, 187, and 249 sec., respectively) roughly fits in with the range of individually chosen presentation durations (132 to 285 sec) in the current experiment with learner-paced presentation (henceforth Experiment 2). It is also notable that the average presentation duration in Experiment 2 (183 sec.) is very close to the medium presentation duration of 187 sec. in Experiment 1. Due to these similarities on the time dimension, differences in the observed viewing behavior of both experiments can be devoted to the issue of learner control.

In both experiments the way visual attention was allocated to reading text and inspecting illustration varied with the presentation pace. *Figure 14* and *Figure 15* depict the ratios of the time spent reading to the time spent inspecting illustrations, and the number of alternations between text and illustrations, respectively. In each figure the left panel depicts eye movement measures for system-paced instruction, the right panel those for learner-paced instruction.

As shown in the left panel of *Figure 14* the ratio of time spent reading to time spent inspecting illustration decreased with increased (system-controlled) presentation duration (*r*=-.42, *p*<.01). Learners spent relatively more time inspecting illustrations than reading text the longer the presentation lasted. In contrast, when the pace of presentation was learner controlled this ratio was not significantly and, if anything, positively correlated with presentation duration (*r*=.32, n.s.). The slight increase in this ratio was caused by the individual reading speed. Individually chosen presentation duration in learner paced instruction only varied with the time spent reading while all participants spent a comparable amount of time inspecting illustrations (see also right panel of *Figure 13*).

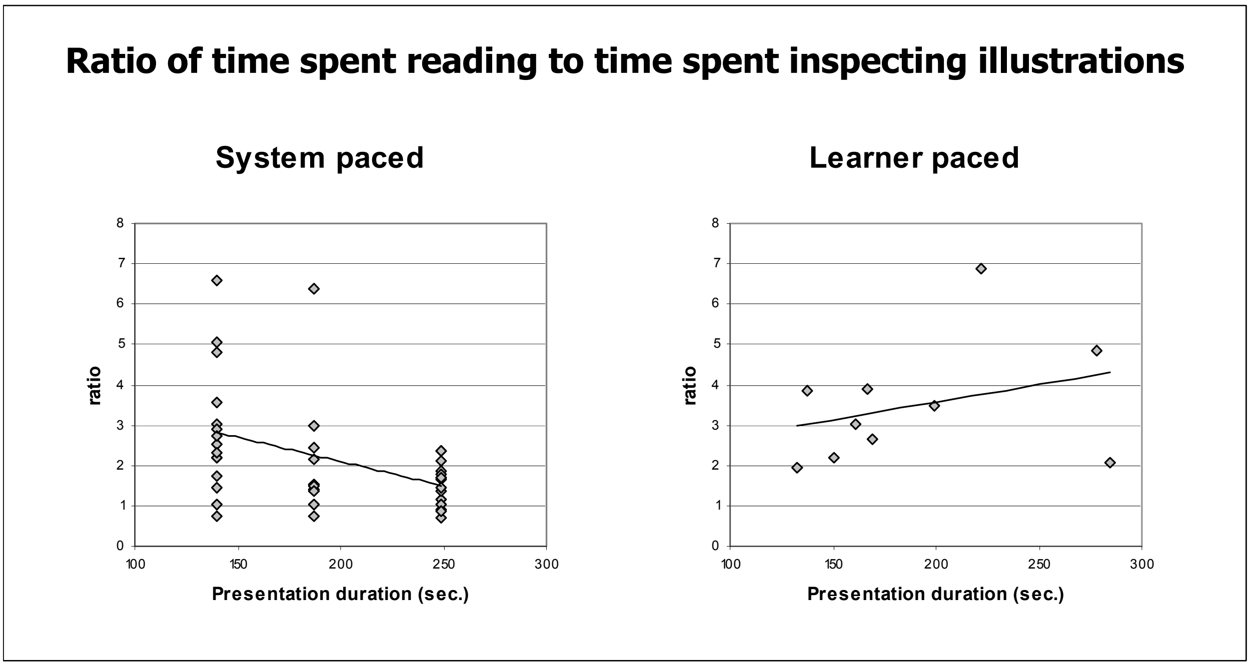

Figure 14. Ratio of time spent reading to time spent inspecting illustrations for written text conditions in system (left panel) vs. learner (right panel) paced presentation.

Also comparing the number of alternations revealed dissimilarities between system- and learner-paced instructions. For system-paced instruction the number of alternations was positively related to presentation duration (*r*=.70, *p*<.001). The longer the presentation lasted, the more often participants looked back and forth between text and illustrations (see left panel of *Figure 15*). In learner-paced instruction this correlation failed statistical significance (*r*=.43, n.s.). The variance of number of alternations was much smaller in learner- compared to system-paced instruction. Thus, the viewing behavior of learners controlling the pace of instruction was more stable over the presentation time than in system-controlled conditions.

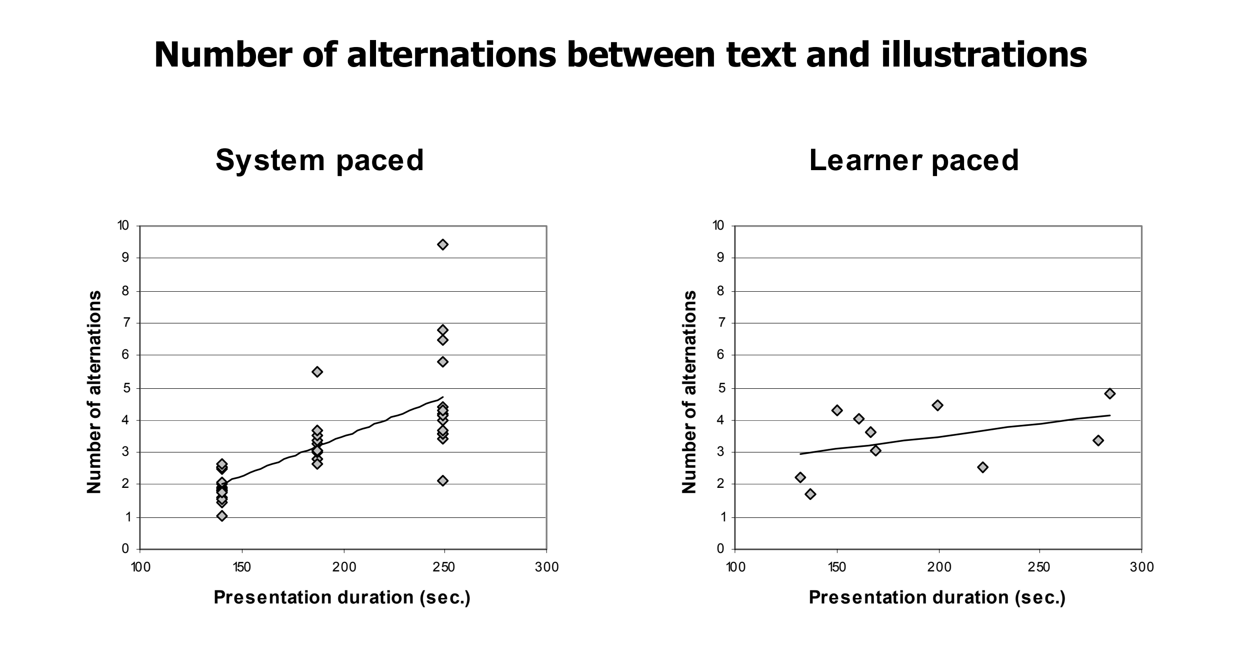

Figure 15. Number of alternations between text and illustrations for written text conditions in system (left panel) vs. learner (right panel) paced presentation.

Taken together, comparing viewing behavior in system- vs. learner-paced instructions revealed remarkable differences. Participants in system-paced instruction used additional presentation time in favor of illustrations while participants in learner-paced instruction used additional presentation time exclusively for reading. Furthermore, participants in system-paced instruction used additional presentation time to alternate their visual attention between text and illustration more often while participants in learner-paced instruction did not. The longer the system-paced presentation duration in Experiment 1 was, the more participants can be assumed to have read (and possibly understood) the written text. They really had additional time to spend on inspecting illustrations and “to look about”. In learner-paced instruction the split of visual attention between text and illustrations appears rather systematic and is best comparable to the viewing behavior shown by the medium system-paced presentation group in Experiment 1. The general strategy was to read (some portion or all of the) text, then switching to inspect the illustration, re-reading some portion of text and going back to the illustration once more. The time needed to process illustrated information appeared rather constant for all participants. The only inter-individual difference was the time spent reading that almost perfectly correlated with the chosen presentation duration. Thus, an optimal fit of presentation pace to the attentional demands in concurrent presentation of text and illustration is driven by the individual reading speed. Giving *all* learners more time might decrease attentional demands. However, some learners were able to follow their “natural” strategy even in medium and fast system paced presentations.

General discussion

The present study revealed that the modality effect in multimedia instructions can be bypassed by giving learners control over the pace of presentation. Participants reached a comparable level of learning performance no matter if the text was written or spoken. Thus, in the absence of time constraints, the modality of text presentation did not influence learning success. Since learning is only impaired if an actual *over*load occurs learning performance does not always identify different levels of cognitive load between the different experimental conditions. However, asking learners to estimate their cognitive load revealed no difference of text modality in self-paced instruction. Controlling the pace of presentation learners also *perceived* a comparable amount of cognitive load.

The levels of learning performance and self-ratings may have been reached by adjusting the pace of presentation to different demands caused by text presentation format. According to cognitive load theory the higher cognitive load of written compared to spoken text presentation can be compensated. If the modality effect in system-paced multimedia learning is due to a cognitive overload in visual working memory, learners in a self-paced multimedia instruction with written text will choose a slower pace (i.e. they spend more time on task) than learners receiving spoken text, in order to keep the load of the visual working memory within bounds. Consequently, differences in the cognitive efficiency between spoken and written text presentation were expected to become evident in learner-controlled paces. Contradictory to this prediction, however, average pace did *not* vary with text presentation format.

How can a lack of modality effects – even in measuring time on task – be explained? The huge variance of resulting presentation durations in both text modalities indicated large inter-individual differences. Overall, time on task did not correlate with learning performance. Learners adjusted the pace of presentation in order to avoid cognitive overload, resulting in comparable levels of subjective load and learning performance. For written text presentation the objected viewing behavior revealed that longer presentation durations almost perfectly correlated with the time spent reading while the time spent inspecting illustrations remained rather constant. Thus, within written text condition individual differences in pacing were mainly triggered by reading speed. Taking the time on task, text, and illustrations as (direct) measures of cognitive load the main source for overall cognitive load (as indicated by time on task) in written text conditions was the expository text (as indicated by the time spent reading). The expository text was read with a mean rate of 150 words per minute. This rate is slower than the often-referenced frequency of 240 words per minute for average readers. The deviation from "normal" reading speed indicates that besides individual differences all learners adjusted their reading behavior to task characteristics. In fact, participants were asked to retrieve information from the learning material in order to perform subsequent tasks. Under these conditions a reading speed below the standard level was expected and indicated deliberate, probably effortful, strategies for remembering expository text (cf. Hartley et al., 1994). Taken together, participants receiving a multimedia instruction with written expository text apparently adjusted the pace of presentation primarily in order to ensure a sufficient text comprehension.

The ability to comprehend textual information can be assumed to be independent from presentation modality (Guthrie, 1973; Jackson & McClelland, 1979; Rost & Hartmann, 1992; Sticht & James, 1984). Hence, also the resulting durations in spoken text presentation reflect comprehension strategies and individual differences in text comprehension abilities. This conclusion is supported by the striking similarities in the distribution of presentation durations in spoken and written text presentation. Not only average duration but also variance and range of durations were almost equal in both text presentation conditions. Thus, the most important moderating variable of cognitive load in self-paced instruction is not the modality of text presentation but text comprehension.

Understanding text is well recognized as a matter of managing working memory load (Graesser & Britton, 1996; Just & Carpenter, 1992). But the cognitive load in multimedia learning must not be reduced to factors of text comprehension. Retrieving information from the material used in the present study did not only require careful reading or listening but also inspecting accompanying illustrations and connecting the verbal and the pictorial information. In fact, all learners spent a considerable amount of time inspecting illustrations. If text was written, participants alternated between text and illustrations several times per scene indicating that they took effort to connect textual and pictorial information. In contrast to their reading speed, however, the actual time learners spent viewing the illustrations in written text condition as well as the number of alternations did not significantly correlate with their total learning time. Thus, unlike text comprehension the cognitive load in terms of time on task induced by illustrations and their referential connections to text did not significantly differ between individual learners. This conclusion is further supported by the low ratings of the respective items in the questionnaire on specific aspects of the multimedia instruction. One can easily imagine that other pictorial information (e.g. statistical graphs) can evoke large inter-individual differences in the same fashion as text comprehension. One direction of further research is, how different levels of complexity of pictorial and/or verbal information interact with text modality and the matter of control.

In the context of this thesis the present study revealed that the modality effect changes with the matter of control. A higher cognitive load of written compared to spoken text presentation in multimedia learning could not be found in a learner-controlled instruction. Thus, the modality effect appears to be restricted to system-controlled instructions (cf. Chapter 2 & 3). Given this conclusion, some substantial shift must have taken place from system- to learner-controlled instruction. Comparing the viewing behavior within written text conditions of system- and learner-controlled instructions revealed such a substantial shift. Apart from the time spent reading the text, learners in a self-paced instruction showed a highly stable fixation pattern. They adjusted the pace of presentation to their individual reading speed and engaged in an otherwise systematic viewing behavior. This pattern underscores the self-paced nature of normal reading. In contrast, learners in system-controlled instructions showed a different viewing behavior. Varying the pace of system-controlled instructions revealed that learners used additional time in favor of illustration over written text and to shift visual attention between text and illustrations more often. This variation of viewing behavior with pace can be explained by a mismatch between pacing and individual needs. None of the different paces in system-controlled presentation conditions has met the needs of all learners. In fast presentation conditions some poor readers surely had problems keeping up with the pace and, thus, had not much time inspecting illustrations. Some skilled readers in slow presentation conditions on the other hand can be expected to have had dispensable time to look (more or less unintentionally) back and forth between text (which they already read) and illustrations (which at least moved). Thus, receiving written text not all learners are doomed to suffer from a fast pace and not all learners need to gain further benefit from a slow pace. Further research is necessary to sharpen the role of individual reading speed and text comprehension abilities in multimedia learning.

In self-paced learning scenarios, apparently all learners gain benefit from minimal control options (Hartley et al., 1994; Mayer & Chandler, 2001). In other words, also in spoken text presentation learners can suffer from a bad system-controlled timing (Mayer & Chandler, 2001). But since spoken text presentation is superior to written text presentation in system-paced multimedia instructions we need to assume that learners receiving written text suffer more easily from a fast pace than learners receiving spoken text. A generally higher cognitive efficiency for spoken over written text presentation was called into question by the total lack of evidence for a modality effect in the present study on self-paced instruction. Differences in cognitive load are presumably not only *compensated* by learner-controlled instruction, but they only *appear* in system-controlled instruction as a matter of particularities of the reading and the listening task. In fact, people are used to adapt to different paces of speech in learning occasions like classroom teaching, educational television, and the like. In contrast, we seldom experience the need to follow a non-self-paced written text presentation in our everyday life. A system-controlled pacing of instruction that is sub-optimal for individual (text) comprehension abilities might be easier to compensate if expository text is spoken rather than written. Stating the initial assumption that the modality effect expresses a mismatch of system-paced instruction with self-paced reading more precisely, the modality effect is rather due to particular task and learner characteristics than to a general cognitive modality principle.

One might also turn around the mismatch argument: Spoken explanations are usually not controlled by the learner and, thus, do not fit to learner-controlled instructions. While in written self-paced material learners can re-read the text as often as they wish, spoken text in the present study was nonrecurring within each scene. In fact, *both* groups (written *and* spoken) agreed that they wished to rewind the animation in order to repeat text and to review illustrations. These requests of further navigation might have been evoked by the minimal option to control the pace of instruction. However, the expressed desire underscores the learner-controlled nature of multimedia learning. Before introducing a maximum of navigation options to multimedia instructions further research is necessary to check for a possible trade-off between benefits of control options for learning and additional cognitive load due to navigation problems. Taken together, concerning the optimal instructional design to foster multimedia learning the present study leads to the following recommendation. In the absence of clear predictions on the optimal pacing of instruction and given that learners benefit from minimal control options anyway, before deciding to present expository text in spoken form, designers of multimedia instructions should implement a control option for pacing in order to assure successful learning: minimal learner control can avoid modality effects.

Chapter 5

General discussion

The main aim of this thesis was to take a closer look at visual attention allocation, cognitive load and learning outcome in learning from dynamic visualizations with accompanying verbal explanations. Five empirical studies were set up to challenge the practical scope as well as the theoretical substantiation of the so-called modality principle: Whenever visualizations are accompanied by verbal explanations, present words as spoken text rather than printed text (Mayer, 2001). The theoretical rationale for this recommendation was that visual processing can become overloaded, when words and pictures are both presented visually. Based on the existing evidences for a modality effect on cognitive load and learning outcome, the thesis raised two questions:

(1) How do learners distribute their visual attention during learning from dynamic visualizations with accompanying verbal explanations?

(2) And which properties of such multimedia instructions *moderate* effects of the modality of text presentation?

The studies examined several instructional design attributes that possibly affect the perception and comprehension of visualizations and verbal explanations. In order to gain direct and objective measures of perceptual and cognitive processes during acquisition, learning outcome measures and indices of cognitive load were complemented by the method of eye tracking. In this chapter I briefly review the empirical results, discuss theoretical and practical implications of the findings, and suggest some directions for further research.

Review of the results

The studies conducted in this thesis were based on the vast empirical literature concerning the modality effect in multimedia learning (Brünken & Leutner, 2001; Jeung, et al., 1997; Kalyuga, et al., 1999, 2000; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi, et al., 1995; Tabbers, et al., 2001; Tindall-Ford, et al., 1997). Due to the strong empirical support for a modality principle in multimedia learning I did not dare to ask for the validity of the modality effect itself. But taking visual processes into account, the empirical studies presented in the previous chapters were set up to test if the modality effect can be *moderated* by properties affecting the perceptibility of a multimedia instruction. Besides the modality of text presentation (Chapters 2, 3, and 4) the studies varied spatial properties of written text presentation (Chapter 2, Experiment 1), the design of visualizations as being animated or static (Chapter 2, Experiment 2; Chapter 3, Experiment 1), the pacing of instruction (Chapter 3, Experiment 2), and its control by the learner (Chapter 4). The instructional material applied in these studies was chosen to be comparable to a common setting of multimedia learning in which the modality effect occurs and consisted of a multimedia explanation on the formation of lightning (cf. Mayer & Moreno, 1998). Visual processes were explored by applying eye tracking as a method previously unexploited in the context of multimedia learning. Observing a learner's viewing behavior was supposed to reveal perceptual as well as cognitive processes during learning. In order to compare visual processes with instructional learning, eye tracking complemented more commonly used measures of learning outcome and cognitive load.

First of all, the results of the studies in this thesis deliver converging evidence for a modality effect. Replicating a study by Moreno and Mayer (Moreno & Mayer, 1999), Experiment 1 in Chapter 2 revealed a superiority of spoken over written text presentation in the applied learning material expressed by higher retention and transfer scores. This potential of the material to be sensitive to the modality of text presentation was confirmed by higher retention and visual memory scores for spoken text presentation in the second experiment of Chapter 2 and by higher visual memory scores in Chapter 3 (Experiment 2). However, a replication of the second experiment of Chapter 2 in Chapter 3 (Experiment 1) failed to show significant differences in learning outcome measures. The ratings on cognitive load and on particular aspects of the presentation collected in the studies of Chapter 3 and 4 also revealed effects of text modality. In both experiments of Chapter 3 learners tended to report less cognitive load when text was spoken rather than written. Asked in more detail, the presentation was perceived as being faster when text was written rather than spoken, and participants in written text conditions felt more distracted from inspecting illustrations by the verbal explanation than participants in spoken text groups. Taken together, when the design attributes of the learning material were comparable to other studies (Mayer & Moreno, 1998; Moreno & Mayer, 1999) the results are fairly in line with the learning gains obtained with spoken text presentation in earlier research (Brünken & Leutner, 2001; Kalyuga et al., 1999; Mayer & Moreno., 1998; Moreno & Mayer, 1999; Tabbers et al., 2001; Tindall-Ford et al., 1997)).

Varying additional design attributes of the instruction, however, moderated the modality effect. Spoken text presentation caused higher retention scores when visualizations were animated but failed to cause differences in the modality of text presentation when the visualizations were static (Chapter 2, Experiment 2). Retention scores in Experiment 2 of Chapter 3 were sensitive to the pacing of instruction but not to modality of text presentation. No modality effect in any learning outcome measure could be found when learners controlled the pace of presentation (Chapter 4). The partial disappearance of the modality effect was further confirmed and elaborated by the ratings on cognitive load and on particular aspects of the presentation taken in Chapters 3 and 4. In Experiment 2 of Chapter 3, higher cognitive load of written compared to spoken text presentation was restricted to a fast pacing. Again, no modality effects occurred in the ratings when learners controlled the pace of presentation in the study of Chapter 4. When learners can adapt the pacing to their individual needs I expected that a higher cognitive load of written compared to spoken text presentation would cause longer learning times. In contrast to this prediction, mean learning times as well as their range and variances for spoken and written text presentation were almost identical. What largely varied, though, were the *individual* learning times.

The “cognitive” effects of text modality, visualization format, pacing and control were mirrored by the participants’ viewing behavior. First of all, the studies consistently revealed an “attentional” or “perceptual” modality effect. Clearly, when verbal explanations were spoken, the visual attention could fully be devoted to the visualizations. In contrast, whenever written text appeared on the screen, visual attention was split between text and visualizations. At least half of the learning time was spent reading independent from the distance between text and visualization (Chapter 2, Experiment 1), the visualizations being animated or static (Chapter 2, Experiment 2; Chapter 3, Experiment 1), the pace of presentation (Chapter 2, Experiment 2), and the control of pace by the learners (Chapter 4). The salience of written text was further highlighted by the fixation patterns. In all studies the learners started with reading at least some portion of text when it occurred before they turned to inspect the accompanying visualization. This highly stable viewing behavior is in accordance with earlier research on the concurrent presentation of written text and pictorial information (Carroll, et al., 1992; d'Ydewalle & Gielen, 1992; d’Ydewalle, et al., 1991; Faraday & Sutcliffe, 1996; Hegarty & Just, 1993; Rayner, et al., 2001; Underwood, et al., 2004).

Taking a closer look, however, the overall fixation pattern in concurrent presentation of written text and visualizations was subject to changes in other design attributes of the multimedia instruction. Altering the attentional salience of the visualizations from animations to static illustrations (Chapter 2, Experiment 2; Chapter 3, Experiment 1) slightly shifted the distribution of visual attention towards written text. Varying the pace of presentation (Chapter 3, Experiment 2) affected the distribution of visual attention in that relatively more visual attention was devoted to visualizations the longer the presentation lasted. And giving the learners control over the pace of presentation (Chapter 4) lead to enormous differences in the times spent reading while the times spent viewing the visualizations remained rather constant across learners. Comparing the viewing behavior between system-paced and self-paced presentation conditions (Chapter 4) revealed remarkable differences. In system-paced presentation conditions primarily visualizations benefited from longer presentation durations and learners alternated more often between text and visualization the longer the presentation lasted. The durations chosen in self-paced presentation only varied with the time spent reading. Alternations were not affected by individual reading speed.

Two main conclusions can be drawn from these results. First, visual attention allocation in learning from visualizations with accompanying verbal explanations follows a fairly stable pattern that can be moderated by design attributes of the instruction. In general, written text drags visual attention away from inspecting illustrations. Thus, written text can be considered to compete with other visual information sources. The degree of competition is influenced by surface characteristics of the visual material (e.g. apparent motion in the visual field) and by the presence and degree of time constraints. Learners adapt to these properties of a multimedia instruction by distributing their visual attention between written text and visualizations differently. Furthermore, they are able to adjust the pace of presentation to a regular reading strategy that only varies in the time taken to read text.

Second, under less attentional competition, less time constraints, and learner control of pace, effects of text modality on visual attention, cognitive load and learning outcome change, decrease, or even disappear. The competition between written text and visualizations was stronger when visualizations were animated rather than static (Chapter 2, Experiment 2; Chapter 3, Experiment 1) and when presentation time was seriously constrained (Chapter 3, Experiment 2). Once learners are relieved from following apparent motion or from weighing trade-offs between text and visualization in time constrained presentation, the need to split visual attention loses much of its impact on learning. These differential effects on cognitive load and learning outcome are associated to particularities of the viewing behavior. In general, presenting written rather than spoken text forces the learner to read text. Consequently, less time can be spent on visualizations, which already may explain the modality effect in time constrained presentation conditions. Most evidently, however, when learners can follow a regular reading strategy by controlling the pace of presentation (Chapter 4) they do not suffer from written text presentation anymore. Thus, the need to read written text may or may not interfere with extracting information from visualizations depending on how seriously reading and viewing visualizations are disturbed by the design of a multimedia instruction. The implications of these results are discussed in the following sections.

Theoretical implications

The findings of this thesis and its interpretations have some implications for the theoretical accounts of the modality effect in multimedia learning as given by the theories of Sweller (Sweller, 1999; Sweller, et al., 1998) and Mayer (2001).

First, Sweller and Mayer explained the modality effect by a possible overload of the visual processing system whenever an expository text accompanying a visualization is written rather than spoken. Besides the fact that the theories do not agree in which kind of information is processed in the visual (Sweller) or visual/pictorial (Mayer) channel, they do not consider a necessary prerequisite for an overload to occur: that more information must be extracted than can actually be processed. Although Sweller and Mayer mention that visual attention has to be split between written text and visualizations they ignore limitations in perception due to this attentional split. As suggested by several theories on visual attention allocation outlined in Chapter 2 (e.g. Allport, 1989; Van der Heijden, 1996) the eye itself is a limiting factor for information processing. The resources of working memory may or may not be sufficient to process all information taken in by the eye. But the eye itself is surely limited in the amount of information that can be fixated and retrieved in a discrete time interval. This perceptual limitation became evident in the viewing behavior observed in the studies of this thesis. Whenever text was written rather than spoken, learners almost immediately started reading and spent at least half of their viewing time on text. Thus, the finding that spoken text leads to less cognitive load and better learning results can be explained by the fact that the amount of time that can be devoted to extract and process pictorial information is decreased when at least some time has to be devoted to written text. This interpretation is supported by the finding that especially the visual memory task proved sensitive to the modality of text presentation. From this viewpoint, spoken text presentation is more efficient due to an increased perceptibility at least of the visualization. Different subsystems in working memory may moderate the modality effect but they are not the initial locus of its appearance.

A crucial aspect of this alternative explanation for the modality effect is a *limited* presentation duration. Written text impairs learning from multimedia instructions only if the learner cannot compensate the loss of processing time for the pictorial information. The results of this thesis support this view in showing that the pace of instruction and its control by the learner are highly relevant factors for the modality effect in multimedia learning. Learners perceived multimedia instructions as “faster” when verbal explanations were written rather than spoken, a higher cognitive load only occurred for fast presentation paces, and no modality effect at all occurred when learners controlled the pace of presentation. Neither cognitive load theory nor Mayer's theory of multimedia learning can account for these results since they do not take time on task into account. This gap may be filled in cognitive load theory by defining pace as an extraneous source of cognitive load. In fact, cognitive load measures turned out to be sensitive to differences in the pacing of instructions. Thus, cognitive load theory can be extended to account for the interaction between pacing and text modality in system-paced instructions.

But the explanation falls short accounting for the results obtained with self-paced instructions. In terms of cognitive load theory the higher cognitive load caused by written text presentation is supposed to be traded for longer time on task in self-paced instructions. This hypothesis could not be confirmed. Controlling the pace of presentation learners compensated for differences between spoken and written text presentation without time costs. However, a particular influence of learner control in multimedia learning became evident in the huge inter-individual differences in time on task independent from the mode of text presentation. These differences can be described in terms of cognitive load theory: "Intrinsic cognitive load (…) is determined by an interaction between the nature of the material being learned and the expertise of the learners" (Sweller et al., 1998, p. 262). One source of cognitive load in multimedia learning can be seen in the expertise of the learner to comprehend text. In fact, text comprehension is well recognized as a matter of managing working memory load (Graesser & Britton, 1996; Just & Carpenter, 1992). This notion is supported by the fixation patterns in self-paced instruction: time on task in written text presentation only co-varied with the time spent reading. Generally, text comprehension can be assumed to be independent from presentation modality (Guthrie, 1973; Jackson & McClelland, 1979; Rost & Hartmann, 1992; Sticht & James, 1984). Thus, in self-paced instructions time on task reflects individual differences in text comprehension abilities. However, the lack of any modality effect in self-paced instructions – even in time on task – also suggests that the matter of control interacts with the comprehension of written and spoken text. This interpretation cannot be drawn from cognitive load theory since the theory considers intrinsic cognitive load (e.g. by individual text comprehension ability) and extraneous cognitive load (e.g. by text presentation mode) as independent and additive factors.

The influence of learner control on the modality effect becomes comprehensible if we consider the qualitative differences between reading and listening. Reading is an inherently self-paced activity while listening typically requires to follow some speaker's pace. In this view, spoken verbal explanations are more compatible with system-paced instructions than written explanations. Thus, the modality effect may be restricted to system-paced instructions due to particularities of the reading task. One may argue that although exceptional in daily life we are also able to adjust our reading behavior to external requirements as for example in subtitled television (d'Ydewalle et al., 1991). However, in the studies of this thesis learners exhibited a different reading behavior in system-paced compared to self-paced instructions, which may indicate a change in cognitive strategies. In the case of instructed learning, individually chosen reading time does not only reflect text comprehension abilities but also the contribution of deliberate, probably effortful, strategies for remembering expository text (Hartley, et al., 1994). Thus, self-paced presentation allows the learner to engage in a more elaborated processing of verbal explanation. Consequently, written text must be assumed to be more compatible with self-paced than with system-paced instructions. Furthermore, written text may be more compatible with self-paced instructions than spoken text since it facilitates strategic behavior for processing and remembering text. In this view, written text presentation may also be superior to spoken text presentation. Actually, there already exists empirical evidence for such a “reversed” modality effect in self-paced instructions (Tabbers, 2002).

Taken together, the theoretical considerations taken in the face of the empirical evidences do challenge cognitive load theory and Mayer’s cognitive theory of multimedia learning. Before turning to the directions for further research in order to advance these theoretical approaches I will shortly consider some implications for the design of multimedia instructions.

Practical implications

The first practical implication that can be derived from the empirical evidences and its theoretical implications is that the scope of application of the modality principle needs to be specified. The recommendation to use spoken rather than written text whenever it is accompanied by a visualization appears to be restricted to learning situations in which the time to retrieve information from both sources is severely limited. In fact, the guideline is derived from results of experiments in which instructions were used with a pacing based on the pace and length of the spoken text (e.g. Brünken & Leutner, 2001; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi et al., 1995; Tindall-Ford et al., 1997). Under these conditions the recommendation still holds.

Sometimes, however, it appears desirable for an instructional designer to use written rather than spoken text, for example under economic considerations. Producing audio and implementing it into multimedia instructions is time-consuming, laborious, and expensive. Furthermore, 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. Hence, the designer of multimedia instructions would like to be sure that there is no alternative to the use of spoken text in order to exploit the technical possibilities and to promote learning.

The studies conducted in this thesis and their theoretical implications allow suggesting when written text can be at least as effective as spoken text. The explanations for the superiority of spoken over written text presentation in time-limited presentation offered in the previous sections were that (a) the perceptibility especially for the visualization is decreased, and (b) written text comprehension is disturbed. As a consequence, a modality effect can be avoided if both information sources are sufficiently perceptible and/or if the design of a multimedia instruction ensures not to bother a regular reading behavior.

In order to make sure that all information sources in a multimedia instruction are sufficiently perceptible the instructional designer must consider the split of visual attention that occurs whenever two or more visual information sources are presented concurrently. In fact, current design guidelines already recommend to present written text near rather than far from visualizations in order to minimize split-attention. However, even if written text is presented in this manner it still drags visual attention away from accompanying visualizations.

The risk to miss important visual information due to this competition can be further decreased by reducing the pace of instruction. How can an instructional designer determine an "appropriate" pacing a priori? First, we need to consider that reading speed is reduced under learning instructions (Hartley et al., 1994) and apparently slower than the normal rate of speech. Furthermore, reading speed varies with text characteristics like word frequency, word length, length of sentences and phrases, etc. (Just & Carpenter, 1987). Based on these characteristics there already exist some metrics to estimate text difficulty (e.g. Smith & Kincaid, 1970; Thomas, Hartley, & Kincaid, 1975; Wagenaar, Schreuder, & Wijlhuizen, 1987). Characteristics that allow estimating the time needed to perceive and process visualizations are less explored and elaborated. Preliminary, however, we may conclude that dynamic visualizations are more difficult to perceive since they are more transient than static ones (Tversky, et al., 2002). In addition, dynamics in a visualization can further reduce the time that is spent reading. Finally, the effectiveness of written text presentation depends on the complexity of referential connections between text and visualization. Some advances to estimate this "element interactivity" have already been taken place (Tindall-Ford et al., 1997).

The thesis, however, highlighted a way to care for an appropriate pacing of instruction without a priori estimates of the above-mentioned characteristics: learner-paced instruction! Instead of *specifying* a system-controlled pace that may be appropriate for an average learner, this minimal form of user interaction allows each learner to *adjust* the pace of presentation to her individual needs. Doing so, the learner can ensure that she captures all information that is displayed. In fact, the thesis revealed that learners were capable of adjusting the pace of presentation in order to learn equally successful no matter if text was written or spoken. Furthermore, an individually chosen pace allows the learner to follow a regular reading strategy. In this view, learner-control is not only recommendable to overcome difficulties with written text presentation in multimedia learning. Since reading is more susceptible to cognitive strategies than listening, learner-paced instructions can even *benefit* from written text presentation.

Directions for further research

The studies in this thesis challenged the theoretical substantiation as well as the practical scope of the modality principle. In the previous sections I pointed out, how the cognitive theories and the design principles derived from them may be further specified in order to account for the current results. 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.

First of all, I concentrated on a well-established learning material to provide a maximal comparability with other studies on the modality effect (e.g. Mayer & Chandler, 2001; Mayer & Moreno, 1998; Moreno & Mayer, 1999). The multimedia instruction on the formation of lightning can surely be considered to be a prototypical case of the application of dynamic visualizations. However, the studies are worth being replicated with other learning material. For example, replicating the studies with the materials used by Sweller and his colleagues will help specifying the impact of my findings on the current theoretical approaches. Getting the same results with their instructions would indicate the generalisability of the interactions of pacing and its control with text modality and further stress the importance of visual processes in multimedia learning.

Moreover, the method of eye tracking can be applied to different learning material. In the material used in the studies of this thesis the patterns of viewing behavior highlighted the role of text comprehension. Concentrating on the verbal explanations appeared reasonable in the present instruction since the visualizations were fairly concrete. However, one can easily imagine more complex and/or abstract pictorial information, for example electric circuits or statistical graphs that require more processing resources. Similarly, text difficulty depends on the text structure and the subject matter. Furthermore, multimedia learning material can differ with respect to the referential connections between text and visualizations. It can be assumed that all these characteristics of a learning material affect the learner’s viewing behavior. The differences in viewing behavior can help estimating the relative load of verbal explanations in comparison to visualizations and the amount of “element interactivity” (Tindall-Ford et al., 1997) across different learning materials.

There is another exciting design characteristic of multimedia instructions in which eye tracking *must* be employed. In order to reduce the perceptual and cognitive load caused by high element interactivity, visual cues can be used to guide visual attention to appropriate referents (Kalyuga et al., 1999). When those design features are purposely introduced, observing the actual fixation paths allows to evaluate if these features were effective in advancing attention allocation and reducing visual search.

Another aspect of eye tracking is that it offers an extensive database. In fact, there are numerous ways in which those data can be analyzed. For example, in reading research viewing behavior is usually described in terms of gaze durations or even single fixations on words and the saccadic movements between these gazes or fixations (e.g. Just & Carpenter, 1980). Such analyses are accompanied by theoretical models accounting for eye movements on the same level of description (e.g. Reichle, Pollatsek, Fisher, & Rayner, 1998). These fine-grained cognitive process models can be tested by tracing the eye-movement protocol (e.g. Salvucci & Anderson, 2001). Matter-of-factly, current theories on the integration of verbal and pictorial information are less elaborated. But they may be advanced in order to allow predictions of fixation paths based on an accurate model of the learning process.

Considering a more practical aspect, the research questions of my thesis must be successively extended to broader classes of multimedia learning material in order to estimate the practical scope of the findings. There already exist some studies that varied the pacing of instruction and the matter of learner-control in multimedia learning with a linearly structured website containing texts and diagrams (Tabbers, 2002). Although the material was much more complex than the instruction used in this thesis, the effects on cognitive load and learning outcomes reported in these studies are fairly in accordance with the present results. Most notably, however, Tabbers found a “reversed” modality effect. With self-paced instructions students learning from a version where text accompanying a diagram was presented on-screen outperformed those students who received spoken text. I deduced the possibility of a reversed modality effect from considerations based on the viewing behavior observed in system- vs. self-paced instructions. The rationale for such an effect is that reading is more accessible for deliberate strategies for remembering expository text than listening. This difference may not affect learning success with single instructions of an approximate length of 3 minutes. However, written text appears superior to spoken text when the amount of displayed information is increased, as done in the studies by Tabbers. The average time on task learners spent in his studies was above 20 minutes. It appears worth examining the amount of content and especially the amount of text that is necessary to evoke such a reversed modality effect in self-paced instructions.

The results obtained with self-paced instructions also underline the importance of extending the research to more interactive learning environments. As supported by the studies referenced in the previous paragraph, effects that apply under more strict system-paced conditions might not work or have different outcomes when learners interact with the program (Tabbers, 2002). Thus, other forms of interactivity than control over the pacing should be investigated as well. For example, giving the learner the choice over the mode of text presentation (Plass, Chun, Mayer, & Leutner, 1998), or offering further navigational interaction like stopping and replaying, allow the learner to adjust a presentation to individual preferences. How those interactions further promote learning or if some of these interactions put an additional cognitive load on the learner has yet to be investigated.

Final remarks

The thesis started with the aim to take a closer look at visual attention allocation, cognitive load and learning outcome in learning from dynamic visualizations with accompanying verbal explanations. Introducing measures of visual attention shifted the view from learning outcomes via cognitive load to *perceptual* aspects of the learning material. The question is not that much if or if not text should be presented aurally instead of visually but if the displayed information can be sufficiently extracted by an individual learner. The studies revealed that under certain circumstances it is still recommendable to present text in spoken rather than written form. Exploiting the possibilities of computer technology, this recommendation appears to be one of minor priority. Understanding the demands of a learning material on the learner’s perception and accounting for individual differences for example by implementing user interaction appears much more promising to advance the design of multimedia instructions in a learner-supporting fashion.

References

Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction, 14,* 241-255.

Allport, A. (1989). Visual attention. In M.I.Posner (Ed.), *Foundations of cognitive science* (pp. 631-682). Cambridge, MA: MIT Press.

Baddeley, A. (1986). *Working memory.* Oxford: Clarendon Press.

Beck, C. R. (1984). Visual cueing strategies: Pictorial, textual and combinational effects. *Educational Communication and Technology Journal, 32,* 207-216.

Bernard, R. M. (1990). Using extended captions to improve learning from instructional illustrations. *British Journal of Educational Technology, 21,* 212-225.

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations*. Learning and Instruction, 14,* 325-341.

Brünken, R., & Leutner, D. (2001). Aufmerksamkeitsverteilung oder Aufmerksamkeitsfokussierung? Empirische Ergebnisse zur "Split-Attention-Hypothese" beim Lernen mit Multimedia. [Split attention or focusing of attention? Empirical results on the split-attention hypothesis in multimedia learning.]. *Unterrichtswissenschaft, 29,* 357-366.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38,* 53-61.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32,* 115-132.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49,* 109-119.

Carroll, P. J., Young, R. J., & 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.

Chandler, P. (2004). The crucial role of cognitive processes in the design of dynamic visualizations. *Learning and Instruction, 14,* 353-357.

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.

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

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.

d'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television. Automatic reading behavior. *Communication Research, 18,* 650-666.

Faraday, P., & Sutcliffe, A. (1996). An empirical study of attending and comprehending multimedia presentations. In: *Proceedings ACM Multimedia 96* (pp. 265-275). Boston MA.

Fink, A. & Neubauer, A. C. (2001). Speed of information processing, psychometric intelligence, and time-estimation as an index of cognitive load. *Personality & Individual Differences, 30,* 1009-1021.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26,* 519-532.

Graesser, A. C., & Britton, B. K. (1996). Five metaphors for text understanding. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text.* (pp. 341-352). Mahwah, NJ: Erlbaum.

Graesser, A. C., Hoffman, N. L., & Clark, L. F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior, 19,* 135-151.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48,* 163-189.

Guthrie, J. T. (1973). Reading comprehension and syntactic responses in good and poor readers. *Journal of Educational Psychology, 65,* 294-299.

Hartley, J. T., Stojack, C. C., Mushaney, T. J., Annon, T. A. K., & Lee, D. W. (1994). Reading speed and prose memory in older and younger adults. *Psychology and Aging, 9,* 216-223.

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

Hegarty, M. (1992b). 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.

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction, 14,* 343-351.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32,* 717-742.

Hegarty, M., Quilici, J., Narayanan, H. N., Holmquist, S., & Moreno, R. (1999). Multimedia instruction: Lessons from evaluation of a theory-based design. *Journal of Educational Multimedia and Hypermedia, 8,* 119-150.

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception and Psychophysics, 55,* 399-411.

Jackson, M. D. & McClelland, J. L. (1975). Sensory and cognitive determinants of reading speed. *Journal of Verbal Learning and Verbal Behavior, 14,* 565-574.

Jackson, M. D. & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology.: General, 108,* 151-181.

Jeung, H.-J., 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. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Newton: Allyn and Bacon.

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

Kaiser, M. K., Proffitt, D. R., & Anderson, K. (1985). Judgments of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 795-803.

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.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3,* 185-226.

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

Large, A., Beheshti, J., Breleux, A., & Renaud, A. (1996). Effect of animation in enhancing descriptive and procedural texts in a multimedia learning environment. *Journal of American Society for Information Science, 47,* 437-448.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11,* 65-100.

Levie, W., & Lentz, R. (1982). Effects of text illustration: A review of research. *Educational Communication and Technology Journal, 30,* 195-232.

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating function of pictures in prose. In D. M. Willows & H. A. Hougthon (Eds.), *The psychology of illustration, Vol.1* (pp. 51-85). New York: Springer.

Macromedia (1999). Flash (Version 4.0) [Computer software]. San Francisco: Author.

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

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

Mayer, R. E. (2001). *Multimedia learning.* Cambridge: UK: 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. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38,* 43-52.

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

McCloskey, M. & Kohl, D. (1983). Naive physics: the curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9,* 146-156.

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

Narayanan, H. N., & Hegarty, M. (2002). Multimedia design for communication of dynamic information. *International Journal of Human-Computer Studies, 57,* 279-315.

Paas, F. G., & Van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35,* 737-743.

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

Paas, F. G., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38,* 63-71.

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

Park, O., & Hopkins, R. (1993). Instructional conditions for using dynamic visual displays: a review. *Instructional-Science, 21,* 427-449.

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

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

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.

Reed, S. K., & Saavedra, N. C. (1986). A comparison of computation, discovery, and graph procedures for improving students' conception of average speed. *Cognition and Instruction, 3,* 31-62.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125-157.

Reitmayr, E. (2003). *Guiding attention in a software tutorial: using an active help system to reduce the split-attention effect.* Unpublished Diploma-Thesis, University Mannheim.

Rieber, L. P. (1989). The effects of computer animated elaboration strategies and practice on factual and application learning in elementary science lesson. *Journal of Educational Computing Research, 5,* 431-444.

Rieber, L. P. (1990a). Using computer animated graphics with science instruction with children. *Journal of Educational Psychology, 82,* 135-140.

Rieber, L. P. (1990b). Animation in computer-based instruction. *Educational Technology: Research and Development, 38,* 77-86.

Rieber, L. P., & Hannafin, M. J. (1988). The effects of computer textual and animated orienting activities and practice on learning from computer-based instruction. *Computers in the Schools, 5,* 77-89.

Rieber, L. P., Tzeng, S.-C., & Tribble, K. (2004). Discovery learning, representation, and explanation within a computer-based simulation: finding the right mix. *Learning and Instruction, 14,* 307-323.

Rost, D. H., & Hartmann, A. (1992). Lesen, Hören und Verstehen. *Zeitschrift für Psychologie, 200,* 345-361.

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

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: learning to tie nautical knots. *Learning and Instruction, 14,* 293-305.

Smith, E. A., & Kincaid, J. P. (1970). Derivation and validation of the automated readability index for use with technical materials. *Human Factors, 12,* 457-464.

Sticht, T. G., & James, J. H. (1984). Listening and reading. In P. Mosenthal, M. L. Kamil, & R. Barr (Eds.), *Handbook of reading research.* (pp. 293-317). New York: Longman.

Swaak, J., & de Jong, T. (2001). Learner vs. system control in using online support for simulation based discovery learning. *Learning Environments Research, 4,* 217-241.

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

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 Merrienboer, J. J. G., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-196.

Tabbers, H. K. (2002). *The modality of text in multimedia instructions. Refining the design guidelines.* Open University of the Netherlands Heerlen.

Tabbers, H. K., Martens, R. L., & Van Merrienboer, 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.

Tabbers, H. K., Martens, R. L., & Van Merrienboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74,* 71-81.

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

Thomas, G., Hartley, R. D., & Kincaid, J. P. (1975). Test-retest and inter-analyst reliability of the Automated Readability Index, Flesh Reading Ease Score, and the Fog Count. *Journal of Reading Behavior, 7,* 149-154.

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

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57,* 247-262.

Underwood, G. (1998). *Eye guidance in reading and scene perception.* Oxford, UK: Elsevier.

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology, 57A,* 165-182.

Van der Heijden, A. H. C. (1996). Visuelle Aufmerksamkeit. In Neumann, O. & Sanders, A. F. (Eds.). *Aufmerksamkeit.* Göttingen: Hogrefe, 1996, S. 7-60.

Van Gerven, P. W. M., Paas, F. G., Van Merrienboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41,* 167-174.

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.

Wagenaar, W. A., Schreuder, R., & Wijlhuizen, G. J. (1987). Readability of instructional text, written for the general public. *Applied Cognitive Psychology, 1,* 155-167.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology, 57A,* 129-163.

Weidenmann, B. (2002). Multicodierung und Multimodalität im Lernprozess. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia und Internet* (3 ed., pp. 45-62). Weinheim: Psychologie Verlags Union.

Yantis, S. (1998). Control of visual attention. In H. Pashler (Ed.), *Attention* (pp. 223-256). Hove, England: Psychology Press.

Zuberbuehler, H.-J. (1999). *Vergleich der Lernwirksamkeit von Animationen und Standbildern.* Unpublished Diploma-Thesis, Eidgenössische Technische Hochschule Zürich.

Author's notes

The research of this thesis would have been impossible to conduct without the support of many people that need to be mentioned here. First of all I whish to thank my advisor Professor Ulrich Glowalla for his manifold contributions to my scientific work in general and my dissertation project in particular. I also thank Michaela Flegel, Bettina Linke, Miriam Herder, Andrea Heiden, Silke Schmidt, and Kathrin Jacobowski for their precious help in collecting the data. You girls did a great job. I thank Magdalena Kowalski, Angela Schröder, Meike Herder, and Maria Siegert for their helpful comments on earlier versions of different parts of this thesis.

An invaluable support to this thesis was supplied by the people and the technical infrastructure of the research group *Instruction and Interactive Media* (IIM). Just to name two examples, Arno Latten introduced me into the eyetracking laboratory, and Alfred Kohnert developed and programmed the initial version of the learning material and supported me in realizing the experimental variations with the actual material.

Alfred Kohnert deserves a special credit. He was quite more than a technical supporter and a reviewer of this manuscript. Without his precious help – intellectual *and* emotional – from the very beginning, I never would have finished this thesis.